

Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction

Energy Exploration & Exploitation

2019, Vol. 37(4) 1426–1451

© The Author(s) 2019

DOI: 10.1177/0144598718822400

journals.sagepub.com/home/eea



Zhijian Liu¹ , Di Wu¹, Yuanwei Liu¹,
Zhonghe Han¹, Liyong Lun², Jun Gao³,
Guangya Jin¹ and Guoqing Cao⁴

Abstract

It is of great significance to achieve the prediction of building energy consumption. However, machine learning, as a promising technique for many practical applications, was rarely utilized in this field. The most important reason is that the predictive structure with best performance is difficult to be determined. To fill the gap, this paper offers one in-depth review, which focuses on the accuracy analyses and model comparisons. Specifically, the accuracy analyses were conducted based on different types of buildings (e.g. residential building, commercial building, government building or educational building), different type of temporal granularity (e.g. sub-hourly, hourly, daily or annual), as well as input/output variables and historical data collections. Further, artificial neural network (ANN) and support vector machine (SVM), as the epidemic models, were compared in terms of their complexity of prediction processes, accuracies of results, the amounts of required historical data, the numbers of inputs, etc. Then the hybrid and single machine learning methods were outlined and compared in terms of their strengths and weaknesses. In addition, several vital defects and further research directions are presented from a multivariate perspective. We hope

¹Department of Power Engineering, School of Energy, Power and Mechanical Engineering, North China Electric Power University, Baoding, China

²Resources & Environment Business Department, China International Engineering Consulting Corporation, Ltd., Beijing, China

³School of Mechanical and Energy Engineering, Tongji University, Shanghai, China

⁴Institute of Building Environment and Energy, China Academy of Building Research, Beijing, PR China

Corresponding author:

Zhijian Liu, Department of Power Engineering, School of Energy, Power and Mechanical Engineering, North China Electric Power University, Baoding 071003, China.

Email: zhjianliu@ncepu.edu.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

that machine learning method could capture more attention from investigators via our introduction and perspective, due to its potential development of accuracy and reliability.

Keywords

Machine learning method, artificial neural networks (ANNs), support vector machines (SVMs), energy consumption prediction, building energy saving

Introduction

How to reduce energy consumption has been considered as a prominent factor that could influence the economic growth, since the energy demand from series of buildings remains at a high level. In both US and EU, energy demand of residential and commercial buildings accounts for 40% of total energy demand (Huebner et al., 2015; Pérez-Lombard et al., 2008). In China, this proportion is up to 30%, moreover, 63% of that is utilized for space heating and cooling (Huebner et al., 2015). Also, energy consumption for building sectors account for approximately 30–40% of the primary energy consumption in China (Aras, 2008; Zhang et al., 2015). Accordingly, various approaches for predicting the building energy consumption have been explored by scientists (Cuce et al., 2016; Huebner et al., 2015). On the basis of the amount of energy consumption predicted, smart energy management and building energy efficiency retrofitting could be achieved. Therefore, the prediction of the building energy consumption with accuracy and convenience has significant importance.

However, to date, there are still some serious obstacles in a predictive process of building energy: (i) Which method is the best or suitable approach to achieve the prediction in terms of the accuracy and convenience? (ii) Which model or algorithm should be recommended or adopted in the prediction process for different cases? Technically, the most important puzzle is to find out the approaches that could perform ultra-fast predictions and/or assessments, with good accuracy. The normal approaches are engineering or statistical methods, which were often used in many domains for performance evaluation. However, other novel methods, such as machine learning, were rarely utilized.

Engineering methods are comprehensive methods, which consists of several partial differential equations or needs to apply the physical principles, as well as thermal dynamics equations (Khosravani et al., 2016; Li et al., 2019) and density functional theory (Li et al., 2018c; Li et al., 2017a; Liu et al., 2018b). Fortunately, due to the dramatic advances of information technologies recently, many software and packages could provide solutions for the complicated shape definition based on partial differential equations. Meanwhile, that promotes the development of numerical analysis of dynamics in engineering practice. Therefore, we could easily obtain the numerical solutions through software products, such as TRNSYS (Razavi et al., 2018), ANSYS (Shen et al., 2017), FLUENT (Yaşa and Ok, 2014), and COMSOL (Embaye et al., 2015). Engineering methods based on large amounts of statistical data could achieve the highly accurate estimation (Jia et al., 2016, 2017, 2018). Many parameters as the inputs are required that may include building construction, external climate conditions, and performances of HVAC equipment for the calculation of energy consumption. Normally, engineering method could be regarded as the

assumptions built on statistical information. Thus, we could try to develop an engineering model or equation through available statistical data for other regions (Hu et al., 2015a, 2015b; Zhang et al., 2018c). However, there is a series of complex processes including the establishment of the models and equations, the setting of parameters and boundary conditions, which are difficult to satisfy the requirement that we aim to acquire the results instantly and conveniently. In addition, many building and environmental parameters as mentioned above are unavailable. Besides, especially in some unknown regions, we have no sophisticated theories for reference, which may become a huge challenge for the development of engineering method.

Statistical methods build the relationship between the energy consumption and influencing variables through the empirical models based on the historical data. For example, Bauer and Scartezzini (1998) built the relationship between the energy consumption and the climatic variables. Ansari et al. (2005) built the relationship between the cooling load and the performance of building envelope. Dhar et al. (1999) utilized Fourier series model to correlate heating and cooling loads with time and dry-bulb temperature. Aydinalp-Koksal and Ugursal (2005) predicted the building energy consumption on the national level through conditional demand analysis. Statistical methods are suitable for the non-linear relations and that avoid the defects of the engineering methods (Lomet et al., 2015; Lü et al., 2015). Nevertheless, large amounts of high-quality historical data required to be collected, moreover, long time-consuming and large computer memory need to be considered. Besides, it would cause inaccurate and insignificant results, if the selection for analytical method was inappropriate.

To avoid the defects of above two approaches, machine learning method was proposed and developed. "Machine Learning" evolved from artificial intelligence, whose aim is to build one good predictor with no complicated process or other strict conditions (Samuel, 1967). Hereinto, artificial neural network (ANN) model and support vector machine (SVM) model as common machine learn method are more reliable than other regression techniques or ordinary simulation models (Chen et al., 2017b). Researchers tend to predict energy utilization or system performance, such as sizing PV power system, air-conditioning, and other hybrid energy systems (Liu et al., 2018a). With the development of computer technology, machine learning method has been significantly developed and widely applied in more and more fields. Nevertheless, the utilization of machine learning method for energy conservation still remains in an infancy stage. Typically, ANN and SVM were used in the prediction of building energy consumption, while accuracy and simplicity remain to be improved. Therefore, we expect that machine learning method could effectively contribute to more troublesome problems in this field. Simultaneously, we hope that this novel method could capture more attentions from investigators via our introduction and perspective, and that the development of machine learning method could be further promoted.

We have already analyzed the advantages and disadvantages of engineering and statistical methods in this section. The characteristics of machine learning method were displayed in the later contents. To be specific, in the following section, the background of machine learning method was expounded, further, the accuracy analyses, in terms of prediction scope (including the type of building and temporal granularity), input and output variables as well as the historical data collected, were carried out. Due to the significant impact of models on predictive accuracies, ANN and SVM models are discussed in third section, which includes their structure, characteristics, prediction process, and accuracy comparison. Then in fourth section, the hybrid machine learning method is outlined. In fifth section, the advantages and

disadvantages of ANN, SVM, and hybrid models are compared. In the final section, the development direction and some puzzles needed to be solved urgently of machine learning are presented. This paper aims to review the approaches to estimate the building energy consumption through machine learning method, and some representative literatures are introduced, summarized, and commented.

Machine learning method

Development background

Machine learning method is a powerful technology that grew out of the exploration of artificial intelligence through the endeavor of scientists, which has huge potential for development (Li et al., 2015, 2016, 2017c, 2018a; Liu et al., 2017d, 2017e). Based on the historical data, with suitable model and algorithms, machine learning method could “learn” the non-linear relationship between the independent variables and target variables. Machine learning method started to flourish in the 1990s. It completes the prediction based on a computer, which is related to the computational statistics. In current technologically advanced period, machine learning method has been employed in many areas. As presented in Figure 1, the application field of machine learning mainly involves medical diagnosis, data analysis, optimization analysis, structural analysis, performance prediction, and information retrieval. Sattlecker et al. (2014) made use of machine learning method to recognize diagnostic

