## REVIEW

**Open Access**

# Machine learning for estimation of building energy consumption and performance: a review

Saleh Seyedzadeh[1*], Farzad Pour Rahimian[2], Ivan Glesk[1] and Marc Roper[3]

**Abstract**

Ever growing population and progressive municipal business demands for constructing new buildings are known as the foremost contributor to greenhouse gasses. Therefore, improvement of energy efficiency of the building sector has become an essential target to reduce the amount of gas emission as well as fossil fuel consumption. One most effective approach to reducing CO2 emission and energy consumption with regards to new buildings is to consider energy efficiency at a very early design stage. On the other hand, efficient energy management and smart refurbishments can enhance energy performance of the existing stock. All these solutions entail accurate energy prediction for optimal decision making. In recent years, artificial intelligence (AI) in general and machine learning (ML) techniques in specific terms have been proposed for forecasting of building energy consumption and performance. This paper provides a substantial review on the four main ML approaches including artificial neural network, support vector machine, Gaussian-based regressions and clustering, which have commonly been applied in forecasting and improving building energy performance.

**Keywords:** Building energy consumption, Building energy efficiency, Energy benchmarking, Machine learning

## MAIN TEXT
## Introduction

Emission of greenhouse gases including carbon dioxide ($CO_2$) in higher layers of the atmosphere are known as the main cause of global warming phenomena. In UK buildings are responsible for 46 percent of all $CO_2$ emissions (Kelly et al. 2012). This figure is 40 percent in the USA and 27 percent in Australia (Filippin 2000). Therefore, enhancement of energy efficiency of the buildings has become an essential issue to reduce the amount of gas emission as well as fossil fuel consumption. An annual saving of 60 billion Euros is estimated by improvement of European Union (EU) buildings' energy performance by 20 percent (Li et al. 2010).

The attempt to decrease the amount of greenhouse gases needs significant alteration in human behaviour in energy consumption, manufacturing of more environmental friendly products and identifying and mitigating the causes of these undesirable gases (Abrahamse et al. 2007). Therefore, enhancement of techniques for construction of more energy efficient buildings and improvement of current buildings' energy usage seem to be great moves in the reduction of global warming menace.

The first step in enhancement of building energy consumption is to calculate this amount using a building energy assessment method which is an informative tool providing a comparative energy performance index to decision makers. Generally, the energy consumption of building during a definite period normalised by floor area is used to express the performance ($kWh/m^2/period$) known as Energy Performance Indicator (EPI) or Energy Use Intensity (EUI) (Hong et al. 2015; Nikolaou et al. 2015).

Building energy assessment are separated into four main categories: engineering calculation, simulation model-based benchmarking and statistical modellings and Machine learning (ML). The engineering methodologies employ physical laws for the derivation of building energy consumption in whole or sub-system levels. The most

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 2 of 20

precise methods apply complex mathematics or building dynamics for the derivation of accurate energy usage for all building components considering internal and external details as the inputs (e.g. climate information, construction fabric, HVAC system). Building energy efficiency simulation includes software and computer models for simulation of performance with predefined status. Generally, computer simulation can be used for a variety of applications such as lighting and HVAC system design.Existing of building energy data has allowed usage of top-down methods for assessment of energy performance. The statistical methods use building historical data and frequently apply regression to model the energy consumption/performance of buildings. These models are also called data-driven surrogate models as they take advantage of existing data instead of relying upon system complex detail. ML as a subset of artificial intransigence provides the ability to learn from data using computer algorithms. The concept of ML is intimately associated with computational statistic. Hence, this method can be also considered as a subcategory of statistical modelling.

This paper reviews state-of-the-art application of ML methods in building energy analysis, estimation and benchmarking by emphasising the advantage and drawbacks, provides the discussion of potential improvement in model efficiency, applications and future recommendations.

First a brief introduction of the motivation and necessity of using ML in building energy filed is presented. Then different ML methods are explained in detail and review of model utilisation in building sector is thoroughly discussed, followed by the summary of these modes providing further information of buildings characteristics (case studies). Based on the discussions on different cases and usages a framework for selecting the most appropriate ML method is proposed. Finally the conclusion is derived to highlight the current challenges in ML and limitations of seminal works, and possible research opportunities for improving energy prediction and benchmarking using ML.

## Motivation

In the last decade, Zero Energy Building (ZEB) has been received huge attention and recognised as the primary design concept for future buildings in most countries (Marszal et al. 2011). On the other hand, building energy efficiency retrofit (BEER) of existing stock is considered as the chief energy reduction factor. In the UK and some European countries, the rate of demolition of existing buildings and constructing new ones is very low as 0.1 percent, whilst having new buildings rate of over 1 percent. It is estimated that at least 70 percent of existing buildings will be still occupied in 2050 (Bell 2004). It has been discussed that finding a sustainable Building

Energy Efficiency Retrofit (BEER) is very challenging and a decision-making tool is essential to propose appropriate retrofit technologies for a specific case (Ma et al. 2012; Ascione et al. 2016).

In order to facilitate decision-making in selecting suitable solutions where there are more than one objectives, there have been some methodologies in place which can be classified under the categories of priori and multi-objective optimisation (MOO) approaches (Wang et al. 2014; Ascione et al. 2014). Most of the developed methods are simulation-based optimisations in which the optimisation algorithms are implemented using a programming language, and the energy-related objectives (energy consumption or gas emission) are calculated by a Building Performance Simulation (BPS) tools such as EnergyPlus (Crawley et al. 2001), TRNSYS (University of Wisconsin-Madison 2015), ESP-r (The Energy Systems Research Unit (ESRU) 2011), etc. This approach limits the computation complexity of the algorithm to BPS's calculation time, in essence when a large number of solutions are defined the process may become extremely costly to handle. This time overhead is the main reason that most related studies have only investigated simple models or retrofitted only one or two parts of the studied envelopes. For the same reason, most of the studies targeted residential buildings, and there are only few reports on optimisation of retrofitting commercial properties (Smarra et al. 2018).

When performed in the early design stages, enhancement of energy efficiency of new stock is more flexible than improving existing buildings, since the structural limitations are far less in new built. Yet, it still requires an enormous amount of simulation, if an optimisation algorithm is utilised. A practical solution to address the design and BEER issues is the development of a data-driven (surrogate) model using historical data. In this method collocated building data (structural characteristics and climate data) is used to predict energy parameters of new samples by applying a learning process.

The application of data-driven models is not limited to only BEER and ZEB design, they are useful tools for optimisation of Energy Management System (EMS) and Heating, Ventilating, and Air Conditioning systems (HVAC) and even a better alternative for traditional building energy benchmarking and rating schemes (Dounis and Caraiscos 2009; Gao and Malkawi 2014; Deb et al. 2016).

