Energy and Buildings 82 (2014) 437–446

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Modeling heating and cooling loads by artificial intelligence for energy-efficient building design

Jui-Sheng Chou*, Dac-Khuong Bui

Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Rd., Taipei, 106, Taiwan

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form 15 July 2014
Accepted 16 July 2014
Available online 27 July 2014

Keywords:
Cooling load
Heating load
Energy performance
Energy-efficient building
Artificial intelligence
Data mining

ABSTRACT

The energy performance of buildings was estimated using various data mining techniques, including support vector regression (SVR), artificial neural network (ANN), classification and regression tree, chi-squared automatic interaction detector, general linear regression, and ensemble inference model. The prediction models were constructed using 768 experimental datasets from the literature with 8 input parameters and 2 output parameters (cooling load (CL) and heating load (HL)). Comparison results showed that the ensemble approach (SVR +ANN) and SVR were the best models for predicting CL and HL, respectively, with mean absolute percentage errors below 4%. Compared to previous works, the ensemble model and SVR model further obtained at least 39.0% to 65.9% lower root mean square errors, respectively, for CL and HL prediction. This study confirms the efficiency, effectiveness, and accuracy of the proposed approach when predicting CL and HL in building design stage. The analytical results support the feasibility of using the proposed techniques to facilitate early designs of energy conserving buildings.

1. Introduction

Since all three primary economic sectors, *i.e.*, industry, transportation, and building, have high energy use, conservation of energy is a critical task [1]. Buildings consume a substantial share of global energy consumption. Therefore, substantial energy savings can be realized by buildings that are properly designed and operated. Heating load (HL) and cooling load (CL) are measures of energy that must be added or removed from a space by heating ventilation and air conditioning (HVAC) system to provide the desired level of thermal comfort within space. Therefore, early predictions of building CL and HL can help engineers design energy-efficient buildings.

Building simulation tools have been widely used to predict and analyze building energy consumption. Building energy consumption and load simulation software developed in recent years include DOE-2 [2], ESP-r [3], Energy Plus [4], and DeST (Designer's Simulation Toolkit) [5]. Although they accurately predict building load in many projects [6–8], the prediction values differ according to the simulation software used to predict the energy use of occupied buildings [9]. Building simulation programs are complicated and time-consuming due to the diverse disciplines involved and the use of varying parameter settings. Current building simulation tools are also difficult to use for identifying and comparing the impacts of variables that affect the observed quantity of interest [10].

Because the building sector accounts for a large proportion of global energy consumption and has enduring adverse environmental impacts, studies of energy performance of buildings (EPB) have recently increased [11]. Rapid increases in building energy consumption in recent decades have also resulted from rising living standards. Notably, the building sector now consumes more than 30% of the total energy worldwide [12]. In Europe, the building sector accounts for 40% of energy use and 36% of CO_2 emissions [13] whereas, in the United States, the building sector accounts for 38.9% of the total primary energy requirement (PER) [14]. In China, building stocks accounted for about 24.1% of total national energy use in 1996 and for 27.5% in 2011. Building stocks are projected to increase to about 35% by 2020 [15].

Moreover, for a modern city like Hong Kong, 60% of carbon emissions are produced by electricity generation, and buildings account for 89% of total electricity consumption [16]. Hence, increasing the energy efficiency and energy performance of buildings is essential for mitigating the increasing demand for additional energy supply as well as CO_2 emission. While many researchers have developed methods of optimizing the operation of various components in HVAC and refrigerating systems [17,18], accurately predicting

* Corresponding author. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.
E-mail addresses: jschou@mail.ntust.edu.tw, rayson.chou@gmail.com (J.-S. Chou), M10205810@mail.ntust.edu.tw (D.-K. Bui).

http://dx.doi.org/10.1016/j.enbuild.2014.07.036
0378-7788/

building HL and CL is critical for effective energy conservation strategies.

In practice, identifying parameters that substantially affect building energy consumption can help optimize a building design. The influential parameters, *e.g.*, relative compactness [19], climate [20], surface area, wall area, and roof area [21], orientation [22], can be grouped into two main categories: the physical properties of a building and meteorological conditions. These factors make the relationship between EPB and its influential parameters very complicated. Consequently, accurately predicting HL and CL of building is a challenging task.

The artificial intelligence (AI) inference model has recently proven to be a viable alternative approach to predict EPB [23]. The AI is employed to develop models that simulate the human inference processes. Thus, AI can infer new facts from previously acquired information and can adaptively change in response to changes in historical data. Tsanas and Xifara [10] stated that AI not only obtains solutions very quickly, it also assists building designers in analyzing the influence of input parameters. Many studies have explored the use of AI models for predicting various interests in the context of EPB [9,24–30]. However, most works have reported unsatisfactory error rates, and most have considered only a few factors that affect building energy use.

Therefore, the objective of this research is to compare the performance of various AI techniques, including support vector regression (SVR), artificial neural network (ANN), classification and regression tree (CART), chi-squared automatic interaction detector (CHAID) and general linear regression (GLR). The best performing models were then combined into ensemble models. A k-fold cross-validation algorithm was used for validation. A synthesized performance index and hypothesis testing were used to compare performance measures between the proposed models and those in previous works. The contribution to the body of knowledge is the development of an AI technique that can predict building CL and HL with improved accuracy in predicting energy consumption and can facilitate early building design for the energy conservation.

The remainder of this paper is organized as below. The following section introduces the study context by reviewing the related literature, including studies of EPB and predictive techniques. Section 3 then describes research methodology and evaluation methods. Section 4 describes the building information and experimental data obtained in this study. Section 5 presents modeling processes, discusses prediction results, and compares model performance. Concluding remarks and research contributions are given in the final section.

2. Literature review

Recently, researchers have studied the use of AI techniques for predicting energy consumption. Many statistical and artificial intelligence techniques for inverse modeling of building heating and cooling loads have been developed [28,31–34]. Particularly, ANNs are very convenient and easy to use by an ordinary operator after the model has been established and are the most popular AI technique in many applications [24,26,27,35–39]. The ANN models do not require definition of explicit relationships between inputs and outputs as in conventional regression. They can model building heating and cooling loads of complex systems from independent or dependent parameters [40].

Olofsson et al., for example, combined ANN with quasi-physical description to predict the annual supplied space heating demand for numerous small single-family buildings [26]. Additionally, Aydinalp et al. [41] summarized the literature on ANN modeling and used outdoor temperature and electric home appliance usage as inputs in the ANN model for predicting annual residential energy usage. Their predictions of the annual energy consumption of 247 houses in Canada had an overall correlation coefficient of 0.909 with the annual actual energy consumption. In Kwok et al., an ANN model was employed to predict energy use by a Hong Kong office building. The best root-mean-squared-percentage-error (RMSPE) was 11.409% [31]. In Hou et al., an ANN based on data-fusion technique was used to forecast air-conditioning load and achieved a small relative error (below 4%) [27]. Although these studies show that ANN can achieve a moderate fit in predicting HL and CL, model performance is generally unsatisfactory.