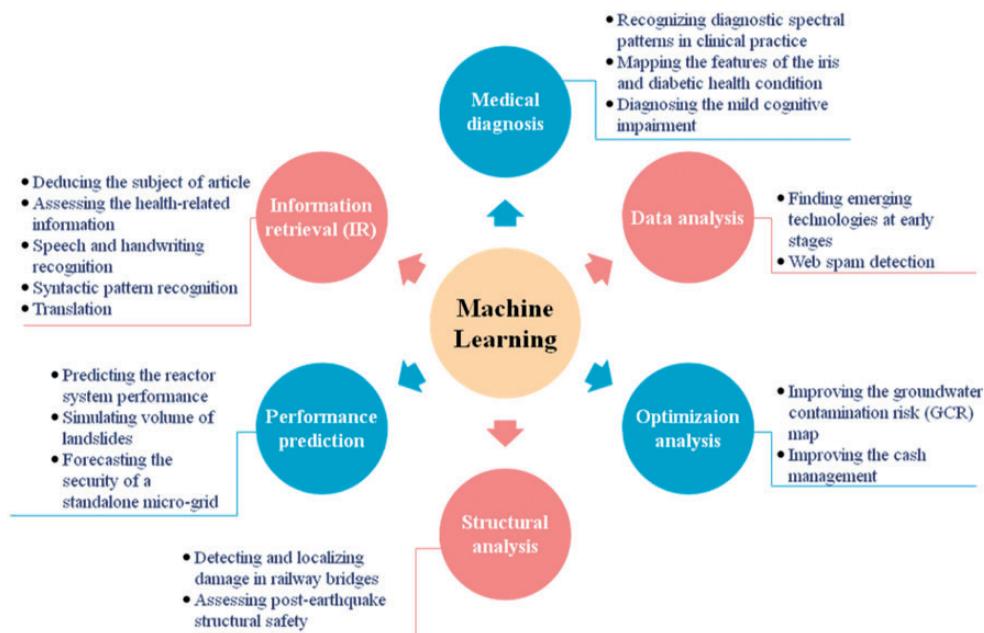


Figure 1. The application fields of machine learning (Azqueta-Gavaldón, 2017; Barzegar et al., 2017; Castiglioni et al., 2018; Chalouhi et al., 2017; Goh and Singh, 2015; Karim et al., 2018; Kim et al., 2017; Lázaro et al., 2017; Lee et al., 2017; Pham et al., 2016; Samant and Agarwal, 2018; Sattlecker et al., 2014; Shirzadi et al., 2017; Zeng et al., 2018b; Zhang et al., 2018a).

spectral patterns in clinical practice, and emphasized the importance of the routine spectral data for the reanalysis process. Within the process of predict genome-wide with complex traits, González-Recio et al. (2014) also suggest that machine learning method could achieve the prediction and classification, and deal with the multidimensional problems. In addition, Pham et al. (2016) and Shirzadi et al. (2017) verified that machine learning method is more accurate than statistical method for simulating volume of landslides (geology field). Moreover, they also indicated that SVM model has the best prediction performances than other models selected in this exploration. Researching large amount of related literature and investigating actual situation, we can see that machine learning method, as a predictive tool, has already occupied a broad market in many field. However, for the domain of a building energy consumption predicted, machine learning has not been widely utilized in practice. The most important reason is that the predictive structure with best performance is difficult to be determined, especially when the various or complex conditions was considered. Therefore, the prediction of building energy consumption should be further explored.

Prediction scope

The different prediction scope in terms of the type of building (residential building, commercial building, government building or educational building) and the type of temporal granularity (sub-hourly, hourly, daily or annual) have the close relationship with the predictive accuracy.

For a commercial building, the energy consumption mainly includes the cooling, heating, and electricity, which follows a certain rule of consumption. For example, the energy consumption accounts for higher proportion during the workdays and the daytime, while energy consumption is less during the weekend and night. Based on the obvious energy consumption period, the energy prediction is easier to achieve (Jiang and O'Meara, 2018). For a residential building, the amount of energy consumption is relied on the occupant behavior more obviously, whereas, there is no regular schedule of activities for different residents or different periods (Huo et al., 2019). Besides, due to the difference in performances of residence (such as thermal insulation, heating or cooling demand, building envelope), various prediction structures need to be provided under different case. That leads to an obstacle that the historical data is difficult to be monitored (Calero et al., 2018). Given above mentioned, the prediction for energy consumption of commercial building is more accurate and easier than that of residential building. However, the importance of prediction for residential building should not be neglected, because of the high proportion of the amount of residential buildings in the total buildings. In addition, the educational building and government building have the similar features with commercial building, thus, their energy consumption is easy to be predicted (Al-Saadi et al., 2017; Heracleous and Michael, 2018).

The determination of the prediction time scale depends on the purpose of the research and the sampling interval of sensors. Most literature focused on the short-term prediction, such as the hourly or daily energy consumption, and that usually studied only one prediction time scale. The research about long-term prediction (e.g. annual prediction) was less. It can be explained that, on one hand, the short-term prediction could satisfy the current requirement of energy consumption. On the other hand, many models (e.g. ANN [Li et al., 2017]) have the higher predictive accuracy only in the short-term scale. Besides, many uncertainties may occur in the long time span, which could influence the overall

accuracy of prediction. Despite these challenges existing, energy consumption prediction models in long-term scale are essential. It is necessary especially in the decision process of long-term, such as energy supply strategy, capacity expansion, and capital investment.

Input/output variables and historical data collected

The suitability of input variables selected determines the predictive accuracy. The suitability mainly refers to which variables should be selected as the input, as well as the number of input variables. It can be seen in Table 1 that the inputs mainly include weather conditions, building envelope, and occupant behaviors. For the weather conditions, it generally includes outside temperature, humidity, solar radiation, and wind speed. For the building envelope, it usually includes window to wall ratio, glazing area, surface area, and heat transfer coefficient of building walls. Noted the input variables mentioned in the literature, it can be found that two essential features of input variables should be satisfied. On one hand, the inputs selected are related with the outputs as far as possible. On the other hand, the inputs should be easy-measured and are all relevant to the targets (Liu et al., 2018c; Wong et al., 2013). The number of inputs has the influence on the complexity of the model structure, therefore, the number of inputs and the scale of database directly affect the suitability of model. Increasing the number of inputs will cause the complexity ulteriorly of models and prolong the time-consumption, while reducing the number will likely decrease the accuracy degree of prediction (Hautier et al., 2010; Zeng et al., 2018a). In order to maintain a balance between these two points, only the inputs that close relate with the targets should be considered. Further, that would be finally determined by the process of training and testing, which was described below. In summary, the suitable input selected has significant importance to the target predicted both the accuracy and the simplicity. As the predictive target or output variables, it could be the heating/cooling load or electricity load or the overall energy for the whole building or for the sub-level components (Table 1). Most researches focus on the prediction for the load of the whole building, which is easier than the prediction on the sub-components level. Because the sub-components of energy system are complex, even there are no clear boundaries between each part.

Collecting the representative historical data is important to determine the reliable model structure. According to the character of data, it can be mainly divided into the real data and simulated data. The real data include the data obtained through collecting device, such as thermocouple, chemical sensors, solar radiation tester, infrared thermal imager, and ambient sensors. The real data collected should be further dealt with before training. Because the raw data may include the inaccurate data, useless variables, non-unified time interval, noisy, and missing data points. Collecting the appropriate data has the contribution to the effectively solve the predictive problems. Accordingly, some aspects, such as the universality and representativeness for historical data for training as well as the appropriateness for model and input variables, should all be deliberated.

Model structure

Different models or algorithms substantially influence the accuracy and reliability of results, thus, their determination is extremely important. ANN and SVM as the epidemic models were utilized frequently by many investigators in different fields. They have strong capacity

Table 1. Prediction characteristics of different literature.

References	Input	Output	Type of data	Building type	Model structure	Temporal granularity	Evaluation criteria
Magalhães et al. (2017)	Building archetype; geographical location; heating pattern; indoor heat gains; set point temperature	Heating energy use; indoor temperatures	Simulation	Residential	ANN	Annual	R2; MSE; MAE; MAPE
Paudel et al. (2014)	Climatic conditions (outside temperature and solar radiation); Occupancy profile; operational Heating power level characteristics	Building heating	Real	One large building	ANN	Several daily	MSE; R2
Neto and Fiorelli (2008)	Dry temperature; relative humidity; solar radiation	Entire energy consumption	Real	Administration building	ANN	Daily	Compared with software
Dong et al. (2005)	Dry-bulb temperature, relative humidity and global solar radiation	Entire energy consumption	Real	Commercial building	SVM	Monthly	MSE; S-MSE; percentage error; CV-RMSE
Mihalakakou et al. (2002)	Air temperature and total solar radiation	Entire energy consumption	Real	Residential building	ANN	Hourly	Compared with measurement data
Domhaci (2010)	Hourly outside temperature values of 4 years	Heating	Real	Residential	ANN	Hourly	RMSE; R2; MAPE
Li et al. (2009)	Dry-bulb temperature, relative humidity and solar radiation intensity	Cooling	Real	Official	SVM	Hourly	RMSE; MRE
Li et al. (2010)	Performances of building envelope (such as mean heat transfer coefficient	Electricity consumption	Real	Residential	SVM; BPNN, RBFNN, GRNN	Annual	RMSE; MRE

(continued)

Table 1. Continued.

References	Input	Output	Type of data	Building type	Model structure	Temporal granularity	Evaluation criteria
Fu et al. (2015)	of building walls, and mean thermal inert index of building walls)	Electricity load (airconditioning system, lighting system, power system, and other equipment)	Real	Public building	SVM; ARIMAX; DT; ANN	Hourly	RMSE; N-MBE
Li et al. (2010)	Weather conditions predicted (hourly dry-bulb temperature; hourly dew point temperature); two previous days' electricity loads	Building cooling load	Real	Non-residential	SVM; FCM-SVM; FCM	Hourly	MAPE; RMSE
Fan et al. (2014)	Dry-bulb temperature; relative humidity; pressure; the amount of cloud; dew point temperature; total rainfall; solar radiation; number of hours of reduced visibility; total evaporation; wind speed	Energy consumption of next day and the peak power demand	Real	Commercial	HybridMLR, SVM, BT, ARIMA, RF, MLP, MARS, kNN	Daily	MAPE
Chou and Bui (2014)	Relative compactness; surface area; wall area; roof area; overall height; orientation; glazing area; glazing area distribution	Cooling	Real	Residential	ANN; SVM; DT; GLR	Hourly	RMSE
		Heating			ANN; SVM; DT; GLR		

R²: predictive coefficient of determination; MSE: mean square error; MAE: mean absolute error; MAPE: mean absolute percent error; S-MSE: mean squared error of scaled value; CV-RMSE: coefficient of variance based on root mean squared error; MRE: mean relative error; N-MBE: normalized mean bias error.