EMSs along with information systems have been utilised for energy data collection and consumption control, which are fundamental operations in the achievement of energy waste reduction and also efficiency awareness advancement. As such, a great amount of data related to sensors and weather information is generated, and there is a demand for analytical tools that enable energy performance measurement assessment and future consumption forecasting. This allows smart energy control (Shaikh et al. 2014),

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 3 of 20

fault detection (Magoulès et al. 2013; He et al. 2011; Liang and Du 2007), potential energy efficiency options and calculation of achieved energy savings. The suitable statistical model is required to learn from flowing data and maintain its accuracy continuously (Yang et al. 2005).

Similarly, accurate estimation of heating and cooling load is the foundation of successful design of HVAC system which leads to reduced operational cost (through saving an amount of energy consumption by end users). Besides, in air-conditioned buildings employing thermal energy storage, this kind of prediction is vital for optimising the system. Kalogirou (Kalogirou et al. 2001) indicated that calculation of loads, especially in non-domestic buildings, is expensive and time consuming for consulting firms. Hence, an alternative solution is required to efficiently operate the HVAC systems, which also can facilitate comfortable temperature and humidity conditions (Kumar et al. 2013). Furthermore, advance forecasting of electricity loads allows determination of excessive usage periods, reduced peak demand and a load of electrical HVAC system.

Short-term energy estimation of individual cases only considers climate information (temperature, humidity of solar radiation), however, the precise prediction of building energy consumption and efficiency becomes a challenge when various affecting features such as structural characteristics (e.g. insulation, glazing, window to wall ratio and orientation), occupancy, appliances, variety of loads, operation hours etc. are taken into account (Zhao and Magoulès 2012b; Ahmad et al. 2014).

In order to highlight the importance of building energy efficiency and increase the public awareness and motivation, in some countries, buildings are assigned by energy labels or ratings (Chung 2011). In the majority of benchmarking schemes, BPS is the critical tool for evaluating building energy performance and then it is compared with a reference building. Hence, a similar issue as mentioned earlier applies in this case as well. Moreover, an expert engineer and building complex building characteristics are required to produce reliable outcomes. Learning models seem to have a promising application in benchmarking as they have the ability to extract the patterns underlying in various features of building data sets, which can be used for smart classification of buildings and determination of realistic reference point for different classes. In addition, they can learn from previous samples to estimate the rating or label of future cases.

Classification can even provide a foundation for evaluation of a specific feature impact in energy loads by first grouping samples based on unrelated (to the intended feature) variables. This method is very beneficial where analysing the impact a parameter such as occupancy behaviour becomes intricate using traditional mathematical or simulation modelling (Yu et al. 2011).

The suggested methodologies use statistical techniques to predict and evaluate energy performance based on collected data from building/s and environment and involve a kind of regression to model the energy characteristics. Simple and multivariate regression (MLR) are among widely used models that relate energy consumption to one or more variables (Hygh et al. 2012). Change-point regression method is also modelled based on the nonlinear impact of parameters that are mainly applied when buildings show a strong correlation between operation time and loads (Ruch et al. 1993). Data envelopment analysis (Mousavi-Avval et al. 2011) and stochastic frontier analysis (Kavousian and Rajagopal 2014) are among the mathematical model applied in this field.
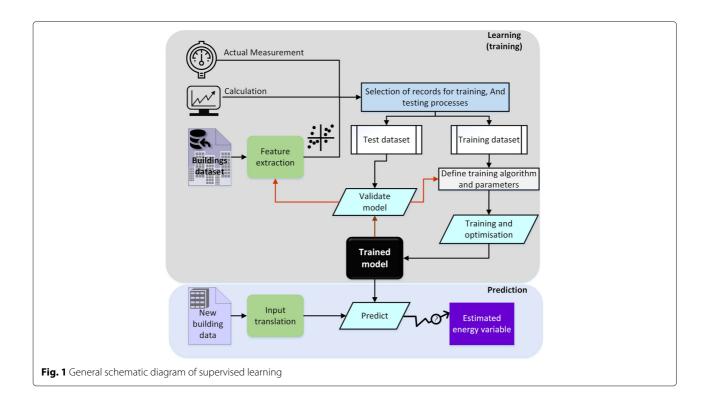
By considerable growth in the amount of valid and attainable dataset of buildings, there is an excellent interest in the utilisation of Artificial Intelligent (AI) methods specifically ML in the construction sector. Moreover, it is indicated that in order to conduct successful projects it is essential to learn and adopt novel technologies in the filed (Pour Rahimian et al. 2014). The most applied ML techniques in this field are Artificial Neural Network (ANN), Support Vector Machine (SVM), Gaussian distribution regression and clustering.

## Review

ML is generally used to describe a computer algorithm that learns from existing data. These algorithms typically use a considerable amount of data and relatively small number of input features for the learning process. In recent years, numerous ML techniques have been proposed in building sector for estimation of heating and cooling loads, energy consumption and performance for various circumstances.

ML models operate as a black box and need no information on building systems. They discover the relation between various input features and output targets (e.g. energy performance) using given data. When the ML models are trained with enough amount of data, they can be used to predict targets for unseen samples, though the relation between the features and the targets is not defined. This procedure is also known as supervised learning in ML field. In this case, the targeted energy parameter is calculated using simulation (in general engineering method) or measured and used for training the model. The general scheme of supervised learning for modelling building energy is illustrated in Fig. 1.

The second method of ML namely as unsupervised learning have received considerable attention in building energy analysis. Unsupervised learning also known as unsupervised classification is mainly applied to unlabeled data to cluster them based on hidden pattern and similarities underlying in features. This method is very beneficial for the application of energy benchmarking where

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 4 of 20



**Fig. 1** General schematic diagram of supervised learning

a determination of baseline buildings is crucial for calculating the energy performance of similar cases. Hence, the clustering algorithms provide more precise tools for grouping various building in comparison with traditional method where mainly relay on building usage type. It should be noted that using the clustering algorithm for forming groups it is not possible to estimate clusters for new buildings. Hence, for the purpose of determining the reference building for other cases, an extra supervised ML technique should be applied. In this approach, all buildings employed for clustering are used as training samples for classification where the generated labels from clustering are considered as learning targets. The flowchart of the overall procedure is demonstrated in Fig. 2.