In recent years, SVR, a variation of support vector machine (SVM), has been widely used in forecasting and regression [42]. Dong et al. evaluated the use of SVR to predict energy consumption in tropical regions [28]. The SVR was used to make hourly forecasts of building cooling load. Their study showed that SVR has superior prediction accuracy compared to conventional back propagation neural networks. Likewise, Li et al. used SVR to predict the hourly cooling load of a building [29,43]. The authors showed that SVR was better than ANNs. In addition, Jain et al. used SVR to forecast the energy consumption of multi-family residential buildings [32]. A sensor-based forecasting model using SVR was applied to an empirical data set for a multi-family residential building in New York City.

Many other AI techniques have been proposed for improving energy consumption prediction accuracy in the energy field. In Yu et al., CART provided accurate predictions of building energy demand with low errors [9]. Li et al. hybridized genetic algorithm with adaptive network-based fuzzy inference system to enhance accuracy in forecasting building energy consumption [33]. Tsanas and Xifara used random forest (RF) technique to estimate CL and HL in residential buildings [10]. Other proposed machine learning techniques for identifying and predicting system behavior include neural network system [44], general regression neural network [24], regression models [34,45], and hybrid system [33].

The above studies agree that AI model performs satisfactorily in predicting energy performance in building. Nevertheless, most works have used individual forecasting models rather than investigating the power of ensemble models. Moreover, the prediction performance of the aforementioned AI techniques needs further study. Therefore, the objective of this study was to fill this gap by using these models individually and in combination (ensemble models) to predict CL and HL *via* cross-fold validation and multiple performance measures.

3. Methodology

3.1. Support vector regression

Support vector machine was first introduced by Vapnik [46]. This supervised learning method generates an SVM by input-output mapping functions from a labeled training dataset. This function solves both classification and regression problems. Typically, regression is performed by using SVR, variation of SVM, to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. The input is first mapped onto an m-dimensional feature space by using non-linear functions as follows:

$$f(x, \omega) = \langle \omega, x \rangle + b \quad \text{with} \quad \omega \in \chi, b \in \chi \tag{1}$$

where $\langle \bullet, \bullet \rangle$ denotes the dot product in χ.

Thus, the goodness of the $f(x)$ can be estimated based on the loss function $L(x)$ as follows:

$$L(x, \omega) = [y, f(x, \omega)] = \begin{cases} 0 & \text{if} \quad |y - f(x, w)| \le \varepsilon \\ |y - f(x, w)| & \text{otherwise} \end{cases} \tag{2}$$

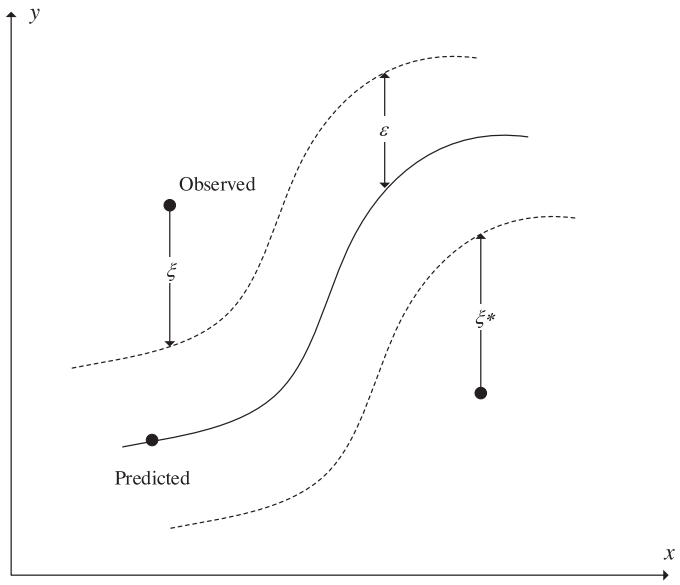

Fig. 1. Support vector regression.

The SVR uses ε parameter (insensitive loss) to compute a linear regression function for a dimensional feature space while concurrently minimizing $\|\omega\|^2$ to reduce model complexity. Two new parameters ξ_i, ξ_i^* are introduced to solve the minimization problem (Fig. 1). Thus, SVR is formulated as minimizing following constraint-optimization problem:

$$\text{Minimize} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}\left(\xi + \xi_i^*\right) \tag{3}$$

$$\text{Subject to:} \quad \begin{cases} y_i - \langle \omega, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \qquad\qquad \ge 0 \end{cases} \tag{4}$$

where the constant $C \ge 0$ determines the trade-off between the flatness of $f(x, \omega)$ and the tolerance for deviations larger than ε.

3.2. Artificial neural network

The ANN imitates the working principles of the human brain to perform learning and prediction tasks. The ANN is based on biological learning and has a structure similar to that of the human nervous system. A neural network uses inter-connected neurons as processing elements, which have similar characteristics as inputs, synaptic strength, activation output and bias. The inter-connections between neurons carry the weights of the network [47]. The neurons in the network can be classified as input, hidden and output neurons. Fig. 2 illustrates an ANN model.

The most widely used and effective neural network model is the back-propagation neural network (BPNN). Activation of each neuron in a hidden output layer is computed as Eq. (5). Notably, during the learning process, the BPNN stores non-linear information between influencing factors and their associated strengths. Connection weighs are adjusted during the training process to deliver predicted values close to target values. Therefore, BPNNs generally have an efficient training process.

$$net_k = \sum w_{kj} O_j \quad \text{and} \quad y_k = f(net_k) = \frac{1}{1 + e^{-net_k}} \tag{5}$$

where net_k is the activation of kth neuron; j is the set of neurons in the preceding layer; w_{kj} is the connection weight between neuron

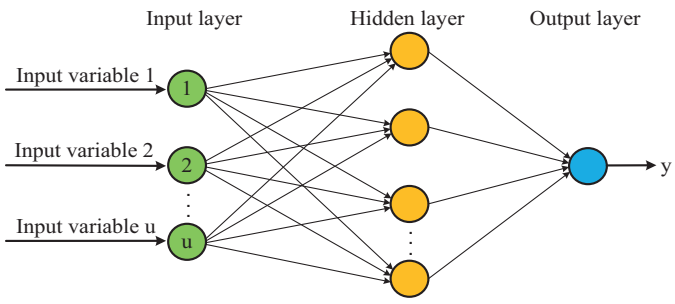

Fig. 2. Structure of ANN.

k and neuron j; O_j is the output of neuron j; and y_k is the sigmoid or logistic transfer function.

3.3. Classification and regression tree

The CART is a decision tree method of constructing a classification or regression tree according to its dependent variable type, which may be categorical or numeric [48]. In the rule-based CART, data are split into two subsets such that the new subset records have higher homogeneity, *i.e.*, higher purity, compared to those in the previous subset. Achieving the homogeneity criterion, however, requires a recursive splitting process. A CART is sufficiently flexible to consider misclassification costs and to specify the prior probability distribution in a classification problem. In terms of their logic rules, decision tree methods are markedly superior to other modeling techniques [49]. In a CART model, purity is defined as the similarity between values and target values and is considered perfect when all subset values are identical.