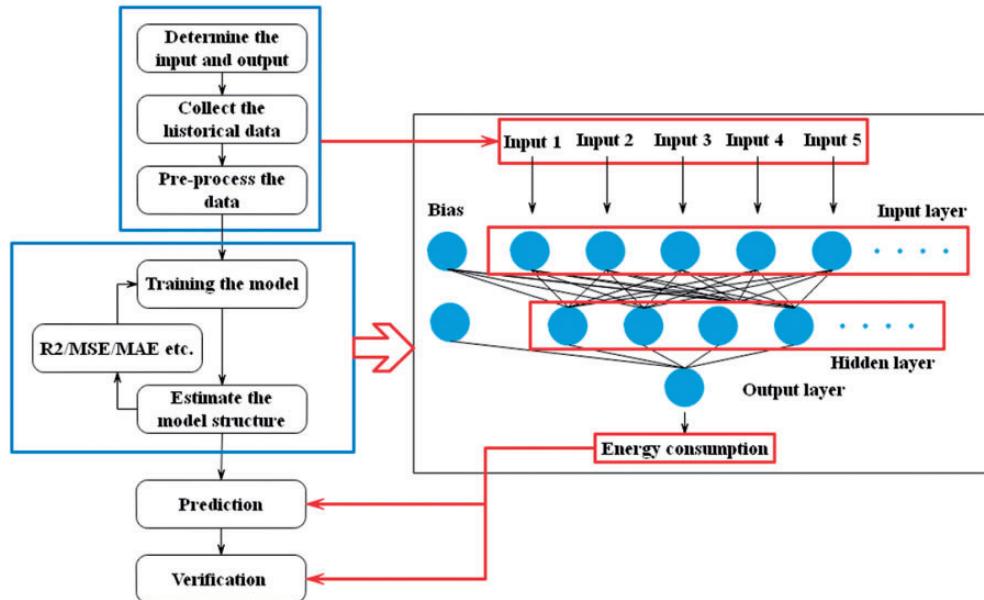


Figure 2. Schematic structure of an ANN model (Medler, 1998).

to capture the complex non-linear relationships between variables of input and output (Kalogirou et al., 2000).

Artificial neural networks

Among the several model structure, ANN model has been the most widely applied in practice. The reason is that ANN model has the robustness, which can effectively solve the non-linear and complex problems. ANN structure is similar with the formation of “neurons”, which could be divided into three parts, i.e. input layer, hidden layer, and output layer (Figure 2; Walker, 1990). Each neuron in a layer is connected to all neurons in the previous layer. In the training process, biases and weights were adjusted to produce the accurate output. Then, the error value between the expected and predictive output was calculated. Based on the error value, whether the model structure is suitable for the studied case was identified. Training cycles would be implemented repeatedly until reaching the accuracy requirement. The schematic form of ANN model is presented in Figure 2.

ANN model could acquire the experimental knowledge in the training process, so that it performs better predictive capacity and resistance ability to noise and errors. Besides, through adjusting the weights between different elements, ANN model could adapt to different cases. Magalhães et al. (2017) predicted the heating energy use and indoor temperatures by utilizing ANN model based on occupant behaviors and reference conditions, with the R^2 value above 0.98. Paudel et al. (2014; Aydin, 2018) predicted the building heating demand through the ANN model based on climatic conditions, occupancy profile, and operational heating power level characteristics, with R^2 value reaching 0.85. Kalogirou et al. (2001) used several climate performances coupled with ANN to predict daily cooling and heating load of the buildings. Their results obtained from prediction and from

simulation software TRNSYS were well agreed. Besides, Mihalakakou et al. (2002) took advantage of ANN model to predict the entire energy consumption of one residential building. Through analyses for the comparison between the predictive data and the measurement data, the reliability of ANN model is further confirmed. Dombayci (2010) achieved the heating energy prediction of one residential building through ANN model. In the testing phase, the RMSE, R^2 , and MAPE value were, respectively, 1.2125, 0.9880, and 0.2081, which also indicates the high accuracy of ANN model.

With regard to non-linear problems, ANN model is better than conventional statistical methods. ANN could implicitly identify the non-linear relationship between input and output with high accuracy. In addition, ANN model has the ability to evaluate the effect of socio-economic factors on energy consumption (Aydinalp et al., 2002). However, to achieve the good predictive power, large amount of historical data is necessary to be provided for training, which may consume a lot of manpower and time. Moreover, ANN model may encounter over-fitting problem in the training process, which is bad for the predictive results (Papadopoulos et al., 2000; Tu, 1996).

Support vector machines

Based on Vapnik–Chernoverkis theory, SVMs have been widely used in the application of forecasting, classification, and regression. SVMs could be regarded as one generalized classifier, which was an extension of the prediction. SVM model has the ability to solve the non-linear problems even with relatively less historical samples. Accordingly, this method is not especially relying on large amount of data for training.

The steps of SVM to predict the target are as follows (Figure 3). Firstly, transform data to the specific format. If the data instance stands for one certain categorical attribute, it needs to be converted into numeric data. Secondly, conduct the scaling on the data. The purpose is reducing the difficulty of data processing and calculation. Thirdly, select the kernel function. For typical kernel function, it includes linear function, polynomial function, radial basis function (RBF), and sigmoid function. Among these kernel functions, RBF could effectively map the input into a high-dimensional feature space, which is better to represent the non-linear relationship between the input and output. Therefore, RBF has been the general function usually in the prediction of the energy consumption. Finally, through cross-validation, penalty parameter, and kernel parameters are determined, which could avoid the over-fitting problem.

For RBF of SVM model, the cardinal steps of transformation of the non-linearity between the input and output are the utilization of linear mapping. The non-linear problem is projected into the high-dimensional space, then the function best fitted the high-dimensional space could be determined. Afterwards, the complex non-linear map could be transferred to linear problem. For RBF, it could not only solve the non-linear problem, but also has the relatively small complexity, which makes it widely applied. Therefore, when we predict the energy consumption by using SVM, the reasonable first choice of kernel function is RBF. Of course, if the number of features is relatively large, the RBF is not suitable, while the linear kernel may be the better selection.

SVM is known as the most accurate and robust machine learning algorithms (Wu et al., 2008). Li et al. (2009) predicted the hourly cooling load of one office building through the SVM model. Hou and Lian (2009) predicted the cooling load of HVAC system by using SVM model. Georgescu et al. (2014) built the relationship between the environmental

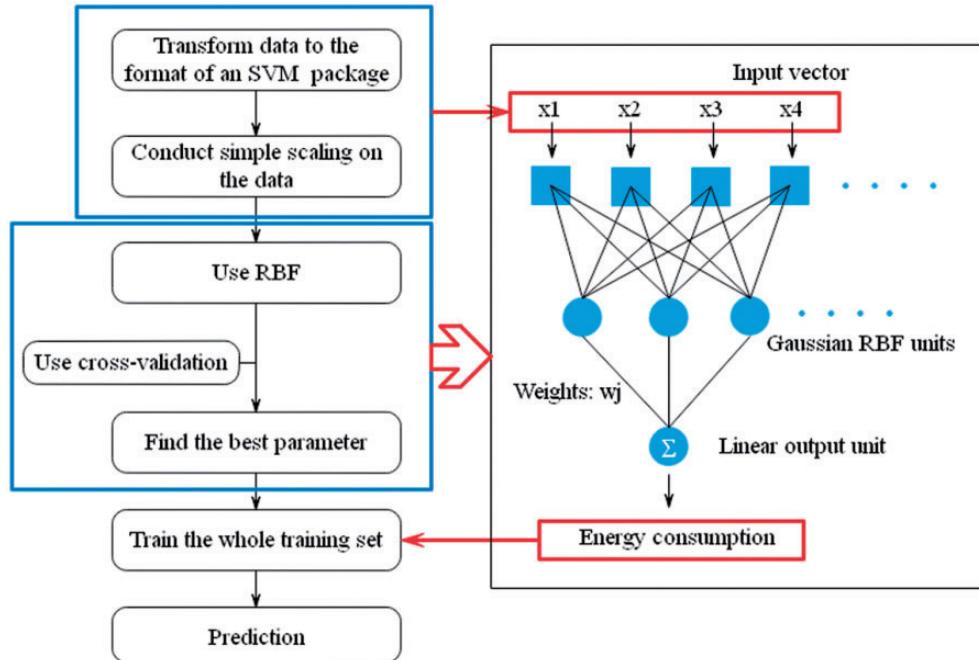


Figure 3. Schematic structure of SVM model.

variables and the building energy consumption through SVM model. Based on the weather forecasting, they predicted the energy consumption for a lot of buildings with higher accuracy that includes offices, laboratories, gymnasiums, dormitories, and restaurants constructed. Based on SVM model, Lai et al. (2008) achieve the accurate prediction of electrical consumption for residential building. For the prediction of electricity consumption, Dagnely et al. (2015) obtained the satisfied results from SVM model with higher accuracy.