Various measurements based on actual and predicted results are calculated, in order to evaluate the performance or accuracy of data-driven models. These include Coefficient of Variance (CV), Mean Bias Error (MBE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Squared Percentage error (MSPE), Mean Absolute Percentage Error (MAPE) and MAE (mean absolute error). CV is the variation of overall prediction error concerning actual mean values. MBE is used to determine the amount over/underestimation of predictions. MSE and MSPE is a good inductor of estimation quality. MAE determines the average value of the errors in a set of forecasts and MAPE is the percentage of error per prediction. RMSE has the same unit of actual measurements.
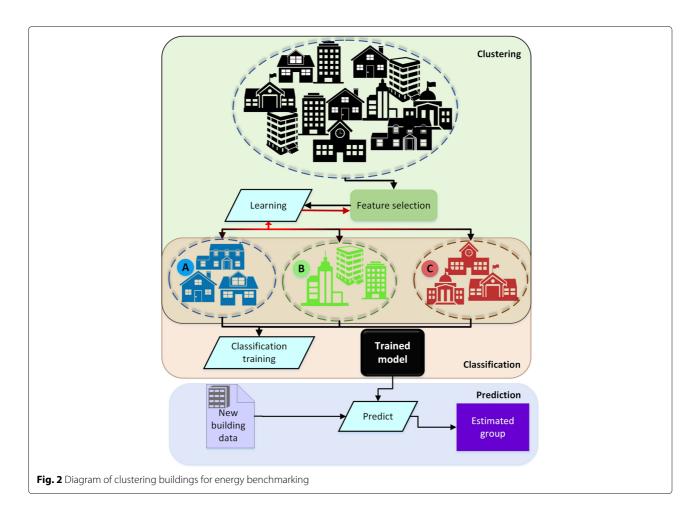
Three main techniques that have widely used in the building sector for supervised learning are ANN, SVM and Gaussian distribution regression models. K-means and hierarchical clustering methods have also utilised for unsupervised learning purposes. These methods are discussed in detail in the following sections, and a summary of other ML techniques is presented subsequently.

**Artificial neural networks**

Neural networks have been broadly utilised for building energy estimation and known as the chief ML techniques in this area. They have successfully used for modelling non-linear problems and complex systems. By applying different techniques, ANNs have the capability to be immune to the fault and noise (Tso and Yau 2007) while learning key patterns of building systems.

The main idea of the ANN is obtained from the neurobiological field. Several kinds of ANN have been proposed for different applications including, Feed Forward Network (FFN), Radial Basis Function Network (RBFN) and recurrent networks (RNN). Each ANN consists of multi-layers (minimum two layers) of neurons and activation functions that form the connections between neurons. Some frequently functions are linear, sigmoid ad hard limit functions (Park and Lek 2016).

In FFN which was the first NN model and as well the simplest one, there are no cycles from input to output neurons and the pieces of information moves in one direction

Seyedzadeh *et al. Visualization in Engineering*   (2018) 6:5

Page 5 of 20



**Fig. 2** Diagram of clustering buildings for energy benchmarking

in the network. Figure 3 illustrate a general structure of FFN with input, output and one hidden layer.

RNN uses its internal memory to learn from preceding experiences by allowing loops from output to input nodes. RNNs have been proposed in various architectures including fully connected, recursive, long short-term memory, etc. This type of neural network usually employed to solve very deep learning tasks (i.e. more than 1000 layers are needed) (Pérez-Ortiz et al. 2003; Gers and Schmidhuber 2000).

In RBFM, a radial basis function is exerted as an activation function providing a linear combination of inputs and neuron parameters as output. This type of network is very effective for prediction of time series estimation (Harpham and Dawson 2006; Leung et al. 2001; Park et al. 1998).

Based on the application and complexity of the task, a structure is decided, and by feeding the adequate amount of records, the activation function updates the weights and bias.

In building sector, ANN models have been applied for fast estimation of heating and cooling loads (Aydinalp et al. 2004; Li et al. 2009b; Alam et al. 2016), energy consumption (Karatasou et al. 2006; Hong et al. 2014a; Ferlito et al. 2015), energy efficiency (Cheng and Cao 2014; Zhang et al. 2015a; Ascione et al. 2017) and space heating (Mihalakakou et al. 2002; Aydinalp et al. 2004). Several successful application of ANN for Automated Fault Detection and Diagnostics (AFDD) in building energy conservation (Magoulès et al. 2013), solar water heater (Kalogirou et al. 2008; He et al. 2011) and HVAC system (Du et al. 2013) have been reported. ANN is also applied in building management systems to provide automatic energy consumption control (Kalogirou 2000; Benedetti et al. 2016), optimisation of heating system (Yang et al. 2003; Ahn et al. 2017) and comfort management (Yang and Wang 2013; Huang et al. 2015).

In 1995, an early study on the application of ANN in prediction of energy consumption using simple FFN model was performed to forecast electric energy usage of a building in tropical climate based on the occupancy and temperature data. Mena et al. 2014 use ANN for short-term estimation of building electricity demand. Targeting the bio-climatic stock, it is shown that outdoor temperature and solar radiation have a notable impact
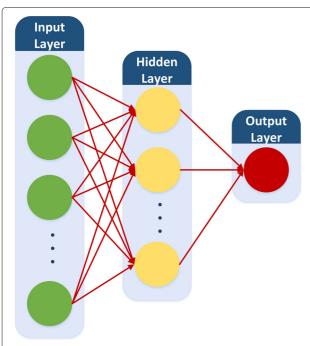
Seyedzadeh *et al. Visualization in Engineering*   (2018) 6:5

Page 6 of 20



**Fig. 3** Conceptual structure of feed forward neural network with three layers

on electricity consumption. Mihalakakou et al. 2002 used FFN and RNN for prediction of hourly electricity energy consumption in a residential building located in Athens. The models consider meteorological variables including air temperature and solar radiation using time series data gathered over six years. Gonzales & Zamarreno 2005 estimated short-term electricity energy consumption using a feedback ANN. Effect of the number of neurons in hidden layers, the best size of data windows and the ANN parameters on the accuracy of the model is investigated. Li et al. 2015 proposed an optimised ANN for prediction of hourly electricity consumption using partial swarm optimisation (PSO) algorithm. PCA is used to remove unnecessary input variables obtained from two datasets: ASHRAE Shootout I and Hanzou library building.

Platon et al. 2015 applied principal component analysis (PCA) to investigate the pre-input variables of ANN in the prediction of hourly electricity consumption of an institutional building. Results from comparison of ANN and case-based reasoning (CBR), reveals that the ANN is superior in term of accuracy. However, as CBR provides more transparency than the ANN and the capability to learn from small data, it can be an alternative approach for complex systems dependent on more variables. Li et al. 2015 proposed an optimised ANN for prediction of hourly electricity consumption using partial swarm optimisation (PSO) algorithm. PCA is used to remove unnecessary input variables obtained from two datasets: ASHRAE Shootout I and Hanzou library building.

Yalcintas (Yalcintas and Ozturk 2007; Yalcintas 2006) used ANN for energy benchmarking in tropical climate contemplate weather and chiller data. The selected building includes office, classroom, laboratory-type buildings, or mixed-use buildings. The accuracy of EUI prediction is compared with multiple linear regression methods showing a remarkable advantage over it. Hong 2014a applies ANN and statistical analysis for energy performance assessment of primary and secondary schools located in the UK by estimating electrical and heating consumption. By comparison of results with DEC benchmarks, it is shown that the ANN is more accurate for the energy assessment. It is concluded that the statistic benchmarks required further advancement and considerations (e.g. number of students and density of the schools) to provide better evaluations in this sector. However, it has been shown that ANN prediction is not as precise as simulation and engineering calculations.