For a given target field, three impurity measures can be used to locate splits for CART models. Additionally, the target field is usually symbolized using Gini while the least squared deviation method is used for automatically selecting continuous targets without explaining the selection. The Gini index $g(t)$ at a node t in a CART is defined by the following equations:

$$g(t) = \sum_{j \ne i} p(j|t)p(i|t) \tag{6}$$

$$p(j|t) = \frac{p(j,t)}{p(t)}; p(j,t) = \frac{\pi(j)N_j(t)}{N_j}; \quad \text{and} \quad p(t) = \sum_j p(j,t) \tag{7}$$

where i and j are target field categories, $\pi(j)$ is the prior probability value for category j, $N_j(t)$ is the number of records in category j of node t, and N_j is the number of records of category j in the root node. Notably, when the Gini index is used to measure the improvement after a split during tree growth, only records in node t and the root node with valid values for the split-predictor are used to compute $N_j(t)$ and N_j, respectively.

3.4. Chi-squared automatic interaction detector

The chi-squared automatic interaction detector (CHAID) is a decision tree classification technique developed by Kass [50]. It tests for independence by using the chi-squared test to assess whether splitting a node significantly improves purity. The predictor with the strongest association (p-value) with the response variable at each node is used as a split node. When the tested predictor shows no statistically significant improvement, no split is performed, the algorithm will stop.

This study, however, proposes the use of exhaustive CHAID to classify the target field, which addresses the limitations of CHAID technique [51]. Specifically, exhaustive CHAID may not optimize the split for a predictor variable because it stops merging categories

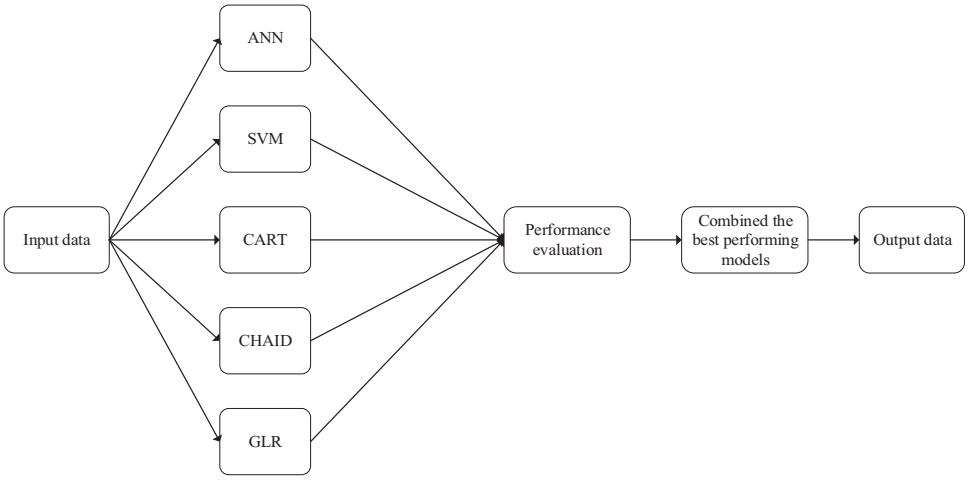

Fig. 3. Ensemble model.

as soon as it finds that all remaining categories significantly differ. Exhaustive CHAID avoids this problem by continuously merging predictor categories until only two super categories remain. After identifying the predictor in the series of merges, it finds the set of categories that has the strongest association with the target variable and computes an adjusted p-value for the association. Thus, exhaustive CHAID finds the best split for each predictor and then chooses which predictor to split by comparing their adjusted p-values [49].

3.5. General linear regression

General linear regression (GLR), a more flexible version of linear regression (LR), assumes that data points have an arbitrary distribution pattern. It constructs the relationship between X (predictor variables) and Y (response variable) by using a link function according to its distribution pattern. The $(X-Y)$ relational model is therefore defined as:

$$g(E(y)) = X \times \beta + O, y \sim F \tag{8}$$

where $g(\bullet)$ is the selected link function, O is the offset variable, F is the distribution model of y, X is the predictor, y is the response variable, and β is the regression coefficient.

The GLR uses Newton–Raphson method to obtain a continuous estimate such that $(X \times \beta + O)$ approaches $g(E(y))$.The final proximal equation is formulated as a $(X-Y)$ relational expression. Although additional parameters in the GLR increase model instability, GLR has a wider application range and obtains a more realistic relationship model compared to LR.

3.6. Ensemble model

An ensemble model is a numerical prediction method for generating a representative sample of the possible future states of dynamical system. This approach ranks a set of the models described above based on its performances and then combines the best performing models into an ensemble model. It can be expressed mathematically as g: $R^d \rightarrow R$ with a dimensional predictor variable X and one-dimensional response Y. Each procedure uses a specified algorithm to yield one estimated function $g(.)$. One estimation by an ensemble-base function $g_{cm}(.)$ is obtained by a linear combination of individual functions as follows:

$$g_{cm}(.) = \sum_{i=1}^{N} c_i \times g(.) \tag{9}$$

where c_i comprises the linear combination coefficients, which are simply based on average values of different weight (Fig. 3).

The artificial intelligence models used in this study to predict CL and HL include ANN, SVM, CART, CHAID, GLR, and ensemble approach. Notably, the ensemble approach method is expected to be superior in prediction performance to that of conventional models [52]. The generalization of predictive models can be enhanced by ensemble averaging. However, the effectiveness of the models needs further examination as described in the following sections. The case study demonstrates that the modeling process can be constructed easily, objectively, and satisfactorily in terms of both utilization and accuracy.

3.7. Evaluation criteria

Multiple evaluating criteria are utilized to compare the performance of prediction models. First, root mean squared error (*RMSE*) is computed to find the square error of the prediction compared to actual values and to find the square root of the summation value. The *RMSE* is thus the average distance of a data point from the fitted line measured along a vertical line. This tool efficiently identifies undesirably large differences. The *RMSE* is stated using the following equation.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \times \sum_{i=1}^{n} [p_i - y_i]^2} \tag{10}$$

where p_i is the predicted value; y_i is the actual values; and n is the sample size.