On one hand, SVM model has higher generalization performance than conventional ANNs. On the other hand, SVM could search the optimal solution in the global region. Unlike other models, solutions from SVM generally do not be trapped into local region, which results in the trouble that take the local optimal solution as the best result in the global area. The disadvantage is that SVM could not achieve the classification directly, which must be conducted through the dimension conversion. Besides, the training period of SVM is long. For the large-size problems, the large amount of computation time should be reserved (Cao and Tay, 2003; Jun et al., 2017). Fortunately, two algorithms (i.e. sub-gradient descent and coordinate descent) were identified that they could deal with large, sparse datasets. Sub-gradient descent is efficient based on large amount of training examples, while coordinate descent is available under the high dimension of the feature space (Du et al., 2017; Nanculef et al., 2014).

Other model structure

Other machine learning models utilized in energy consumption of building mainly include genetic algorithm (GA), non-dominated solution GA, decision tree algorithm, Bayesian

network algorithm, and cluster algorithm. In addition, autoregressive (AR) moving average model, AR model, and autoregressive integrated moving average (ARIMA) model were also utilized.

Taking the GA as the example, GA has been successfully utilized for the optimization problems (Chen et al., 2017a). GA can imitate the process of biological evolution. To be specific, the next generations which have better properties could be reproduced from parent chromosomes by selection, crossover, and mutation operation. Further, the new offspring produces constantly until the termination condition is satisfied. Eventually, the last generations with better genes are the best solution for optimization problem. It is specialized in solving complex problems with multivariable or non-linear characteristics. GA could achieve the global searching, but they do not calculate objective function of all possible combinations (Abdmouleh et al., 2017), so its solutions may be trapped into local region. Moreover, the value of characteristic parameters, such as number of population and generations, mutation rate, and crossover rate has significant influence on the final results. In practice, Reynolds et al. (2018) combined the ANN and GA model to minimize the building energy consumption. Garshasbi et al. (2016) predicted the hourly energy consumption based on the hybrid GA method and Monte Carlo simulation. Reynolds et al. (2019) predicted the energy supply and demand relying on the ANN–GA model. However, the above parameters were determined only by experience without training process. Additionally, complicated programming also restricted their extensive application.

Hybrid machine learning method

Theory

According to the Gartner's 2016 hype cycle (Forni and Meulen, 2016), machine learning method is at its peak of inflated expectations. Because of the complexity of objective and the huge amount of samples, the application of hybrid machine learning method has developed promptly. Moss et al. (2012) combine the discriminated analysis and machine learning method for enhancing the classification and prediction ability for a range of chemicals. Through the prediction process for solar radiation, Voyant et al. (2017) suggest that hybrid method could give better results than single predictors. Marasco and Kontokosta (2016) indicated that machine learning method (based on falling rule lists) is a rapid estimated tool for the potential of energy efficiency. Cramer et al. (2017) proposed that machine learning method based on intelligent systems has the ability to predict rainfall. In the current era of big data, which instruments have the ability to classify the huge amount of data with quickly response speed will get the long-term development. Therefore, hybrid method based on machine learning method is regarded as the potential technology.

Hybrid machine learning method as the advanced data-mining technique has been also developed in the field of energy consumption. The hybrid machine learning could take advantages of every algorithm and simultaneously avoid their weaknesses to obtain the predictive target. In Wang et al.'s review, they divide the hybrid machine learning into two forms including the heterogeneous models and homogeneous models. The heterogeneous models are built based on different types of basic models, while the homogeneous models are built based on the same types of basic models (Hao et al., 2018; Wang et al., 2017). The steps of the utilization of hybrid machine learning are similar to the steps of the

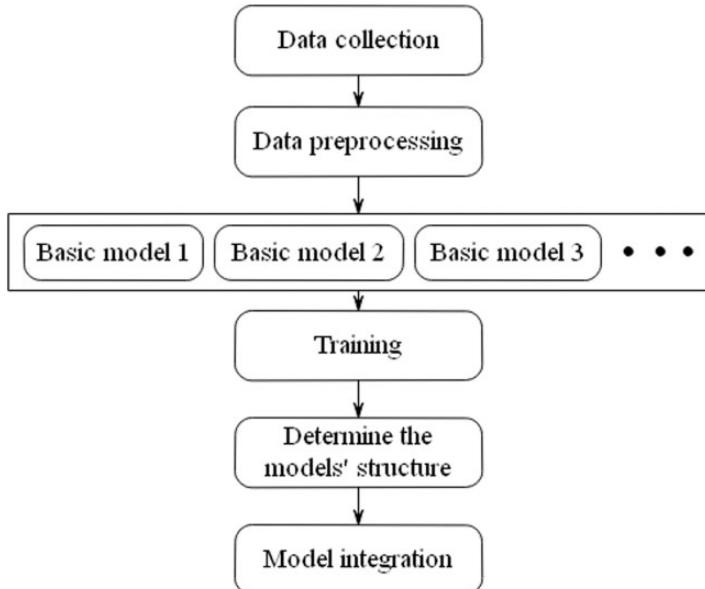


Figure 4. Schematic structure of hybrid machine learning method.

single model (Figure 4). The difference is mainly in the training process for the determination of models' structure, which can refer to (Wang et al., 2017b).

Applications

In order to develop the SVM model, parallel SVM and least squares SVM (LS-SVM; Zhao and Magoulès, 2010) were proposed and adopted in the prediction for building energy consumption. Xuemei et al. (2010) conjunctive use SVM and cluster algorithm to predict the cooling load of HVAC system based on the historical measuring data. Through training and testing, the forecasting model was adopted. On the basis of the MAPE and RMSE, fuzzy C-mean clustering algorithm (FCM) coupled with fuzzy support vector machine (FSVM) has the best performances than the single SVM and FCM-SVM.

Robinson et al. (2017) took advantage of machine learning method based on gradient boosting regression models to predict commercial building energy-consumption. Although this method was only successfully performed for building group in entire metropolitan areas, some limitations are existing through the experiments and verification. However, it still indicated the universality of machine learning method. Mateo et al. (2013) verified that several machine learning methods and classical technologies could predict the short-term indoor temperature of buildings. They also indicated that the combination of multilayer perceptron (MLP) with non-linear autoregressive techniques is the best predictor by comparison.

Furthermore, Fan et al. (2014) implemented the prediction for the energy consumption of next day and the peak power demand based on eight typical models. They are respectively multiple linear regression (MLR), support vector regression (SVR), boosting tree (BT), ARIMA, random forests (RF), MLP, multivariate adaptive regression splines (MARS),

and k-nearest neighbors (kNN). Further, they proposed the ensemble model that combines the eight base models, and GA was utilized to optimize the weights of each models. In the practice, based on the ensemble model, the MAPE of energy consumption of next day and the peak power demand reach 2.32% and 2.85%, respectively. Chou and Bui (2014) predicted the cooling load of one residential building by ensemble model (combine the SVR and ANN), which indicates the feasibility of the combined data-mining techniques for forecasting cooling consumption of building. It can be concluded that the hybrid machine learning method could obtain the energy consumption predictive values, which is more accurate than that obtained from individual base models. But it does not say that as many models as possible can improve the accuracy of the predicted results. According to Zhou et al.'s (2002) research, the effective ensemble model is built based on limited number of appropriate models.

ANN model based on HTS

Li et al. (2017b, 2018b) and Liu et al. (2016b, 2017f) have done most of the primary works on the machine learning based high-throughput screening (HTS) for prediction and optimization of solar energy systems. The successful exploration illustrates that hybrid machine method can also optimize the energy sectors, which extends the scope of application of machine learning. HTS could screen huge amounts of combinations to obtain the candidates that satisfy the required properties, through high-throughput experimental together with computational techniques. It can promptly single out the promising candidates among millions of possible samples, which dramatically accelerated the progress for science (Liu et al., 2018b; Wang et al., 2011; Wójcik et al., 2015). In addition, HTS development is in the mature period, which is normally employed for the screening of the material with the best properties. Due to the parallel and automatic disposition process, HTS has become one valuable selection technology (Sundermann and Gerlach, 2016).

To integrate the advantages of above two technologies, machine learning method based on HTS was proposed by Li and Liu et al. Flow chart for this novel method is presented in Figure 5. In their research, heat collection rate (HCR) was served as the target, while the input variables include eight parameters (e.g., tube length, tube number, tube center distance, tank volume, collector area, angle between tubes and ground) that could be obtained easily. Two types of structure with higher performance than previous one were singled out, and their performances were verified through the verification in practice. It indicates the feasibility of the novel integrated method. In summary, in view of the complicated correlations between HCR and extrinsic properties, ANN model based on HTS could be considered as the efficient approach to predict and optimize the performance of solar collector.