Wong et al. 2010 used ANN for assessing the dynamic energy performance of a commercial building with daylighting in Hong Kong. EnergyPlus software along with algorithms for calculation of interior reflection is applied to generate the building daily energy usage. Nash–Sutcliffe Efficiency Coefficient (NSEC) is used as the primary measurement to investigate ANN accuracy in predicting cooling, heating, electric lighting and total electricity consumption.

ANN can be used for determination of parameters for energy performance assessment of buildings. Lundin et al. 2004 proposes a method for prediction of total heat loss coefficient, the total heat capacity and the gain factor that are key elements in the estimation of energy efficiency. Buratti et al. 2014 employs ANN as a tool for evaluation of building energy certificates accuracy using 6500 energy labels in Italy. The study investigates a different combination of input variables to minimise the number of training features. Using the outcome of the ANN, a new index is proposed to check the accuracy of declared data for energy certificates with a low error of 3.6%.

Hong et al. 2014 applied ANN for benchmarking of schools buildings in the UK and investigate the limitations of the assessment. An extensive database including 120000 DEC records is used for training and testing the model (Hong et al. 2014b). Reviewing outcomes of the research and comparison with bottom-up models, authors suggest the combinational use of top-down and bottom-up methods to achieve higher accuracy.

Khayatian et al. 2016 predicts energy performance certificates for residential building using an ANN model and Italian CENED database as training records. A combination set of direct and calculated features are used as inputs and heat demand indicators (derived using CENED software) as the output target of ANN.

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 7 of 20

Ascinoe et al. 2017 proposed an ANN for evaluation of energy consumption and inhabitants' thermal comfort to predict energy performance of the building. Energy assessment of the buildings are performed using Energy-Plus software, and a simulation-based sensitivity/ uncertainty analysis is proposed for further improvement of network parameters. New buildings and retrofitted stock in presence of energy retrofit measures are considered separately. For the latter case, ANN is employed for optimisation of retrofit parameters. For the first one, three single output ANN is developed to predict primary energy consumption of space heating and cooling and the ratio of yearly discomfort hours by setting whole-building parameters as network inputs (i.e. geometry, envelope, operation and HVAC). At the same time, Beccali et al. 2017 propose the use of ANN fast forecasting as a decision support tool for optimising the retrofit actions of buildings located in Italy.

Kalogirou & Bojic (Kalogirou and Bojic 2000; Kalogirou 2000) applies RNN to predict hourly energy demand of a passive solar building. ZID software has been employed to calculate the output target. Although results demonstrate high accuracy of estimation, the number of input features (season, insulation, wall thickness and time of the day) and total training records (forty simulated cases) are insufficient. Later in 2001, Kalogitrou (Kalogirou et al. 2001) applies ANN for estimating the daily heat loads of model house buildings with different calumniations of the wall (single and double) and roof (different insulations) types using a typical meteorological data for Cyprus. In this study, TRNSYS software was used as an energy evaluation engine for all cases and the data validated by comparison of one building energy consumption with the actual measurement. Karatasou et al. 2006 develops an FFN model for hourly prediction of energy loads in residential buildings. The impact of various parameters on the accuracy of a trained model is also investigated, and it is shown that parameters such as humidity and wind speed are less significant and can be eliminated from training features. Furthermore, the application of statistical analysis for enhancement of ANN model and 24 hours ahead prediction of energy consumption is demonstrated. These methods consist of hypothesis testing, information criteria and cross-validation in pre-processing and model development. However, there is less enlightenment about the main distinctions of applied FFN models. In 2010, Dombayci (Dombayci 2010) used ANN to prediction hourly energy consumption of a simple model house based on Turkish standards. The degree-hour method is applied to derive the hourly energy consumption to be used in ANN training. The models are suitable for single building energy management of simple residential buildings as it does not take many characteristics into account.

Kialashaki & Reilsel 2013 compared an ANN with MLR for estimation of the US domestic buildings energy demand. Seven independent variables (population, gross domestic product, house size, median household income, cost of residential electricity, natural gas and oil) as selected from different data sources (1984–2010) to represent the building characteristics. Antanasijevic et al. 2015 compare ANN with multiple linear and polynomial regression models for forecasting the energy consumption and energy-related greenhouse gas emission using building data from 26 European countries. The results show 4.5% improvement in term of ANN accuracy (mean absolute percentage error) in both cases.

Neto & Fiorelli 2008 compared predicted energy demand of a building in Brazil using ANN model and simulation software, EnergyPlus. The research investigates the impact of using hidden layer showing an insignificant difference in accuracy of the models. Furthermore, it reveals that external temperature is more important than humidity and solar radiation in estimating energy consumption of the study case. The authors show that ANN is more accurate that detailed simulation model, especially in short-term prediction. They conclude that improper assessment of lighting and occupancy would be the main reason for uncertainty in engineering models. Popesco et al. 2009 developed an original simulation and ANN-based models for predicting hourly heating energy demand of buildings connected to district heating system. Climate and mass flow rate variables of prior 24h are used as inputs. Deb et al. 2016 also used five previous day's data as ANN model inputs to forecast daily cooling demand of three institutional buildings in Singapore.

Olofsson & Anderson 2001 predicated daily heating consumption of six building family in Sweden constructed in the 1970's. The building went through the retrofitting in the early 1990's, and the measurements were performed before and after the renovation procedure. ANN makes an accurate long-term prediction of energy demand based on short-term measured data. PCA is also applied to reduce the number of input features to four (i.e. construction year, number of floors, framework, floor area, number of inhabitants and ventilation system). Ekici & Aksoy 2009 used back-propagation ANN to predict heating loads of three different buildings by taking climate information into account. Heating energy demand of the sample buildings is calculated using a finite difference approach of transient state one-dimensional heat conduction problem. Paudel et al. 2014 used dynamic ANN to predict heating energy consumption focusing on building occupancy profile and operational short-term heating power level characteristics.

Ben-Nakhi 2004 used a general RNN for prediction of public buildings profile of the next days using hourly energy consumption data, intending to optimise HVAC

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 8 of 20

thermal energy storage. Data from a public office building in Kuwait constructed from 1997 to 2001 is used for training and testing the ANN model. Energy consumption value of buildings is calculated using ESP-r simulation software and considering climate information, various densities of occupancy and orientation characteristics. The results show that ANN only needs external temperature for accurate prediction of cooling loads, whereas simulation software demand for intricate climate detail.