In contrast to *RMSE*, the mean absolute error (*MAE*) is a quantity used to measure how close forecasts are to the eventual outcomes. It computes the average magnitude of errors between predicted and actual values with disregard of direction of over or under of errors. The mean absolute error is given by:

$$MAE = \left(\frac{1}{n}\right) \times \sum_{i=1}^{n} \left(\left|p_i - y_i\right|\right) \tag{11}$$

The mean absolute percentage error (*MAPE*) is a statistical measure of predictive accuracy and is usually presented as a percentage. The *MAPE* is useful for evaluating the performance of predictive models because it gives relative values. The *MAPE* is effective for identifying the relative differences between models because it is

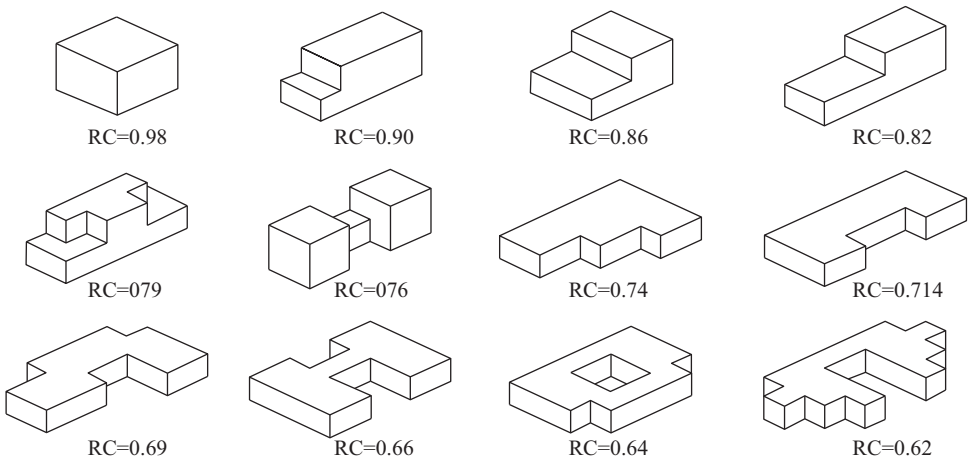

Fig. 4. Building shapes.

unaffected by the size or unit of actual and predicted values. The *MAPE* can be calculated using the following equation.

$$MAPE = \left(\frac{1}{n}\right) \times \sum_{i=1}^{n} \left(\frac{|p_i - y_i|}{y_i} \right) \times 100 \tag{12}$$

The linear correlation coefficient, *R*, is a common measure of how well the curve fits the actual data. A value of 1 indicates a perfect fit between actual and predicted values, meaning that the values have the same propensity. The mathematical formula for computing *R* is

$$R = \frac{n\sum_{i=1}^{n} y_i \times p_i - \sum_{i=1}^{n} y_i \times \sum_{i=1}^{n} p_i}{\sqrt{n \times \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} \times \sqrt{n \times \sum_{i=1}^{n} p_i^2 - \left(\sum_{i=1}^{n} p_i\right)^2}} \tag{13}$$

Moreover, to obtain a comprehensive performance measure, measures *RMSE*, *MAE*, *MAPE*, and 1-*R* are combined into a synthesis index (*SI*) as follows.

$$SI = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{P_i - P_{\min,i}}{P_{\max,i} - P_{\min,i}} \right) \tag{14}$$

where n is the number of performance measures and P_i is performance measure. The *SI* range is 0–1; an *SI* value close to 0 indicates a highly accurate prediction model.

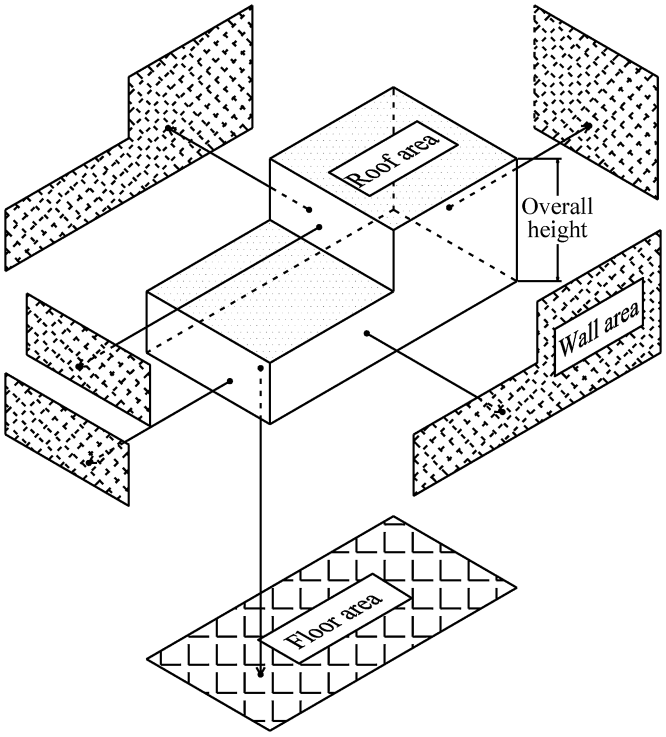

Fig. 5. Genetic definition of building areas.

4. Building information and data description

This work investigated twelve building types simulated in Ecotect [10]. Each type consists of various elementary blocks (see Fig. 4). All buildings have the same volume (771.75 m^3) and used the same materials for each block but have different surface areas and dimensions. The simulation data assumed that activities of the buildings were sedentary. The experiments investigated eight input parameters and two output parameters of residential buildings, *i.e.*, relative compactness (RC), surface area, wall area, roof area, overall

Table 1
Data description.

Description	Variable	Type of input/output	Min.	Max.	Mean
Relative compactness	X_1	Set	0.62	0.98	0.76
Surface area	X_2	Set	514.5	808.5	671.71
Wall area	X_3	Set	245	416.5	318.50
Roof area	X_4	Set	110.25	220.5	176.60
Overall height	X_5	Set	3.5	7	5.25
Orientation	X_6	Set	2	5	3.50
Glazing area	X_7	Set	0	0.4	0.23
Glazing area distribution	X_8	Set	0	5	2.81
Heating load (HL)	Y_1	Range	6.01	43.1	22.31
Cooling load (CL)	Y_2	Range	10.9	48.03	24.59

height, orientation, glazing area, glazing distribution, and cooling load/heating load.

Parameters known to have a significant influence on EPB were selected as input parameters [10,19,21]. The RC indicator [53] is used to show different building types. The RC of a building is derived by comparing its volume to surface area ratio to that of a building with the same volume but a more compact shape. Most building have orthogonal polyhedral shapes. Using the cube as the reference shape, RC is given by

$$RC = 6V^{2/3}A^{-1} \quad (15)$$

where V is the volume of building; A is the surface area of building.

Surface area is calculated as the total of wall area, roof area and floor area. Fig. 5 shows the details of the wall area, roof area, floor area and overall building height. Four major orientations are considered in the experiments: north, east, west and south. Three percentages of glazing area to floor area were 10%, 25% and 40%. Besides, five different glazing distributions are simulated: (1) uniform: with 25% glazing for each face, (2) north: 55% for the north face and 15% for each of the other faces, (3) east: 55% for the east face and 15% for each of the other faces, (4) south: 55% for the south face and 15% for each of the other faces, and (5) west: 55% for the west face and 15% for each of the other faces. Additionally, no glazing areas are simulated in the experiment. Therefore, 768 data cases were simulated. Table 1 provides the detailed input and output parameters in this study.