Predictive accuracy comparison

There are numerous ANN model like multilayer feed-forward neural network (MLFN), general regression neural network (GRNN), recurrent network, and self-organizing maps (Haykin, 1994). All of these networks have their own learning algorithm. There are many algorithms such as gradient descent with momentum, Broyden–Fletcher–Goldfarb–Shanno quasi-Newton, Levenberg–Marquardt, and conjugate gradient. For SVM, there are four basic kernel functions as mentioned above including linear function, polynomial function, RBF, and sigmoid function. Besides, different combination of penalty parameter and kernel

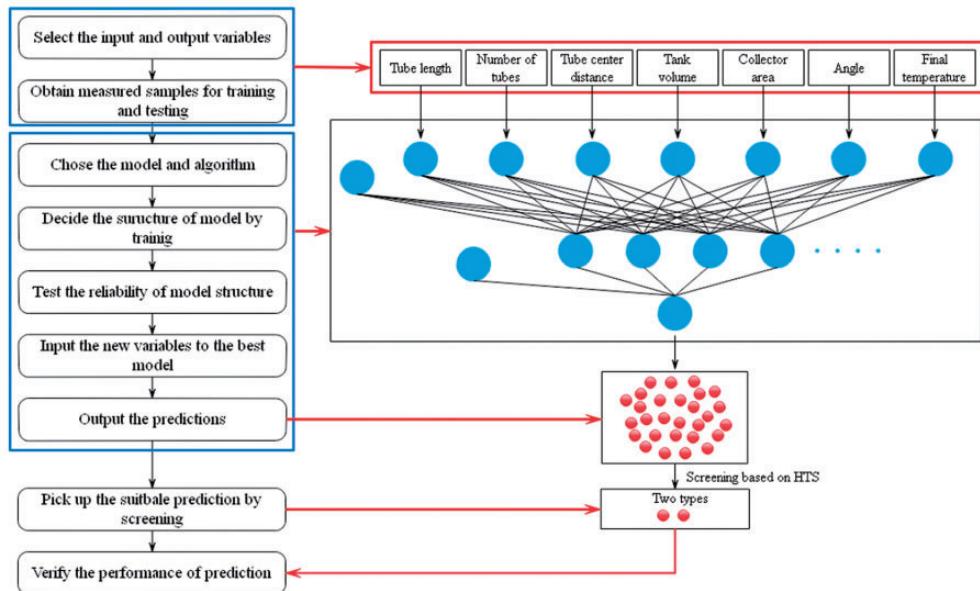


Figure 5. Flow chart for prediction process for solar collector by using ANN model based on HTS (Li et al., 2017).

parameters also generate different results, which has significant influence on the predictive accuracy. Therefore, the training for certain model is particularly necessary.

The training of ANN model and SVM model are in favor of acquirement for the suitable structure of model, which could avoid the risk of over- or under-fitting. Many factors, e.g. historical data for training, input variables, and different models, have significant influences on the algorithmic structure. The aim for training is to determine the ANN structure including input variables, numbers of hidden nodes, hidden layer(s), and weight values. Therefore, the training is the important process to optimize the model structure and make the relative parameters meet the studied case.

Many investigators were dedicated to explore the accurate models that could reach the best predictive values. Fu et al. (2015) predicted the hourly electricity load through different SVM model based on weather conditions and two previous days' hourly electricity loads. Further, the results from the SVM model were compared with the results obtained from the ARIMAX, decision tree, and ANN model, which showed the higher predictive accuracy of SVM. Li et al. (2010) predict the electricity consumption of residential buildings based on the SVM, BPNN, RBFNN, GRNN, respectively, showing that SVM model was more applicable to the problem than other models. Azadeh et al. (2008) performed the prediction of the annual electricity consumption, which presents the applicability of the ANN model with supervised MLP in this field. Through the analysis of variance, it was verified that ANN model outperforms than conventional regression significantly.

Furthermore, Neto and Fiorelli (2008) used climate variables and ANN model to predict the total daily energy consumption of building. They compared the predictive results with the results simulated from Energy-Plus, which indicates that the ANN model is more accurate than Energy-Plus. Based on SVM model, Dong et al. (2005) predicted the monthly

Table 2. Comparative information about ANN, SVM, and hybrid model.

Models	ANN	SVM	Hybrid model
Accuracy	Average	Good	Best
Model complexity	Medium	High	Fairly high
Computation speed	Medium	Medium	Low
Computation difficulty	Medium	Medium	Difficult
Energy sampling type	Long-term; short-term	Long-term; short-term	Long-term
Determination of parameters/models	Average	Simple	Difficult

ANN: artificial neural network; SVM: support vector machine.

energy consumption of one commercial building located in tropical region. In terms of the mean square error (MSE) and coefficient of variance (CV), it verifies that the performances of SVM model are better than other neural network models in prediction. In the Li's study, it was indicated that the performance of SVM model is better than the conventional neural networks in the prediction of the cooling load of the office building (Gao et al., 2012; Li et al., 2009; Zhang et al., 2017a). It can be obviously seen that ANN model and SVM model have been the most popular method to predict the energy building regarding the heating, cooling, or overall load. From the given above researches, it can be concluded that which model should be recommended with higher predictive accuracy depends on the studied objection and the input parameters collected.

For the single model prediction, there are three advantages: (1) easy to determine model structure, (2) easy to implement, and (3) short computation time. The disadvantage is the limitation of accuracy and reliability. On the contrary, hybrid machine learning method improves the predictive accuracy and stability remarkably, because of the utilization of multiple basic models. The irrelevance of these basic models could reduce the overall prediction error of the hybrid model. Besides, the hybrid predictive method needs more knowledge to determine the model structure and more calculation time. The performances of the hybrid prediction method are depending on each basic model selected. However, there is no reference about the selection for the basic models. That is usually determined through own experience. The detailed comparative information about ANN, SVM, and hybrid model is listed in Table 2.

Discussion

The example about solar collector describes the successful application of building energy system design through machine learning method based on HTS with MLFN to predict the better HCR. Simultaneously, Li and Liu et al. also seek for the better solar collector structure with lower heat loss coefficient HCL (Liu et al., 2018; Zeng and Gao, 2017). They indicated that GRNN is the acceptable model to predict the HLC value. Therefore, the selection for model is various, which depends on the object studied. Whether the selection of model is suitable should be identified through the training process.

Machine learning method has strong ties to mathematical optimization. Simultaneously, it delivers methods, models, and applications to many fields. The optimization of solar collector based on the machine learning and HTS illustrates that hybrid machine learning

method could provide the accurate prediction, and dramatically shorten the test period and saves the manpower. In addition, machine learn method based on HTS can be also utilized in taxonomy of plant (Singh et al., 2016), drug discovery (Scheeder et al., 2018; Simm et al., 2018), and new alloys discovery (Gubaev et al., 2019). Obviously, the search speed increases by orders of magnitude based on this hybrid method, which greatly reduces the waste of resources. Nevertheless, the novel method only indicates the availability in experiment (Li et al., 2018d; Peng et al., 2008; Liu et al., 2017g). Therefore, the suitability of that in other field remains to be verified so far, furthermore, its feasibility remains to be developed.

In order to improve the predictive efficiency, some simplified model and grey model were usually utilized. For example, in the research done by Wang et al., they proposed the simplified building energy model with desirable prediction of dynamic thermal performance of building. That was achieved through simplifying the model of the building envelopes and the building internal mass (Wang and Xu, 2006). If there were no completed data groups, the energy consumption prediction could be achieved through grey models, while the research about grey models is relatively less. Based on the improved grey system, Guo et al. predict the energy consumption of heat pump water heaters in residential buildings. Through the analyses of the measured and predicted values, they found that the accuracy of the model increases with the increase of the data sample interval. Further, the time of four weeks is the best data sample interval in the prediction. The distinct characteristic of the method is grey model could deal with the sample with unequal gap sequence and provide the accurate prediction (Guo et al., 2011). Zhou et al. (2008) also predict the hourly building load of the next day through the two weather prediction modules including the air temperature, relative humidity, and solar radiation based on the grey-box model. In addition, extreme learning machine (ELM) belongs to MLFNs, which was usually used for classification, regression or feature learning, etc. Unlike the parameters setting of other machine learn methods, the parameters of hidden nodes in ELM have no necessity of optimizing. ELM can be also employed to predict energy performance of building, which was based on the selected highly correlated parameters and efficient model to achieve (Kumar et al., 2018).

Perspective

Although there were many promising predictive results for different cases, there are some obstacles that are different to determine, such as the type and number of inputs, the scale of historical data for training or testing, and the number of models in hybrid method. Therefore, a specific guidance should be edited or summarized to assist in solving the problems, even if it is based on the experience. The guidance should have the ability to provide some suggestions to improve the predictive accuracy, such as which building was recommended to be monitored, and in which measurement locations to ensure data representativeness. These suggestions could significantly eliminate duplication of efforts.

In addition, three aspects about the energy consumption prediction should be concerned. Firstly, the occupancy factors (including the occupancy behaviors, and the number of occupants) could obviously affect the energy consumption. While the researches about the effect of occupants factors on the energy consumption prediction results are very few. Secondly, there were fewer references about the prediction with long-term period that has important significance on energy management. Thirdly, the researches about the energy consumption prediction of the sub-components were less. Thus, the exploration of energy consumption prediction mentioned above should be implemented.

Further, the more robust and effective model or method for the energy consumption should be developed. For example, based on ELM, deep learning method or decision tree method, optimize their structure to suit the prediction of building energy consumption. The influence of input variables on the prediction should be evaluated, which could balance the computation difficulty and the accuracy of results in practice. Furthermore, with the development of smart meters and automatic acquisition system, more representative monitoring data with large sizes will become available, which has the benefit to the build model. Moreover, the development of the smart energy management and building energy efficiency retrofitting will be more advanced and have more clear direction (Mathew et al., 2015; Zhou and Yang, 2016). It is also expected that machine learning could be applied to more other energy conversion processes (Langley, 2011) and some challenging topics in optimization and analysis related topics (Chen et al., 2016; Dai et al., 2014, 2015; Hu et al., 2012, 2016, 2017; Liu et al., 2016a, 2017a, 2017b; Peng et al., 2016a, 2016b, 2016c; Peng and Zhang, 2017; Shen et al., 2017; Wang et al., 2019; Yuan et al., 2011; Zhang et al., 2017b, 2017c, 2018b; Zhou et al., 2014).