Hou et al. 2006 predicted hourly cooling loads in an air-conditioned building integrating rough set theory and ANN. Input features of ANN are determined and optimised by analyses relevant parameters to cooling load using rough set theory. The proposed model with different combinations of input sets is compared with the autoregressive integrated moving-average model all showing better accuracy. Yokoyama et al. 2009 used back-propagation ANN to predict cooling load demand by introducing a global optimisation method for the improvement of network parameters. The effect of the number of hidden layers and the number of neurons in each layer is investigated to optimise the accuracy of the proposed ANN.

Yan & Yao 2010 has proposed an investigation of the climate information effect on energy consumption in various climate zones. Back-propagation ANN is used to predict heating and cooling load to assist new building designs. Later, Biswas et al. 2016 applied the similar approach on residential sector and demonstration houses in the USA using Matlab toolbox.

Aydinalp et al. 2002 models the Appliance, Lighting and space Cooling (ALC) in residential buildings located in Canada. ANN for prediction of energy consumption shows better accuracy in comparison with engineering calculation methods. Later, they used ANN to predict Space heating and domestic hot water for the same buildings (Aydinalp et al. 2004).

Azadeh et al. (Azadeh and Sohrabkhani 2006; Azadeh et al. 2008) demonstrate the application of ANN based electricity consumption prediction model in the manufacturing industry. The model is used to predict the annual long-term consumption of industries in Iran using a multilayer perception model. The results compare with the traditional regression model using ANOVA and show superiority for the application. Later in 2014, (Kialashaki 2014 ) foretasted energy demand of the industrial sector in the US considering gross domestic and national products and population.

## Support vector machine

SVMs are highly robust models for solving non-linear problems and used in research and industry for regression and classification purposes. As SVMs can be trained with few numbers of data samples, they could be right solutions for modelling study cases with no recorded historical data. Furthermore, SVMs are based on the Structural Risk Minimisation (SRM) principle that seeks to minimise an upper bound of generalisation error consisting of the sum of training error and a confidence level. SVMs with kernel function acts as a two-layer ANN, but the number of hyper-parameters is fewer than that. Another advantage of SVM over other ML models is uniqueness and globally optimality of the generated solution, as it does not require non-linear optimisation with the risk of sucking in a local minimum limit. One main drawback of SVM is the computation time, which has the order almost equal to the cube of problem samples.

Suppose every input parameter comprises a vector $X_i$ ($i$ denotes the $i$th input component sample), and a corresponding output vector $Y_i$ that can be building heating loads, rating or energy consumption. SVM relates inputs to output parameters using the following equation:

$$Y = W \cdot \phi(X) + b \qquad (1)$$

where $\phi(X)$ function non-linearly maps $X$ to a higher dimensional feature space. The bias, $b$, is dependent of selected kernel function (e.g. $b$ can be equal to zero for Gaussian RBF). $W$ is the weight vector and approximated by empirical risk function as:

$$Minimise: \frac{1}{2}\|W\|^2 + C\frac{1}{1}\sum_{i=1}^{N} L_\varepsilon(Y_i, f(X_i)) \qquad (2)$$

$L_\varepsilon$ is $\varepsilon$-intensity loss function and defined as

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

Here $\varepsilon$ denotes the domain of $\varepsilon$-insensitivity and $N$ is the number of training samples. The loss becomes zero when the predicted value drops within the band area and gets the difference value between the predicted and radius $\varepsilon$ of the domain, in case the expected point falls out of that region. The regularised constant $C$ presents the error penalty, which is defined by the user.

SVM rejects the training samples with errors less than the predetermined $\varepsilon$. By acquisition slack variables $\xi$ and $\xi_i^*$ for calculation of the distance from the band are, Eq. (3) can be expressed as:

$$\underset{\xi,\xi_i^*,W,b}{Minmise}: \frac{1}{2}\|W\|^2 + C\frac{1}{N}\sum_{i=1}^{N} \xi + \xi_i^* \qquad (4)$$

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 9 of 20

subject to

$$\begin{cases} Y_i - W \cdot \phi(x_i) - b \leq \varepsilon + \xi \\ W \cdot \phi(x_i) + b - Y_i \leq \varepsilon + \xi_i^* \\ \xi \geq 0, \quad \xi_i^* \geq 0 \end{cases} \tag{5}$$

The SVM problem using a kernel function of $K(X_i, X_j)$ ($\alpha_i, \alpha_i^*$ as Lagrange multipliers) can be simplified as:

$$\underset{\{\alpha_i\},\{\alpha_i^*\}}{Maximise} : - \varepsilon \sum_{i=1}^{N} \left( \alpha_i^* + \alpha_i \right) + \sum_{i=1}^{N} Y_i \left( \alpha_i^* - \alpha_i \right)$$
$$- \frac{1}{2} sum_{i=1}^{N} \sum_{j=1}^{N} \left( \alpha_i^* - \alpha_i \right) \left( \alpha_j^* - \alpha_j \right) K \left( X_i, X_j \right) \tag{6}$$

subject to

$$\sum_{i=1}^{N} \left( \alpha_i^* - \alpha_i \right) = 0, \qquad 0 \leq \alpha_i, \alpha_i^* \geq C \tag{7}$$

In building sector, SVM has been used for forecasting of cooling and heading loads (Li et al. 2009a; 2009b; Hou and Lian 2009), electricity consumption (Dong et al. 2005; Xing-ping and Rui 2007), energy consumption (Lai et al. 2008; Li et al. 2010; Zhao and Magoulès 2010; Jung et al. 2015), and classification of energy usage of buildings (Li et al. 2010).

In 2005, at first in building sector SVM was applied for estimation monthly electricity usage for non-domestic building in tropical country of Singapore (Dong et al. 2005). In this study, Dong et al. considers three input parameters including temperature, humidity and solar radiation and targets four different buildings. The data is collected over three years and used for training and testing the developed model. Results of using RBF kernel indicates that SVM model has excellent accuracy in predicting the electrical loads and the low error rate of 4%. The conclusion declares the superiority of SVM over previously derived ANN models in terms of selection of small model parameters and accuracy. This initial worked was followed by Lai et al. 2009a applying SVM for forecasting monthly and short-term (i.e. daily) prediction of electricity consumption of a domestic building located in Japan. They used outdoor, living and bedroom temperature and humidity as well as water temperature as input parameters and collected electricity usage data over a year. Massana et al. 2015 compare SVM, ANN and MLR in short-term prediction of non-domestic buildings' electricity demand and conclude that SVM provide higher accuracy and lower computational cost.

Later in 2010, Li et al. 2010 used SVM for long-term prediction (yearly) of electricity consumption of domestic buildings. They consider fifteen building envelope parameters collected from 59 different cases along with the annual electricity consumption which is normalised by unit area. Besides, they compare the accuracy of the SVM model with three types of ANNs including propagation, RFB and general regression. Testing the trained model over 20% of study cases provides results that show SVM outperforms ANNs for all samples. Solomon et al. 2011 predict weekly electricity consumption of a massive commercial building considering previous electricity usage, temperature data and wind velocity.