The simulation used materials with lowest U-values contemporaneously as follows: walls—1.78; floors—0.86; roofs—0.50; windows—2.26. The internal conditions of buildings were designed as follows: clothing—0.6 clo; humidity—60%; air speed—0.30 m/s, lighting level 300 lx. Sensitivity of internal gains was 5 and latency was 2 W/m^2. The change rate and wind sensitivity of infiltrated air were set to 0.5 and 0.25, respectively. The simulation used a mixed mode with 95% efficiency and a thermostat range of 19–24 °C. Operational time was 15–20 h on weekdays and 10–20 h on weekends. Additional details of the simulation experiments are provided by Tsanas and Xifara [10].

Table 2
Prediction model parameter settings.

Model	Parameter	Setting
ANN	Alpha (momentum)	0.9
	Initial Eta (initial learning rate)	0.3
	High Eta (high learning rate)	0.1
	Low Eta (low learning rate)	0.01
	Eta decay (learning rate decay)	30
	Transfer function	Sigmoid
	Hidden layers	One
	No. of neurons in the hidden layer	3
	Persistence	200
SVR	Stopping criteria	1.00E − 03
	Regularization parameter (C)	10
	Regression precision (epsilon)	0.1
	Kernel type	RBF
	RBF gamma	0.1
CART	Levels below root	5
	Mode	Simple
	Maximum surrogates	5
	Minimum change in impurity	0.0001
	Impurity measure for categorical targets	Gini
	Minimum records in parent branch (%)	2
	Minimum records in child branch (%)	1
CHAID	Mode	Simple
	Alpha for splitting	0.05
	Alpha for merging	0.05
	Chi-square method	Pearson
	Minimum records in parent branch (%)	2
	Minimum records in child branch (%)	1
	Epsilon for convergence	0.001
	Maximum iterations for convergence	100
	Allow splitting of merged categories	False
	Use Bonferroni adjustment	True
GLR	Distribution	Normal
	Singularity tolerance	1.00E − 07
	Value order for categorical inputs	Ascending
	Scale parameter method	Maximum likelihood estimate
	Covariance matrix	Model-based estimator
	Confidence interval level (%)	95

5. Prediction results and performance evaluation

5.1. Data preprocessing and model application

Model performance was assessed by a stratified k-fold cross-validation approach. For model evaluation, this approach mitigates bias caused by randomness in choosing testing cases [54]. Kohavi showed that 10 folds were optimal in terms of bias and variance associated with the validation process and in terms of time [55]. The present study thus used a stratified 10-fold cross-validation method to compare the average performance of the models. The average prediction results for 10 testing folds can then be used to appraise model performance. To do so, randomly selected data were divided into 10 distinct folds. Each fold was used in turn as a testing set with the remaining folds used as a training set so as

Fig. 6. Ten-fold cross-validation.

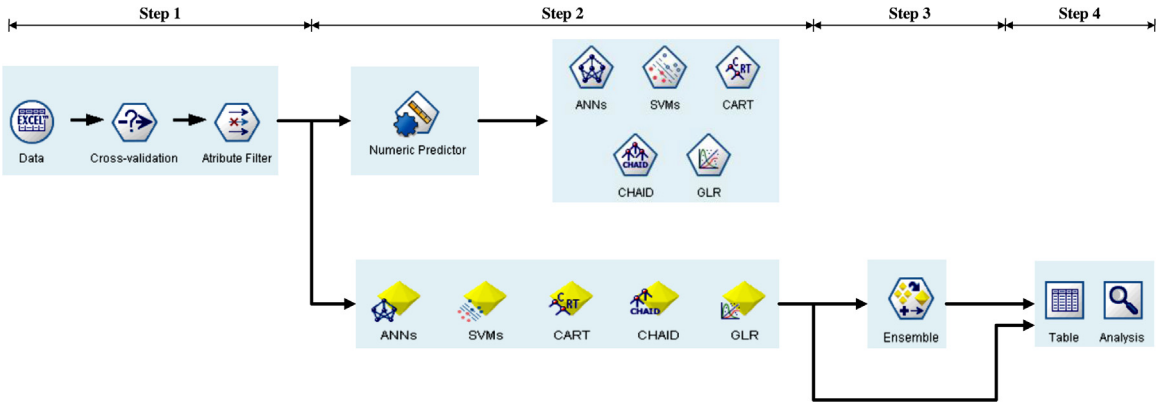

Fig. 7. Modeling stream for predicting heating and cooling loads.

to ensure that all data instances were applied in both training and testing phases (see Fig. 6). Algorithm accuracy is then expressed as the average accuracy of the ten models in ten validation rounds.

The predictions results were obtained using the Numeric predictor node (*i.e.*, ANN, CART, CHAID, GLR and SVR model) in Clementine (now IBM SPSS Modeler) [49]. The Numeric predictor node estimates and compare models for continuous numeric range outcomes using several different methods. The node allows the choice of which algorithms to use and allows experimentation with multiple combinations of options in a single modeling path. Model comparisons can be based on correlation, relative error, or number of variables used. The parameters of prediction models (Table 2) were set to default in Clementine to establish a baseline for validation. For example, the ANN's parameters were set as sigmoid transfer function, hidden layers = 1, No. of neurons in the hidden layer = 3, momentum = 0.9, initial learning rate = 0.3, high learning rate = 0.1, low learning rate = 0.01, learning rate decay = 30 and persistence = 200.

Each of the 2 outputs in the experiment was run 10 times with 10 fold-cross validation. The analytical results were then averaged. In each fold, the training data were used to create a prediction model. The model was then evaluated using testing data. The steps for each fold are similar and are briefly listed below (Fig. 7):

- Step 1: Input phase: add data to the source node based on the cross-validation algorithm.
- Step 2: Cross training and testing model: use the numerical predictor node to train data and use the created nugget node to evaluate the model.
- Step 3: Combine the best models *via* an ensemble node.
- Step 4: Output phase: assess analytical results through table and analysis nodes.

5.2. Analytical discussions

Table 3 shows the performance results for the prediction models, including ANN, SVR, CART, CHAID, GLR and ensemble models. The accuracy measures were used to evaluate the predictive AI techniques. Table 3 lists the summary of cross-fold modeling performance for cooling load and heating load during the test period. For example, SVR (0.00) had the best results based on the *SI* values for heating load.

Particularly, the ensemble model (*i.e.*, SVR + ANN) had an even better overall performance in predicting CL, and SVR was the best model for predicting HL (Table 3). Aggregate performance of the ensemble models was evaluated by identifying the top two, the top three, the top four, and the top five single models. All ensemble models obtained excellent correlations between actual values and predicted output (higher than 0.97, see Table 3). In CL phases, the best ensemble model was derived from the two best single models (SVR + ANN). Notably, the best performance of the ensemble model was superior to that of the best performance of the individual predictive models. However, SVR had the best performance in predicting HL.

The energy industry clearly recognizes the importance of predicting the energy performance of buildings. Several models have been proposed for comparing and improving the accuracy in predicting CL and HL (Table 4). For example, Tsanas and Xifara proposed a random forest (RF) method to estimate the CL and HL in residential buildings [10]. Their models performed well. For predicting CL, they obtained high values for *RMSE* = 2.567 (kW), *MAE* = 1.42 (kW) and *MAPE* = 4.62 (%). Their comparisons showed that the prediction performance of the RF model was superior to that of iteratively reweighted least squares (IRLS) models [56].