Conclusion

An overview of machine learning method adopted in the field of building energy consumption prediction was presented in this review. Some most typical literatures reported in recent years were introduced, summarized, and commented. The accuracy analyses were conducted based on different type of building, different type of temporal granularity, as well as input/output variables and historical data collected. Further, the strengths and weaknesses of several models structure were compared, especially for ANN and SVM. Several main conclusions can be summarized as follows:

1. Compared to other statistical or physical methods, machine learning could acquire the prediction results with higher reliability, which can significantly save manpower and reduce time-consumption. However, the predictive structure with best performance is difficult to be determined for different cases.
2. For residential buildings, there is no regular schedule of activities for different residents or different periods. Therefore, it is difficult to predict for them, which is related to a lot of individual factors. Moreover, the long-term predictive results are vulnerable, which is easily affected by the uncertainties, thus, which is more difficult to predict than that on short-term.
3. Through the comparison between single (ANN and SVM) and hybrid machine learning methods, it can be seen that different models have different features that are applicable to different situations. They all have their respective strengths and weaknesses. Therefore, for different cases, suitable models needed to be selected on the basis of their individual characteristics and training processes.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The National Key R&D Program of China-Technical System and Key Technology Development of Nearly Zero Energy Building (Grant No. 2017YFC0702600), the National Science Foundation of China (Grant No. 51708211), the Opening Funds of State Key Laboratory of Building Safety and Built Environment National Engineering Research Center of Building Technology (Grant No. BSBE2017-08), Natural Science Foundation of Hebei (Grant No. E2017502051), and the Fundamental Research Funds for the Central Universities (Grant No. 2018MS103, 2018MS108 and 2017MS119).

ORCID iD

Zhijian Liu  <http://orcid.org/0000-0003-1841-2671>

References

- Abdmouleh Z, Gastli A, Ben-Brahim L, et al. (2017) Review of optimization techniques applied for the integration of distributed generation from renewable energy sources. *Renewable Energy* 113: 266–280.
- Al-Saadi SNJ, Ramaswamy M, Al-Rashdi H, et al. (2017) Energy management strategies for a governmental building in Oman. *Energy Procedia* 141: 206–210.
- Ansari FA, Mokhtar AS, Abbas KA, et al. (2005) A simple approach for building cooling load estimation. *American Journal of Environmental Sciences* 1(3): 209–212.
- Aras N (2008) Forecasting residential consumption of natural gas using genetic algorithms. *Energy Exploration & Exploitation* 26(4): 241–266.
- Aydinalp M, Ugursal VI and Fung AS (2002) Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. *Applied Energy* 71(2): 87–110.
- Aydinalp-Koksal M and Ugursal VI (2005) Comparison of neural network, conditional demand analysis and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy* 85: 271–296.
- Aydin L (2018) Effects of increasing indigenous coal share in Turkey's electricity generation mix on key economic and environmental indicators: An extended input–output analysis. *Energy Exploration & Exploitation* 36(2): 230–245.
- Azadeh A, Ghaderi SF and Sohrabkhani S (2008) Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management* 49(8): 2272–2278.
- Azqueta-Gavaldón A (2017) Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters* 158: 47–50.
- Barzegar R, Moghaddam AA, Deo R, et al. (2017) Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *The Science of the Total Environment* 621: 697–712.
- Bauer M and Scartezzini JL (1998) A simplified correlation method accounting for heating and cooling loads in energy-efficient buildings. *Energy and Buildings* 27(2): 147–154.
- Calero M, Alameda-Hernandez E, Fernández-Serrano M, et al. (2018) Energy consumption reduction proposals for thermal systems in residential buildings. *Energy and Buildings* 175: 121–130.
- Cao LJ and Tay FEH (2003) Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks* 14: 1506–1518.
- Castiglioni I, Salvatore C and Ramirez J (2018) Machine-learning neuroimaging challenge for automated diagnosis of mild cognitive impairment: Lessons learnt. *Journal of Neuroscience Methods* 302: 10.

- Chalouhi EK, Gonzalez I, Gentile C, et al. (2017) Damage detection in railway bridges using machine learning: Application to a historic structure. *Procedia Engineering* 199: 1931–1936.
- Chen W, Panahi M and Pourghasemi HR (2017a) Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *Catena* 157: 310–324.
- Chen W, Pourghasemi HR, Kornejady A, et al. (2017b) Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* 305: 314–327.
- Chen Y, Wang Z and Lu L (2016) Remix : On-demand live randomization. In: *Proceedings of the sixth ACM conference on data and application security and privacy*, 9 March, New Orleans, Louisiana, USA, 2016, pp.50–61.
- Chou JS and Bui DK (2014) Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings* 82: 437–446.
- Cramer S, Kampouridis M, Freitas AA, et al. (2017) An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications* 85: 169–181.
- Cuce E, Harjunowibowo D and Cuce PM (2016) Renewable and sustainable energy saving strategies for greenhouse systems: A comprehensive review. *Renewable and Sustainable Energy Reviews* 64: 34–59.
- Dagnely P, Ruette T, Tourwé T, et al. (2015) Predicting hourly energy consumption. Can you beat an autoregressive model. In: *Proceeding of the 24th annual machine learning conference of Belgium and the Netherlands*, Benelearn, Delft, The Netherlands, 2015, p.19.
- Dai Z, Peng Y, Mansy HA, et al. (2014) Comparison of poroviscoelastic models for sound and vibration in the lungs. *Journal of Vibration and Acoustics* 136(5): 051012.
- Dai Z, Peng Y, Mansy HA, et al. (2015) A model of lung parenchyma stress relaxation using fractional viscoelasticity. *Medical Engineering & Physics* 37(8): 752–758.
- Dhar A, Reddy TA and Claridge DC (1999) A Fourier series model to predict hourly heating and cooling energy use in commercial buildings with outdoor temperature as the only weather variable. *Journal of Solar Energy Engineering* 121: 47–53.
- Dombayci ÖA (2010) The prediction of heating energy consumption in a model house by using artificial neural networks in Denizli–Turkey. *Advances in Engineering Software* 41(2): 141–147.
- Dong B, Cao C and Lee SE (2005) Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings* 37(5): 545–553.
- Du B, Gao J, Zeng L, et al. (2017) Area optimization of solar collectors for adsorption desalination. *Solar Energy* 157: 298–308.
- Embaye M, Al-Dadah RK and Mahmoud S (2015) Thermal performance of hydronic radiator with flow pulsation – Numerical investigation. *Applied Thermal Engineering* 80: 109–117.
- Fan C, Xiao F and Wang S (2014) Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy* 127: 1–10.
- Forni A and Meulen R (2016) Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage. Available at: <http://www.gartner.com/newsroom/id/3412017> (accessed 13 July 2018).
- Fu Y, Li Z, Zhang H, et al. (2015) Using support vector machine to predict next day electricity load of public buildings with sub-metering devices. *Procedia Engineering* 121: 1016–1022.
- Gao P, Lü X, Zhang J, et al. (2012) Distribution characteristics and significance of hydrocarbon shows to petroleum geology in Kalpin area, NW Tarim Basin. *Energy Exploration & Exploitation* 30(1): 89–108.
- Garshasbi S, Kurnitski J and Mohammadi Y (2016) A hybrid genetic algorithm and Monte Carlo simulation approach to predict hourly energy consumption and generation by a cluster of net zero energy buildings. *Applied Energy* 179: 626–637.

- Georgescu M, Eccles E, Manjunath V, et al. (2014) Machine learning methods for site-level building energy forecasting and data rectification. *Building Simulation and Optimization – The Second IBPSA – England Conference*, London, UK.
- Goh KL and Singh AK (2015) Comprehensive literature review on machine learning structures for web spam classification. *Procedia Computer Science* 70: 434–441.
- González-Recio O, Rosa GJM and Gianola D (2014) Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livestock Science* 166(1): 217–231.
- Gubaev K, Podryabinkin EV, Hart GLW, et al. (2019) Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Computational Materials Science* 156: 148–156.
- Guo JJ, Wu JY and Wang RZ (2011) A new approach to energy consumption prediction of domestic heat pump water heater based on grey system theory. *Energy and Buildings* 43(6): 1273–1279.
- Hao X, An X, Wu B, et al. (2018) Application of a support vector machine algorithm to the safety precaution technique of medium-low pressure gas regulators. *Journal of Thermal Science* 27(1): 74–77.
- Hautier G, Fischer CC, Jain A, et al. (2010) Finding natures missing ternary oxide compounds using machine learning and density functional theory. *Chemistry of Materials* 22(12): 3762–3767.
- Haykin S (1994) *Neural Networks: A Comprehensive Foundation*. 1st ed. Upper Saddle River, NJ: Prentice Hall PTR.
- Heracleous C and Michael A (2018) Assessment of overheating risk and the impact of natural ventilation in educational buildings of Southern Europe under current and future climatic conditions. *Energy* 165: 1228–1239.
- Hou Z and Lian Z (2009) An application of support vector machines in cooling load prediction. Intelligent systems and applications. ISA 2009. *International Workshop on IEEE* (2009): 1–4.
- Hu W, Choi KK and Cho H (2016) Reliability-based design optimization of wind turbine blades for fatigue life under dynamic wind load uncertainty. *Structural and Multidisciplinary Optimization* 54(4): 953–970.
- Hu W, Han I, Park SC, et al. (2012) Multi-objective structural optimization of a HAWT composite blade based on ultimate limit state analysis. *Journal of Mechanical Science and Technology* 26(1): 129–135.
- Hu W, Pryor SC, Letson F, et al. (2017) Investigation of gust-seismic relationships and applications to gust detection. *Journal of Geophysical Research* 122(1): 140–151.
- Hu Z, Lu W, Thouless MD, et al. (2015a) Effect of plastic deformation on the evolution of wear and local stress fields in fretting. *International Journal of Solids and Structures* 82: 1–8.
- Hu Z, Lu W, Thouless MD, et al. (2015b) Simulation of wear evolution using fictitious eigenstrains. *Tribology International* 82: 191–194.
- Huebner GM, Hamilton I, Chalabi Z, et al. (2015) Explaining domestic energy consumption – The comparative contribution of building factors, socio-demographics, behaviours and attitudes. *Applied Energy* 159: 589–600.
- Huo T, Ren H and Cai W (2019) Estimating urban residential building-related energy consumption and energy intensity in China based on improved building stock turnover model. *The Science of the Total Environment* 650: 427–437.
- Jia X, Jin C, Buzzia M, et al. (2016) Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves. *Renewable Energy* 99: 1191–1201.
- Jia X, Jin C, Buzzia M, et al. (2018) A deviation based assessment methodology for multiple machine health patterns classification and fault detection. *Mechanical Systems and Signal Processing* 99: 244–261.
- Jia X, Zhao M, Di Y, et al. (2017) Investigation on the kurtosis filter and the derivation of convolutional sparse filter for impulsive signature enhancement. *Journal of Sound and Vibration* 386: 433–448.
- Jiang A and O'Meara A (2018) Accommodating thermal features of commercial building systems to mitigate energy consumption in Florida due to global climate change. *Energy and Buildings* 179: 86–98.