In addition, Li et al. 2009b apply SVM to forecast hourly cooling leads of an office building located in China. They use three similar input parameters which were used by Dong et al. 2005 and collected from local climate database. The target samples are gathered during summer and one month used for training and four months for testing the model. In the meantime, they present a comparison with ANN models and indicate that SVM and general regression ANN have more potential to be used in the field. Hou & Lian (Hou and Lian 2009) examine the accuracy of SVM with an autoregressive integrated moving average based model (MacArthur et al. 1989) and demonstrate the supremacy of SVM regarding maximum and minimum error values. Xuemei et al. 2009 developed a model based on Least Square SVM (LS-SVM) and used the same input parameters. This approach contributes to learning correction for limited training sets and enhanced prediction time efficiency to traditional SVM model in load forecasting. Jinhu et al. 2010 and Li et al. 2010 apply improved PCA to find the significant parameters and show better accuracy. However, the information about original and selected features are missing. The further improvement of similar SVM based cooling load prediction has been demonstrated using a fuzzy C-mean algorithm for clustering samples (Xuemei et al. 2010), simulated annealing particle swarm optimisation to prevent premature convergence (Li et al. 2010) and Markov chains to the farther forecast of the interval after primitive prediction (Zhang and Qi 2009).

Zhao & Magoules 2010 predicted energy consumption of office building using parallel implementation of SVM. They aim to optimise the building characteristics of a model case. They utilised EnergyPlus software to calculate the energy demands. The results show a slight improvement regarding accuracy. Later in 2012, the authors apply gradient guided feature selection and the correlation coefficients methods to decrease the number of features for RBF and polynomial based SVM models (Zhao and Magoulès 2012a).

In 2014, Jain et al. 2014 used sensor-based data of multi-family domestic building located in New York City to develop an SVM model. The aim is to investigate the effect of a different time interval and building spaces of data collection on energy consumption forecasting. The authors point out that the optimum efficiency of the derived

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 10 of 20

model is obtained when hourly intervals collected at floor level is utilised. Edwards et al. 2012 present a comparison of SVM, LS-SVM and ANN in forecasting hourly energy consumption of small residential buildings and find ANN as the least accurate model.

**Gaussian process and mixture models**

Since early 2000, Gaussian process (GP) regression has been employed by researchers in different application (Jiang et al. 2010; Grosicki et al. 2005; Bukkapatnam and Cheng 2010). In building energy field, GP has been recently utilised due to its potentiality in determining the uncertainty of predictions. In building energy modelling, there are usually uncertainties in the section of appropriate values for some characteristics (e.g. envelope insulation). Hence, evaluation of input uncertainty on foretasted results has made the GP as an alternative approach to model building energy rather than conventional and other ML regression models. The main drawback of GP modelling is expensive computational cost, especially with the increase of training samples. This high cost is due to the fact that GP constructs a model by determining the structure of a covariance matrix composed of $N \times N$ input variable where the matrix inversion required in predictions has a complexity of $O\left(N^3\right)$

Given a set of $n$ independent input vector $X_j$ ($j = 1, \cdots, n$), the corresponding observations of $y_i$ ($i = 1, \cdots, n$) are correlated using covariance function $K$ with normal distribution equal to (Li et al. 2014):

$$
\begin{aligned}
P(y; m; k) = & \frac{1}{(2\pi)^{n/2} |K(X,X)|^{1/2}} \\
& \times exp\left(-\frac{1}{2}(y-m)^T K(X,X)^{-1}(y-m)\right)
\end{aligned}
\tag{8}
$$

The covariance or kernel function can be derived as

$$
K = \begin{vmatrix}
k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\
k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\
\vdots & \vdots & \ddots & \vdots \\
k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n)
\end{vmatrix}
\tag{9}
$$

A white noise, $\sigma$, is presumed in order to consider the uncertainty. It is assumed that the samples are corrupted (lets suppose as new inputs as $x^*$) by this noise. In this case covariance of $y$ is expressed as

$$
cov(y) = K(X,X) + \sigma^2
\tag{10}
$$

Then $y^*$ can be estimated as below.

$$
y^* = \sum_{i=1}^{n} \alpha_i k(x_i, x^*)
\tag{11}
$$

$$
\alpha_i = \left(K(X,X) + \sigma^2 I\right)^{-1} y_i
\tag{12}
$$

A Gaussian Mixture Model (GMM) is parametric probability density (PDF) function which is expressed as (Reynolds 2015):

$$
p(x, y) = \sum_{k=1}^{K} \pi_k p\left(x, y | \mu_k, \sum_k\right)
\tag{13}
$$

here $\pi_k p\left(x, y | \mu_k, \sum_k\right)$ is PDF of $K$ Gaussian components and $\mu_k$ is the mean function of $k$th component. For regression proposes the multivariate non-linear function from the model is derived. Indeed, Gaussian mixture regression constructs a series of Gaussian mixture to unite the density of data and calculate regression function for each model as presented in Eq. 13 .

Heo (Heo et al. 2012; Heo and Zavala 2012) applies GP model to calculate the building energy saving after retrofitting by forecasting the total energy consumption. The model uses outside temperature, relative humidity, and occupancy count as an input variable and considers output measurement errors to approximate uncertainty levels. Later in 2013, Zhang et al. 2013 use GP regression for predicting the energy demand of an office building cooling and heating in the post-retrofit phase. They show that the accuracy of the GP model is very dependant on training and testing data range.

Noh & Rajagopal 2013 propose a long-term GP prediction model for total energy consumption of a campus building using smart meter measurements and weather data. Nghiem & Jones 2017 propose a GP based model for demand response service by predicting building energy consumption. Rastogi et al. 2017 compare the accuracy of GP and linear regression in emulating of a building performance simulation and show that the accuracy of GP is four times better than linear regression testing on EnergyPlus simulated case studies located in the US.

Burkhart et al. 2014 integrate GP with a Monte Carlo expectation maximisation algorithm to train the model under data uncertainty. The aim is to optimise office building HVAC system performance by predicting its daily energy demand. Relative humidity and ambient temperature are considered as specific input variables and daily occupancy with two different scenarios (moderate and vigorous) as uncertain data. The results indicate that the models can be trained even with limited data or sparse measurements employing rough approximation and data range instead of sensor data.

Manfren et al. 2013 develop a method for calibration and uncertainty analysis of building energy simulation

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 11 of 20

model. They used detailed simulation, GP with RFB kernel and MLR to predict monthly electricity and gas usage of heating and cooling systems. The results indicate that GP not only provides a tool for optimisation and uncertainty analysis of building energy models but also shows higher accuracy in comparison with a piece-wise regression model.