Table 4 describes the analytical results for the proposed best single and ensemble models comparing to results reported in other

Table 3
Average 10-fold cross-validation results for five single models and ensemble models.

Model	Cooling load					Heating load				
	RMSE (kW)	*MAE* (kW)	*MAPE* (%)	*R*	*SI*	*RMSE* (kW)	*MAE* (kW)	*MAPE* (%)	*R*	*SI*
ANN	1.678	1.158	4.403	0.984	0.53 (2)	0.610	0.432	2.362	0.998	0.34 (2)
SVR	1.647	0.890	2.985	0.985	0.14 (1)	0.346	0.236	1.132	0.999	**0.00 (1)**
CART	1.841	1.157	4.020	0.981	0.76 (3)	0.800	0.437	2.104	0.996	0.51 (3)
CHAID	1.859	1.174	4.104	0.981	0.82 (5)	0.909	0.469	2.407	0.995	0.64 (4)
GLR	1.740	1.292	4.966	0.983	0.79 (4)	1.039	0.787	4.591	0.995	1.00 (5)
Combined 5 single models	1.614	1.030	3.539	0.986	0.23	0.539	0.345	1.610	0.998	0.20
Combined 4 best single models	1.624	1.021	3.525	0.985	0.24	0.526	0.329	1.593	0.999	0.19
Combined 3 best single models	1.610	1.000	3.474	0.986	0.20	0.488	0.315	1.581	0.999	0.15
Combined 2 best single models	1.566	0.973	3.455	0.986	**0.11**	0.428	0.300	1.557	0.999	0.11

Note: Underline value denotes the best overall prediction performance (*SI*) for the cooling and heating loads, respectively; (.): denotes performance ranking for single models.

primary works [10,56]. Based on the numerical evaluation, Fig. 8 further shows that the performance measures (*RMSE*, *MAE*, and *MAPE*) of best models identified in Table 3 are superior to those models reported in literature. In CL phase, the ensemble model had a higher prediction accuracy compared to the other individual models used in this study and compared to those reported in literature. Table 5 presents the improvement achieved by the proposed models across CL and HL data. In tests of the prediction performance of CL, the ensemble model (SVR + ANN) had error rates (*i.e.*, *RMSE*, *MAE*, and *MAPE*) 25.2–63.3% better than those reported in previous works. For the HL data, the SVR model had error rates 48.1–88.8% better than models reported in previous works.

For further comparisons of the performance of ensemble models in predicting CL and HL, a *t*-test was used to assess whether they were significantly better than models reported in previous works. In this sense, the null hypothesis H_0 associated with prediction errors of ensemble model (μ) is assumed $\geq$ than single model/previous work (μ_0) in *MAE*, *RMSE* and *MAPE* test cases. The rejection region must be in the form ($\mu < \mu_0$), such that $P = (\mu < \mu_0) = \alpha$ (significance). Therefore, the alternative hypothesis is the denial of H_0. The hypothesis test results confirmed the significantly improved performance measures obtained for the ensemble models in CL and HL phase. For predicting CL, the ensemble models were also superior to the single model and RF. However, SVR was the best model for predicting HL.

Table 4
Performance comparison of the best models and other primary works.

Phase	The best models and models reported in other primary work	*RMSE* (kW)	*MAE* (kW)	*MAPE* (%)
Cooling load	SVR + ANN	**1.566**	**0.973**	**3.455**
	IRLS [56]	3.385	2.210	9.410
	RF [10]	2.567	1.420	4.620
Heating load	SVR	**0.346**	**0.236**	**1.132**
	IRLS [56]	3.142	2.140	10.090
	RF [10]	1.015	0.510	2.180

Note: underline value denotes the best performance measure for the cooling and heating loads, respectively.

Notably, computing efficiency is another critical factor that should be evaluated during model building. The evaluation of a model should consider both computing and prediction accuracy. For each single model, Table 6 shows the computing time in predicting CL and HL and the computer specifications. The computation data shows that all of the proposed AI models can predict CL and HL within seconds. Therefore, these models provide a good basis for developing near-real-time energy performance management systems.