- Jun J, Kang J, Jeong D, et al. (2017) An efficient approach for optimizing full field development plan using Monte-Carlo simulation coupled with genetic algorithm and new variable setting method for well placement applied to gas condensate field in Vietnam. *Energy Exploration & Exploitation* 35(1): 75–102.
- Kalogirou S, Eftekhari M and Marjanovic-Halburd L (2001) Estimation of the daily heating and cooling loads using artificial neural networks. In: *Proceedings of CLIMA 2000 International Conference*, Naples, Italy, 2001. Available at: <http://ktisis.cut.ac.cy/handle/10488/883> (accessed 13 July 2018).
- Kalogirou SA (2000) Applications of artificial neural-networks for energy systems. *Applied Energy* 67: 17–35.
- Karim MA, Currie J and Lie TT (2018) A machine learning based optimized energy dispatching scheme for restoring a hybrid microgrid. *Electric Power Systems Research* 155: 206–215.
- Khosravani H, Castilla M, Berenguel M, et al. (2016) A comparison of energy consumption prediction models based on neural networks of a bioclimatic building. *Energies* 9(1): 57.
- Kim EY, Lee MY, Kim SH, et al. (2017) Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 76: 65–71.
- Kumar S, Pal SK and Singh RP (2018) Intra ELM variants ensemble based model to predict energy performance in residential buildings. *Sustainable Energy, Grids and Networks* 16: 177–187.
- Lai F, Magoules F and Lherminier F (2008) Vapnik's learning theory applied to energy consumption forecasts in residential buildings. *International Journal of Computer Mathematics* 85(10): 1563–1588.
- Langley P (2011) The changing science of machine learning. *Machine Learning* 82(3): 275–279.
- Lázaro JL, Jiménez ÁB and Takeda A (2017) Improving cash logistics in bank branches by coupling machine learning and robust optimization. *Expert Systems with Applications* 92: 236–255.
- Lee C, Kwon O, Kim M, et al. (2017) Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change* 127: 291–303.
- Li G, Xuan Q, Pei G, et al. (2018a) Effect of non-uniform illumination and temperature distribution on concentrating solar cell-A review. *Energy* 144: 1119–1136.
- Li G, Zhou K, Song Z, et al. (2018b) Inconsistent phenomenon of thermoelectric load resistance for photovoltaic–thermoelectric module. *Energy Conversion and Management* 161: 155–161.
- Li H, Chen FL and Cheng K (2015) Prediction of zeta potential of decomposed peat via machine learning: Comparative study of support vector machine and artificial neural networks. *International Journal of Electrochemical Science* 10(8): 6044–6056.
- Li H and Henkelman G (2017a) Dehydrogenation selectivity of ethanol on close-packed transition metal surfaces: A computational study of monometallic, Pd/Au, and Rh/Au catalysts. *The Journal of Physical Chemistry C* 121(49): 27504–27510.
- Li H and Liu Z (2018c) Performance prediction and optimization of solar water heater via a knowledge-based machine learning method. In: *Handbook of Research on Power and Energy System Optimization*. Pennsylvania, USA: GI Global, pp.55–74.
- Li H, Liu Z, Liu K, et al. (2017b) Predictive power of machine learning for optimizing solar water heater performance: The potential application of high-throughput screening. *International Journal of Photoenergy* 2017: 1–10.
- Li H, Luo L, Kunal P, et al. (2018d) Oxygen reduction reaction on classically immiscible bimetallics: A case study of RhAu. *The Journal of Physical Chemistry C* 122(5): 2712–2716.
- Li H, Tang X, Wang R, et al. (2016) Comparative study on theoretical and machine learning methods for acquiring compressed liquid densities of 1,1,1,2,3,3-heptafluoropropane (R227ea) via Song and Mason equation, support vector machine, and artificial neural networks. *Applied Sciences* 6(1): 25.

- Li H, Zhang Z and Liu Z (2017c) Application of artificial neural networks for catalysis: A review. *Catalysts* 7(10): 306.
- Li P, Han Z, Jia X, et al. (2019) Analysis and comparison on thermodynamic and economic performances of an organic Rankine cycle with constant and one-dimensional dynamic turbine efficiency. *Energy Conversion and Management* 180: 665–679.
- Li Q, Meng Q, Cai J, et al. (2009) Applying support vector machine to predict hourly cooling load in the building. *Applied Energy* 86(10): 2249–2256.
- Li Q, Ren P and Meng Q (2010) Prediction model of annual energy consumption of residential buildings. In: *International conference on advances in energy engineering*. Beijing, China, 19–20 June 2010, pp.223–226.
- Liu H, Yue S, Wang Y, et al. (2018a) Unsteady study on the effects of matching characteristic of tandem cascade on the performance and flow at large angle of attack. *Journal of Thermal Science* 27(6): 505–515.
- Liu T, Jagota A and Hui CY (2016a) Effect of surface tension on the adhesion between a rigid flat punch and a semi-infinite neo-Hookean half-space. *Extreme Mechanics Letters* 9: 310–316.
- Liu T, Jagota A and Hui CY (2017a) A closed form large deformation solution of plate bending with surface effects. *Soft Matter* 13(2): 386–393.
- Liu T, Nadermann N, He Z, et al. (2017b) Spontaneous droplet motion on a periodically compliant substrate. *Langmuir* 33(20): 4942–4947.
- Liu X, Zhou Y and Chen X (2018) Mining outlier data in mobile Internet-based large real-time databases. *Complexity* 2018(4): 1–12.
- Liu Y, Li H, Cen W, et al. (2018b) A computational study of supported Cu-based bimetallic nanoclusters for CO oxidation. *Physical Chemistry Chemical Physics* 20(11): 7508–7513.
- Liu Z, Cheng K, Li H, et al. (2017d) Exploring the potential relationship between indoor air quality and the concentration of airborne culturable fungi: A combined experimental and neural network modeling study. *Environmental Science and Pollution Research* 25(4): 3510–3517.
- Liu Z, Li H and Cao G (2017e) Quick estimation model for the concentration of indoor airborne culturable bacteria: An application of machine learning. *International Journal of Environmental Research and Public Health* 14(8): 857.
- Liu Z, Li H, Liu K, et al. (2017f) Design of high-performance water-in-glass evacuated tube solar water heaters by a high-throughput screening based on machine learning: A combined modeling and experimental study. *Solar Energy* 142: 61–67.
- Liu Z, Li H, Tang X, et al. (2016b) Extreme learning machine: a new alternative for measuring heat collection rate and heat loss coefficient of water-in-glass evacuated tube solar water heaters. *SpringerPlus* 5(1): 626.
- Liu Z, Wu D, Li J, et al. (2018c) Optimizing building envelope dimensions for passive solar houses in the Qinghai-Tibetan region: Window to wall ratio and depth of sunspace. *Journal of Thermal Science*, pp.1–14.
- Liu Z, Xu W, Zhai X, et al. (2017g) Feasibility and performance study of the hybrid ground-source heat pump system for one office building in Chinese heating dominated areas. *Renewable Energy* 101: 1131–1140.
- Lomet A, Suard F and Chèz D (2015) Statistical modeling for real domestic hot Water consumption forecasting. *Energy Procedia* 70: 379–387.
- Lü X, Lu T, Kibert CJ, et al. (2015) Modeling and forecasting energy consumption for heterogeneous buildings using a physical-statistical approach. *Applied Energy* 144: 261–275.
- Magalhães SMC, Leal VMS and Horta IM (2017) Modelling the relationship between heating energy use and indoor temperatures in residential buildings through artificial neural networks considering occupant behavior. *Energy and Buildings* 151: 332–343.
- Marasco DE and Kontokosta CE (2016) Applications of machine learning methods to identifying and predicting building retrofit opportunities. *Energy and Buildings* 128: 431–441.