Sirvastav et al. 2013 employ GMM to predicts daily/hourly energy consumption of commercial buildings (a DOE reference model for supermarket and a retail store building). This parametrised model allows locally adaptive uncertainty quantification for building data.

Zhang et al. 2015b compare change-point models, GP, GMM and FF-ANN models for prediction of an office building's HVAC system hot water energy usage considering weather data (ambient dry bulb temperature) as an input variable. The ANN utilised in this work has one hidden layer activated using tangent sigmoid transfer function. The results show that the best performance is achieved using GMM and the worst by ANN. The authors conclude that as the ANN is not fed by adequate data, it is not a suitable model for the case study. Although the accuracy of GMM and GP is slightly better than the change-point regression, the later is recommended due to the simplicity of the approach. It should be noted that the Gaussian methods are the best choice for analysing uncertainty and capturing complex building behaviour.

### Clustering algorithms

Clustering is one of the well-known ML techniques that identifies implicit relations, patterns and distributions in data sets. Clustering is an unsupervised learning method that can describe the hidden structure in a collection of unlabeled data. In building energy, the primary application of this technique is to classify buildings using various features and characteristics instead of only use type or topology is very advantageous in building energy benchmarking. Clustering for such an application implicates four steps (Gao and Malkawi 2014): (a) data collections, (b) feature identification and selection, (c) adaptation of appropriate clustering algorithm and (d) benchmarking each building within classified groups. The most common clustering algorithm is k-means that iteratively seeks for a local maximum. The algorithm begins with a random selection of *k* centroids (centre of cluster), and each data is assigned to the nearest centre point. Then all centroids are recalculated using the mean of all data points in a group. This process continues until it satisfies a stopping criterion (e.g. a minimum aggregation of distances is reached).

Targeting 320 schools in Greece, Santamouris et al. 2007 propose a building energy classification method using fuzzy clustering (Gath and Geva 1989). Total energy consumption (heating and electricity) over three years along with information on operating hours, number of pupils, structure characteristics, etc., are collected. By applying a clustering algorithm, five building energy rating classes are determined. The clustering based classification is then compared to similar frequency rating process indicating that clustering offers more robust classes resolving the problem of low and unbalanced or very large class constitution. The authors apply outcomes to ten study cases to investigate the potential energy conservation. Gaitani et al. 2010 use 1100 school samples for the development of a framework for heating energy consumption rating, aiming at evaluation of potential energy savings. A k-mean clustering incorporating PCA algorithm is utilised to form five rating classes and determine representative building of each cluster. Pieri et al. 2015 propose a cluster-based energy audit considering cooling and heating loads of hotels in Greece.

Gao & Malkawi 2014 demonstrate that energy performance benchmarking using clustering algorithm is more accurate and robust than the US Energy Star scheme due to the ability in integrating all the building features that affect energy consumption. The feature extraction is made using ordinary least squares regression and clusters are generated using the k-means algorithm. Lara et al. 2015 also apply k-means clustering to assess the energy performance of schools in Italy and characterise reference building for each group. First an MLR method, as a mean of correlation analysis, is used to identify the most appropriate quantities and variables for representation of energy demand and building properties. Then clustering algorithm cluster similar buildings regarding the defined variables. Finally, the building having the minimum distance from the centroid is selected as the representative for each cluster. These reference buildings are useful tools for optimising retrofit solutions.

Yu et al. 2011 use clustering technique to demonstrate the impact of occupancy behaviour in building energy consumption. A similarity of building features unrelated to occupants behaviour is used for creating clusters, and the impact of users action in energy demand is investigated for each cluster. Petcharat et al. 2012 propose a clustering algorithm to asses potential energy saving regarded to the lighting system in Thailand non-domestic stock. The authors indicate that cluster-based analysis is more effective than the only comparison of target building power density with reference cases that are defined by the country's Energy Act.

Yang et al. 2017 apply a k-shape (proposed for clustering time series) algorithm to identify energy usage patterns and then employ SVM for enhancing the accuracy of building energy demand prediction. Jalori & Reddy 2015 propose clustering of days based on daily/hourly energy consumptions to detect and remove outlier data point. This process further improves data-driven energy

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 12 of 20

**Table 1** Summary of machine learning techniques for prediction of building energy consumption and performance

| Target | ML | Prediction term | Building case and data | Features | Ref |
|---|---|---|---|---|---|
| Energy Performance | ANN | Month | schools in England and Wales (120,253 DEC records) | Construction year, Phase of education, Number of pupils, Internal environmental conditioning, Site exposure, Orientation, North facade adjacency, South facade adjacency, East facade adjacency, West facade adjacency, Floor area, Building depth ratio, Compactness ratio, Surface exposure ratio, North glazing ratio, South glazing ratio, East glazing ratio, West glazing ratio, Glazing type, Roof shape, Roof glazing, Heating degree-days, Cooling degree-days | Hong et al. (2014a) |
| | ANN | - | Educational building (previous preliminary energy assessments (PEA) reports for over 60 buildings in Hawaii) | Operation hours, Age, Square feet area, Yearly electricity usage, Percentage electricity used for lighting, air conditioning, plug loads | Yalcintas (2006) |
| | ANN | Year | Office buildings in Italy (8800 building stock simulated using EnergyPlus) | geometry(9), envelope(30), operation (6) and HVAC (3) | Ascione et al. (2017) |
| | ANN | Year | Schools in UK (120,253 DEC records) | North glazing ratio, South glazing ratio, East glazing ratio, West glazing ratio, Glazing type, Roof shape, Roof glazing, Heating degree days, Cooling degree days | Hong et al. (2014) |
| | ANN | - | Residential buildings (the online CENED database) | Degree days, Net volume, Net floor area, Dispersant surface, Opaque to glazed ratio, Year of construction, Thermal conductivity, Average floor height, Opaque surface area, Glazed surface area, Construction period, Non-linear features | Khayatian et al. (2016) |
| | ANN | Day | An generic reference office building in Hong Kong (8760 hourly records calculated using EnergyPlus) | External weather conditions (daily average dry-bulb temperature, daily average wet-bulb tempera-ture, daily global solar radiation and daily average clearness index), Building envelope designs (solar aperture, daylight aperture, overhang and side-fins projections), Day type | Wong et al. (2010) |
| | Clustering | - | 5215 commercial building samples (CBECS database) | Area, Percent heated, Percent cooled, Wall materials, Roof materials, Window materials, Window percent, Shape, Number of floors, Construction year, Weekly operation hours, Occupants, Variable air volume, Heating unit, Cooling unit, Economizer, Refrigerators, Number of servers, Office equipment, Heating and cooling degree day | Gao and Malkawi (2014) |
| | Clustering | - | 1100 school in Greece (data gathered over one) | Heated surfac, Age of the building, Insulation of the building, Number of classrooms, Number of students, School's operating hours per day, Age of the heating system, Energy consumption per unit | Gaitani et al. (2010) |
| | Clustering | - | 320 schools in Greece (Energy data have been collected for a three years) | Temperature, Solar radiation, Energy consumption per unit, Operational period, Number of students, Construction characteristics, Installed equipment | Santamouris et al. (2007) |
| | Clustering | - | 60 schools in Italy (data collected over 5 years) | Area of the floor in thermal contact with the ground, Opaque envelope area, Transparent envelope area, Windows tp vertical walls ratio, Windows to floor area ratio, Transparent to opaque envelope ratio, Envelope average thermal transmittance, Shape, Heating system Capacity | Arambula Lara et al. (2015) |