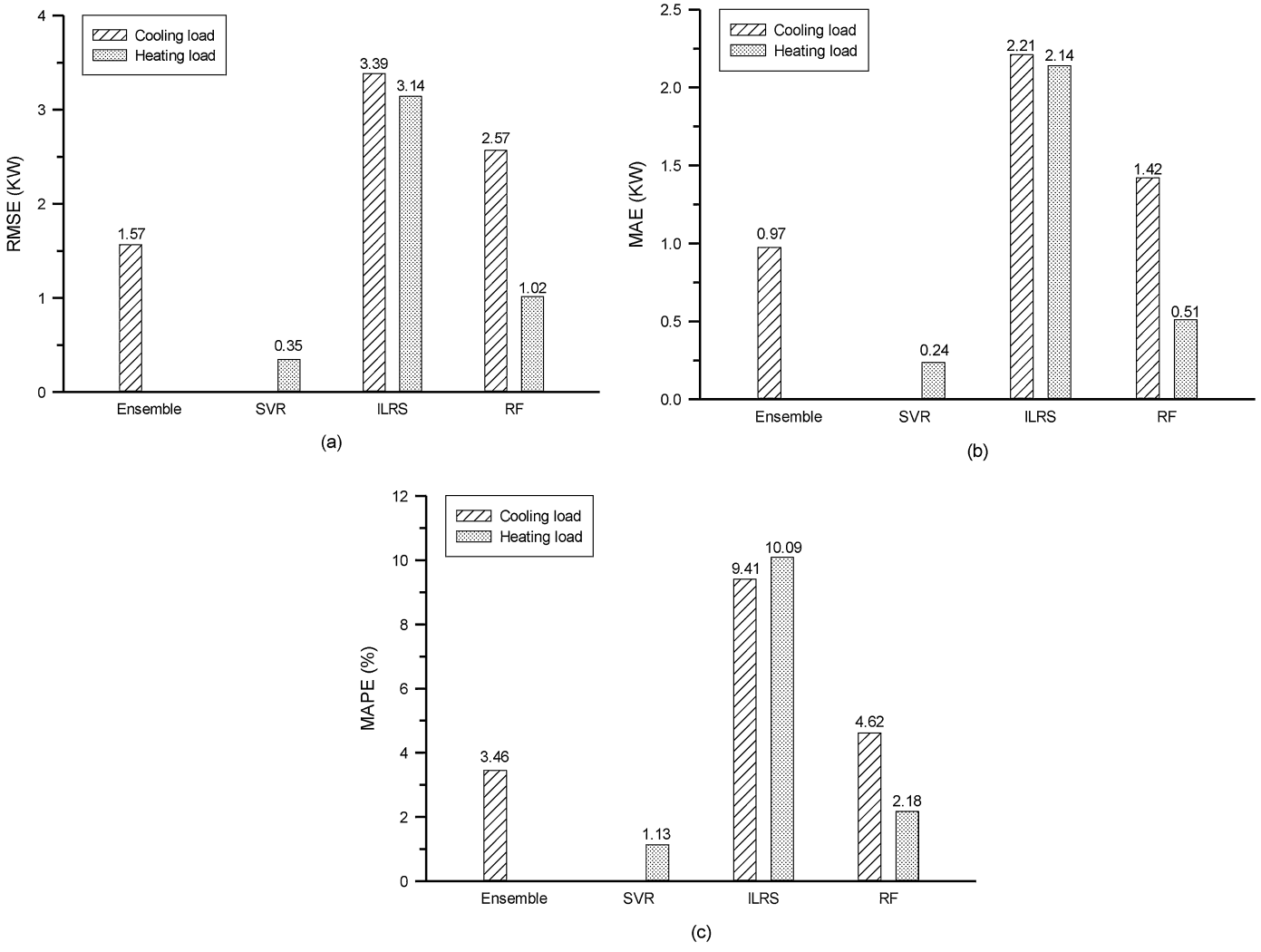

Fig. 8. Criteria for evaluating comparative models.

Table 5
Hypothesis testing results for improvement rates in the best models.

Phase	Second best model and models reported in other primary work	Best model	Improved by the best model (%)			
			RMSE	*MAE*	*MAPE*	*R*
Cooling load	SVR	SVR + ANN	4.9	–	–	0.2
	IRLS [56]		53.7***	56.0***	63.3***	N/A
	RF [10]		39.0***	31.5***	25.2***	N/A
Heating load	SVR + ANN	SVR	19.1**	21.4***	27.3***	0.0
	IRLS [56]		89.0***	89.0***	88.8***	N/A
	RF [10]		65.9***	53.8***	48.1***	N/A

The improvement and hypothesis testing are calculated using average performance measures.
** Represents the level of significance is higher than 1%.*** Represents the level of significance is higher than 0.1%.

Table 6
Computing time and hardware specification.

Model	Average time (s)		Hardware specification & software adopted	
	CL	HL		
ANN	0.51	0.56	CPU	Intel Celeron 2.0 GHz
SVR	1.00	1.03	Ram	2GB
CART	0.62	0.61	OS	Window 7 32-bit
CHAID	0.59	0.58	Software	IBM SPSS Modeler®
GLR	0.58	0.58		
Ensemble	1.01	1.06		

6. Conclusions

Various data-mining techniques, including ANN, SVR, CART, CHAID, GLR and ensemble model, were compared in terms of speed and performance in predicting building CL and HL. The proposed approach is easily performed and has many potential applications in building energy prediction. Various building characteristics were used as input to CL and HL. Data for 768 cases of CL and HL were used to construct the prediction models. A 10-fold cross-validation method was used to mitigate bias in comparisons of model performance. The analytical results demonstrate the applicability of data-mining techniques for forecasting energy consumption by buildings.

This work also confirms that the ensemble model (SVR + ANN) and SVR substantially improve performance in predicting CL and HL, respectively. They also reduce the amount of work required to develop complex computing algorithms. Specifically, the robustness and prediction accuracy of the proposed model are superior to those of other models (Table 4), and the *MAPE*s are improved between 25.2% and 88.8% with a confidence level of 99.9%. The comparison results further indicate that the ensemble models for predicting CL have *MAPE*s below 4% and are an average of 25.2% better than those of RF. For HL prediction, the SVR improved *RMSE* by 65.9%, *MAE* by 53.8%, and *MAPE* by 48.1% compared to RF.

Notably, the proposed SVR requires only 1.03 s to predict HL, and the ensemble model (SVR + ANN) requires 1.01 s to predict CL. Therefore, these models provide a good basis for developing a real-time building energy performance management system in future works. The proposed models are easy to use, and have fewer parameters and experiments to control compared to simulation software. The ensemble approach (SVR + ANN) can be the default prediction model for CL while the SVR model is most appropriate for HL prediction.

In practice, energy engineers may apply ensemble model and SVR for early predictions of CL and HL, respectively, during the design process. Using these models would provide designs with adequate CL and HL. By using the proposed artificial intelligence techniques as an effective tool for designing and analyzing energy efficient buildings, they can play an important role in the energy conservation. Moreover, fast and accurate predictions of building CL and HL can alleviate engineers' design loading and learning curves. Hence, they gain less cost in energy-efficiency design and more time for other building works. Additionally, accurate prediction at the design stage can help building systems work properly, and thus lead to improved occupant comfort, longer life cycle of equipment, energy efficiency and reduced maintenance cost.

However, the limitation of this work is that it used default settings in the single and ensemble models. Additionally, the developed models are only applicable to the twelve specified building types with a controlled experiment setup. In recent works, individual AI has been integrated with other computing paradigms such as genetic algorithm, fuzzy logic, particle swarm optimization, and firefly colony algorithm to enhance the performance of single AI models in many civil engineering fields. Further studies can focus on investigating parameters optimization in single and ensemble models by using evolutionary computing methods and swarm intelligence algorithms to achieve further improvements in their accuracy in predicting CL and HL in buildings.

References

[1] M.S. Al-Homoud, Computer-aided building energy analysis techniques, Building and Environment 36 (4) (2001) 421–433.
[2] D.A. York, E.F. Tucker, C.C. Cappiello, DOE-2 Reference Manual Version 2.1, Los Alamos, New Mexico, 1980.
[3] P.A. Strachan, G. Kokogiannakis, I.A. Macdonald, History and development of validation with the ESP-r simulation program, Building and Environment 43 (4) (2008) 601–609.
[4] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, J. Glazer, EnergyPlus: creating a new-generation building energy simulation program, Energy and Buildings 33 (4) (2001) 319–331.
[5] D. Yan, J. Xia, W. Tang, F. Song, X. Zhang, Y. Jiang, DeST—an integrated building simulation toolkit Part I: Fundamentals, Building Simulation 1 (2) (2008) 95–110.
[6] D. Thevenard, K. Haddad, Ground reflectivity in the context of building energy simulation, Energy and Buildings 38 (8) (2006) 972–980.
[7] J.C. Lam, K.K.W. Wan, C.L. Tsang, L. Yang, Building energy efficiency in different climates, Energy Conversion and Management 49 (8) (2008) 2354–2366.
[8] N. Eskin, H. Türkmen, Analysis of annual heating and cooling energy requirements for office buildings in different climates in Turkey, Energy and Buildings 40 (5) (2008) 763–773.
[9] Z. Yu, F. Haghighat, B.C.M. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy and Buildings 42 (10) (2010) 1637–1646.
[10] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy and Buildings 49 (0) (2012) 560–567.
[11] Z. Li, G. Huang, Re-evaluation of building cooling load prediction models for use in humid subtropical area, Energy and Buildings 62 (0) (2013) 442–449.
[12] IEA, Key World Energy Statistics, 2009.
[13] European Parliament and Council, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), Official Journal of the European Union L153 (0) (2010) 13–35.
[14] A.G. Kwok, N.B. Rajkovich, Addressing climate change in comfort standards, Building and Environment 45 (1) (2010) 18–22.
[15] R. Yao, B. Li, K. Steemers, Energy policy and standard for built environment in China, Renewable Energy 30 (13) (2005) 1973–1988.
[16] M.C. Leung, N.C.F. Tse, L.L. Lai, T.T. Chow, The use of occupancy space electrical power demand in building cooling load prediction, Energy and Buildings 55 (0) (2012) 151–163.