- Mateo F, Carrasco JJ, Sellami A, et al. (2013) Machine learning methods to forecast temperature in buildings. *Expert Systems with Applications* 40(4): 1061–1068.
- Mathew PA, Dunn LN, Sohn MD, et al. (2015) Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *Applied Energy* 140: 85–93.
- Medler D (1998) A brief history of connectionism. *Neural Computing Surveys* 1(2): 18–72.
- Mihalakakou G, Santamouris M and Tsangrassoulis A (2002) On the energy consumption in residential buildings. *Energy and Buildings* 34(7): 727–736.
- Moss GP, Shah AJ, Adams RG, et al. (2012) The application of discriminant analysis and machine learning methods as tools to identify and classify compounds with potential as transdermal enhancers. *European Journal of Pharmaceutical Sciences* 45(1): 116–127.
- Neto AH and Fiorelli FAS (2008) Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings* 40(12): 2169–2176.
- Nanculef R, Frandi E, Sartori C, et al. (2014) A novel Frank-Wolfe algorithm. Analysis and applications to large-scale SVM training. *Information Sciences* 285(C): 66–99.
- Papadopoulos G, Edwards PJ and Murray AF (2000) Confidence estimation methods for neural networks: A practical comparison. *IEEE Transactions on Neural Networks* 12(6): 1278–1287.
- Paudel S, Elmtiri M, Kling WL, et al. (2014) Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. *Energy and Buildings* 70: 81–93.
- Peng H and Ling X (2008) Optimal design approach for the plate-fin heat exchangers using neural networks cooperated with genetic algorithms. *Applied Thermal Engineering* 28(5–6): 642–650.
- Peng Y, Dai Z, Mansy HA, et al. (2016a) Sound transmission in porcine thorax through airway insonification. *Medical and Biological Engineering and Computing* 54(4): 675–689.
- Peng Y, He J, Khavari R, et al. (2016b) Functional mapping of the pelvic floor and sphincter muscles from high-density surface EMG recordings. *International Urogynecology Journal* 27(11): 1689–1696.
- Peng Y, He J, Yao B, et al. (2016c) Motor unit number estimation based on high-density surface electromyography decomposition. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 127(9): 3059–3065.
- Peng Y and Zhang Y (2017) Improving the repeatability of Motor Unit Number Index (MUNIX) by introducing additional epochs at low contraction levels. *Clinical Neurophysiology* 128(7): 1158–1165.
- Pérez-Lombard L, Ortiz J and Pout C (2008) A review on buildings energy consumption information. *Energy and Buildings* 40(3): 394–398.
- Pham BT, Pradhan B, Bui DT, et al. (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software* 84: 240–250.
- Razavi SH, Ahmadi R and Zahedi A (2018) Modeling, simulation and dynamic control of solar assisted ground source heat pump to provide heating load and DHW. *Applied Thermal Engineering* 129: 127–144.
- Reynolds J, Ahmad MW, Rezgui Y, et al. (2019) Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm. *Applied Energy* 235: 699–713.
- Reynolds J, Rezgui Y, Kwan A, et al. (2018) A zone-level, building energy optimisation combining an artificial neural network, a genetic algorithm, and model predictive control. *Energy* 151: 729–739.
- Robinson C, Dilksina B, Hubbs J, et al. (2017) Machine learning approaches for estimating commercial building energy consumption. *Applied Energy* 208: 889–904.
- Samant P and Agarwal R (2018) Machine learning techniques for medical diagnosis of diabetes using iris images. *Computer Methods and Programs in Biomedicine* 157: 121–128.
- Samuel AL (1967) Some studies in machine learning using the game of checkers. II – Recent progress. *IBM Journal of Research and Development* 11(6): 601–617.

- Sattlecker M, Stone N and Bessant C (2014) Current trends in machine-learning methods applied to spectroscopic cancer diagnosis. *TrAC – Trends in Analytical Chemistry* 59: 17–25.
- Scheeder C, Heigwer F and Boutros M (2018) Machine learning and image-based profiling in drug discovery. *Current Opinion in Systems Biology* 10: 43–52.
- Shen W, Li D, Zhang S, et al. (2017) Analysis of wave motion in one-dimensional structures through fast-Fourier-transform-based wavelet finite element method. *Journal of Sound and Vibration* 400: 369–386.
- Shirzadi A, Shahabi H, Chapi K, et al. (2017) A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena* 157: 213–226.
- Simm J, Klambauer G, Arany A, et al. (2018) Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chemical Biology* 25: 611–618.
- Singh A, Ganapathysubramanian B, Singh AK, et al. (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science* 21: 110–124.
- Sundermann A and Gerlach O (2016) High-throughput screening as a supplemental tool for the development of advanced emission control catalysts: Methodological approaches and data processing. *Catalysts* 6(2): 23.
- Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology* 49(11): 1225–1231.
- Voyant C, Notton G, Kalogirou S, et al. (2017) Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* 105: 569–582.
- Walker SF (1990) A brief history of connectionism and its psychological implications. *AI & Society* 4(1): 17–38.
- Wang P, Yang R, Liu W, et al. (2011) Reservoir characteristics and favorable area prediction of Devonian Kizirtag formation in Bashituo area, Tarim Basin. *Energy Exploration & Exploitation* 29(5): 575–596.
- Wang S and Xu X (2006) Simplified building model for transient thermal performance estimation using GA-based parameter identification. *International Journal of Thermal Sciences* 45(4): 419–432.
- Wang X, Qi Y, Wang Z, et al. (2019) Design and implementation of SecPod, a framework for virtualization-based security systems. *IEEE Transactions on Dependable and Secure Computing* 16 (1): 44–57. DOI: 10.1109/TDSC.2017.2675991
- Wang Z and Srinivasan RS (2017b) A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews* 75: 796–808.
- Wong KI, Wong PK, Cheung CS, et al. (2013) Modeling and optimization of biodiesel engine performance using advanced machine learning methods. *Energy* 55: 519–528.
- Wójcik M, Telzerow A, Quax WJ, et al. (2015) High-throughput screening in protein engineering: Recent advances and future perspectives. *International Journal of Molecular Sciences* 16(10): 24918–24945.
- Wu X, Kumar V, Quinlan JR, et al. (2008) Top 10 algorithms in data mining. *Knowledge and Information Systems* 14(1): 1–37.
- Xuemei L, Yuyan D, Lixing D, et al. (2010) Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering. In: *2010 international conference on computer and communication technologies in agriculture engineering*, 12–13 June 2010, Chengdu, China, Vol. 1, pp.438–441.
- Yaşa E and Ok V (2014) Evaluation of the effects of courtyard building shapes on solar heat gains and energy efficiency according to different climatic regions. *Energy and Buildings* 73: 192–199.
- Yuan S, Zhang H, Wang N, et al. (2011) Motion study of a redundant 7-DOF operation robot. In: *2011 international conference on electrical and control engineering*, ICECE 2011 – Proceedings, Yichang, China, 16–18 Sept 2011, pp.3056–3060.
- Zeng L and Gao J (2017) Contaminant source identification within an air duct system: Using a backward probabilistic model. *Procedia Engineering* 205: 3485–3492.

- Zeng L, Gao J, Du B, et al. (2018a) Probability-based inverse characterization of the instantaneous pollutant source within a ventilation system. *Building and Environment* 143: 378–389.
- Zeng Y, Liu J, Sun K, et al. (2018b) Machine learning based system performance prediction model for reactor control. *Annals of Nuclear Energy* 113: 270–278.
- Zhang Y, Burton HV, Sun H, et al. (2018a) A machine learning framework for assessing post-earthquake structural safety. *Structural Safety* 72: 1–16.
- Zhang Y, He CQ, Tang BJ, et al. (2015) China's energy consumption in the building sector: A life cycle approach. *Energy and Buildings* 94: 240–251.
- Zhang Y, Long E, Li Y, et al. (2017a) Solar radiation reflective coating material on building envelopes: Heat transfer analysis and cooling energy saving. *Energy Exploration & Exploitation* 35(6): 748–766.
- Zhang Z, Drapaca C, Zhang Z, et al. (2018b) Leakage evaluation by virtual entropy generation (VEG) method. *Entropy* 20(1): 14.
- Zhang Z, Li H, Chang H, et al. (2018c) Machine learning predictive framework for CO₂ thermodynamic properties in solution. *Journal of CO₂ Utilization* 26: 152–159.
- Zhang Z, Ou J, Li D, et al. (2017b) Optimization design of coupling beam metal damper in shear wall structures. *Applied Sciences* 7(2): 137.
- Zhang Z, Ou J, Li D, et al. (2017c) A thermography-based method for fatigue behavior evaluation of coupling beam damper. *Frattura Ed Integrità Strutturale* 11(40): 149–161.
- Zhao HX and Magoulès F (2010) Parallel support vector machines applied to the prediction of multiple buildings energy consumption. *Journal of Algorithms and Computational Technology* 4: 231–250.
- Zhou K and Yang S (2016) Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renewable and Sustainable Energy Reviews* 56: 810–819.
- Zhou Q, Wang S, Xu X, et al. (2008) A grey-box model of next-day building thermal load prediction for energy-efficient control. *International Journal of Energy Research* 32: 1418–1431.
- Zhou Y, Wang X, Chen Y, et al. (2014) ARMlock: Hardware-based fault isolation for ARM Yajin. In: *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 3–9 Nov 2014, New York, USA, Vol. 14, pp.558–569.
- Zhou ZH, Wu J and Tang W (2002) Ensembling neural networks: Many could be better than all. *Artificial Intelligence* 137: 239–263.