[17] M. Ning, M. Zaheeruddin, Neuro-optimal operation of a variable air volume HVAC&amp: R system, Applied Thermal Engineering 30 (5) (2010) 385–399.
[18] B.C. Ahn, J.W. Mitchell, Optimal control development for chilled water plants using a quadratic representation, Energy and Buildings 33 (4) (2001) 371–378.
[19] W. Pessenlehner, A. Mahdavi, Building morphology, transparence, and energy performance, in: Eighth International IBPSA Conference, Eindhoven, The Netherlands, 2003, pp. 1025–1030.
[20] K.K.W. Wan, D.H.W. Li, D. Liu, J.C. Lam, Future trends of building heating and cooling loads and energy consumption in different climates, Building and Environment 46 (1) (2011) 223–234.
[21] S. Schiavon, K.H. Lee, F. Bauman, T. Webster, Influence of raised floor on zone design cooling load in commercial buildings, Energy and Buildings 42 (8) (2010) 1182–1191.
[22] J. Parasonis, A. Keizikas, A. Endriukaitytė, D. Kalibatienė, Architectural solutions to increase the energy efficiency of buildings, Journal of Civil Engineering and Management 18 (1) (2012) 71–80.
[23] N. Fumo, A review on the basics of building energy estimation, Renewable and Sustainable Energy Reviews 31 (0) (2014) 53–60.
[24] A.E. Ben-Nakhi, M.A. Mahmoud, Cooling load prediction for buildings using general regression neural networks, Energy Conversion and Management 45 (13–14) (2004) 2127–2141.
[25] B.B. Ekici, U.T. Aksoy, Prediction of building energy consumption by using artificial neural networks, Advances in Engineering Software 40 (5) (2009) 356–362.
[26] T. Olofsson, S. Andersson, R. Östin, A method for predicting the annual building heating demand based on limited performance data, Energy and Buildings 28 (1) (1998) 101–108.
[27] Z. Hou, Z. Lian, Y. Yao, X. Yuan, Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique, Applied Energy 83 (9) (2006) 1033–1046.
[28] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy and Buildings 37 (5) (2005) 545–553.
[29] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, Applied Energy 86 (10) (2009) 2249–2256.
[30] A.C. Menezes, A. Cripps, R.A. Buswell, J. Wright, D. Bouchlaghem, Estimating the energy consumption and power demand of small power equipment in office buildings, Energy and Buildings 75 (0) (2014) 199–209.
[31] S.S.K. Kwok, E.W.M. Lee, A study of the importance of occupancy to building cooling load in prediction by intelligent approach, Energy Conversion and Management 52 (7) (2011) 2555–2564.
[32] R.K. Jain, K.M. Smith, P.J. Culligan, J.E. Taylor, Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy, Applied Energy 123 (0) (2014) 168–178.
[33] K. Li, H. Su, J. Chu, Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study, Energy and Buildings 43 (10) (2011) 2893–2899.
[34] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy and Buildings 57 (0) (2013) 302–312.
[35] S.A. Kalogirou, M. Bojic, Artificial neural networks for the prediction of the energy consumption of a passive solar building, Energy 25 (5) (2000) 479–491.
[36] H.-T. Pao, Comparing linear and nonlinear forecasts for Taiwan's electricity consumption, Energy 31 (12) (2006) 2129–2141.
[37] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy and Buildings 40 (12) (2008) 2169–2176.
[38] A. Yezioro, B. Dong, F. Leite, An applied artificial intelligence approach towards assessing building performance simulation tools, Energy and Buildings 40 (4) (2008) 612–620.
[39] J. Zhang, F. Haghighat, Development of artificial neural network based heat convection algorithm for thermal simulation of large rectangular cross-sectional area earth-to-air heat exchangers, Energy and Buildings 42 (4) (2010) 435–440.
[40] A. Rabl, A. Rialhe, Energy signature models for commercial buildings: test with measured data and interpretation, Energy and Buildings 19 (2) (1992) 143–154.
[41] M. Aydinalp, V. Ismet Ugursal, A.S. Fung, Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks, Applied Energy 71 (2) (2002) 87–110.
[42] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (1) (2004) 113–126.
[43] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks, Energy Conversion and Management 50 (1) (2009) 90–96.
[44] N. Kashiwagi, T. Tobi, Heating and cooling load prediction using a neural network system, in:, IJCNN '93—Nagoya. Proceedings of 1993 International Joint Conference on Neural Networks 931 (1993) 939–942.
[45] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, V.I. Hanby, Regression models for predicting UK office building energy consumption from heating and cooling demands, Energy and Buildings 59 (0) (2013) 214–227.
[46] V.N. Vapnik, The nature of statistical learning theory, Springer-Verlag, New York, NY, 1995.
[47] T.N. Singh, S. Sinha, V.K. Singh, Prediction of thermal conductivity of rock through physico-mechanical properties, Building and Environment 42 (1) (2007) 146–155.
[48] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Chapman & Hall/CRC, New York, NY, 1984.
[49] SPSS, Clementine 12.0 Algorithm Guide, Integral Solutions Limited, Chicago, IL, 2007.
[50] G.V. Kass, An exploratory technique for investigating large quantities of categorical data, Journal of the Royal Statistical Society, Series C: Applied Statistics 29 (2) (1980) 119–127.
[51] D. Biggs, B. De Ville, E. Suen, A method of choosing multiway partitions for classification and decision trees, Journal of Applied Statistics 18 (1) (1991) 49–62.
[52] P.J.L. Adeodato, A.L. Arnaud, G.C. Vasconcelos, R.C.L.V. Cunha, D.S.M.P. Monteiro, M.L.P. ensembles improve long term prediction accuracy over single networks, International Journal of Forecasting 27 (3) (2011) 661–671.
[53] A. Mahdavi, B. Gurtekin, Shapes, numbers, and perception: aspects and dimensions of the design performance space., in: Proceedings of the 6th International Conference: Design and Decision Support Systems in Architecture, Ellecom, The Netherlands, 2002, ISBN 90-6814-141-4, pp. 291–300.
[54] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York Inc., New York, NY, 2006.
[55] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence—vol. 2, Morgan Kaufmann Publishers Inc., Montreal, QC, 1995, pp. 1137–1143.
[56] D.B. Rubin, Iteratively reweighted least squares, Encyclopedia of Statistical Sciences 4 (1983) 272–275.