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 13 of 20

**Table 1** Summary of machine learning techniques for prediction of building energy consumption and performance (*Continued*)

| Target | ML | Prediction term | Building case and data | Features | Ref |
|---|---|---|---|---|---|
| HVAC Loads | GPR | Day | An office building in Chicago (Loads calculated using simulation) | Weather, Occupancy count | Burkhart et al. (2014) |
| | GPR, GMM, ANN | Day | Office building (three months data collected) | Outside dry bulb air temperature, Day | Zhang et al. (2015b) |
| Heating & Cooling Loads | ANN | Year | Model house with 9 combination of wall and roof type (loads are calculated using TRNSYS simulation) | Wall and Roof type, Maximum and mean daily direct and global radiation, Maximum and mean temperature of the day , Mean wind speed and direction (degrees) | Kalogirou et al. (2001) |
| | GPR | Hour | Office building in Philadelphia | Outdoor temperature | Zhang et al. (2013) |
| | GPR | Year | Typical buildings in the US (loads calculated using EnergyPlus) | Building characteristics, Climate data (28 features) | Rastogi et al. (2017) |
| | GPR | Month | Retrofitted office building (Actual measurements and simulation) | Building envelope characteristics, Solar shading control system | Manfren et al. (2013) |
| Heating Loads | ANN | Hour | Simulation models (Data collected from a District Heating Company of the city of Iasi) | Solar radiation, Wind speed, Outside temperature of previous 24h, Mass flow rate of hot water of previous 24h, Hot water temperature exit from plant system | Popescu et al. (2009) |
| | ANN | Hour | Schools in UK (120,253 DEC records) | Glazing ratio in all cardinal directions, Roof shape and glazing, Heating and cooling degree days | Hong et al. (2014) |
| | ANN | Day | Six single-family buildings, constructed in Stockholm (The measurements performed before and after retrofitting) | Construction year, Stories, Framework, Floor area, Number of inhabitants, Ventilation system | Olofsson and Andersson (2001) |
| | ANN | Hour | An institutional building in Nantes (The data is taken from data acquisition system for 1.5 months) | Climate and heating energy data, Occupancy profile | Paudel et al. (2014) |
| | SVM | Day | Single-story mass-built buildings (Simulated using EnergyPlus) | Outdoor dry bulb and relative humidity, Wind speed, Direct solar, Ground temperature, Outdoor air density, Water mains temperature, Number of occupants, Total heat gain of lights, electric equipment and window, Heat loss for walls, Mean air temperature, Infiltration volume, Heating outlet temp | Zhao and Magoulès (2012a) |
| | ANN | Month | Three sample buildings (Heating loads demand calculated using finite difference approach of transient state) | Transparency ratio, Insulation thickness, Building form factors | Ekici and Aksoy (2009) |
| Cooling Loads | ANN | Hour | Parking space (data gathers over 23 weekdays) | Temperature, Relative humidity | Yokoyama et al. (2009) |

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 14 of 20

**Table 1** Summary of machine learning techniques for prediction of building energy consumption and performance (*Continued*)

| Target | ML | Prediction term | Building case and data | Features | Ref |
|---|---|---|---|---|---|
| | ANN | Day | Public office building in Kuwait (data for three building types) | External temperature | Ben-Nakhi and Mahmoud (2004) |
| | ANN, SVM | Hour | A model building in China (measurements from an existing HVAC system) | Temperature, Relative humidity | Hou et al. (2006); Hou and Lian (2009); Xuemei et al. (2009) |
| | ANN | Day | Three institutional buildings (The energy data is obtained through the facility management office) | Five previous day | Deb et al. (2016) |
| | GPR | Hour | office building in Lemont city (data obtained from baselining and post-retrofit days) | Outdoor temperature | Heo et al. (2012) |
| | GPR | Day | An School building in Stanford city (data obtained from baselining and post-retrofit days) | Outdoor temperature | Noh and Rajagopal (2013) |
| Energy Demand | ANN | Hour | holiday home which is used only during weekends (forty cases generated by the program ZID) | Season, Insulation, Wall thickness, Time of day, Energy calculating function | Kalogirou and Bojic (2000) |
| | ANN | Hour | Two datasets (Great Building Energy Predictor Shootout I (5 months), office building located in Athens, Greece (one year)) | Temperature, Solar radiation, Humidity ratio, Wind speed, Day | Karatasou et al. (2006) |
| | ANN | Year | the US domestic buildings (energy consumption is taken from U.S. Energy Information Administration) | Population, Gross domestic product, House size, Median household income, Cost of residential electricity, Natural gas and oil | Kialashaki and Reisel (2013) |
| | ANN | Day | An office building in University of Sao Paulo | Daily maximum and minimum external dry-bulb temperatures | Neto and Fiorelli (2008) |
| | SVM | Month | Four office buildings in Singapore (energy consumption is obtained from utility bills over 4 years) | Dry bulb temperature, Relative humidity, Global solar radiation | Dong et al. (2005) |
| | SVM | Hour | Multi-family domestic building in New York City (data from the Great Energy Predictor Shootout) | Temperature, Humidity, Wind speed | Jain et al. (2014) |
| Electricity Demand | ANN | Hour | An institutional facility in Calgary (data collected over 15 month) | Outside temperature and relative humidity, Boiler outlet water temperature and flow-rate, Chiller outlet water temperature and flow-rate, Supply air temperatures for hot, cold duct, Supply and return control settings, Indoor air temperatures of 2 different zones | Platon et al. (2015) |
| | ANN | Hour | A building in Athens (time series of hourly values are collected over 6 years) | Air temperature, Solar radiation | Mihalakakou et al. (2002) |
| | SVM, ANN | Month, Day | A residential building in Japan (data is collected over one year) | Date, Outdoor temperature, Bedroom temperature, Living temperature, Living humidity, Bedroom humidity, Outdoor humidity, Water temperature | Li et al. (2009a) |

Seyedzadeh *et al. Visualization in Engineering*   (2018) 6:5

Page 15 of 20

**Table 1** Summary of machine learning techniques for prediction of building energy consumption and performance (*Continued*)

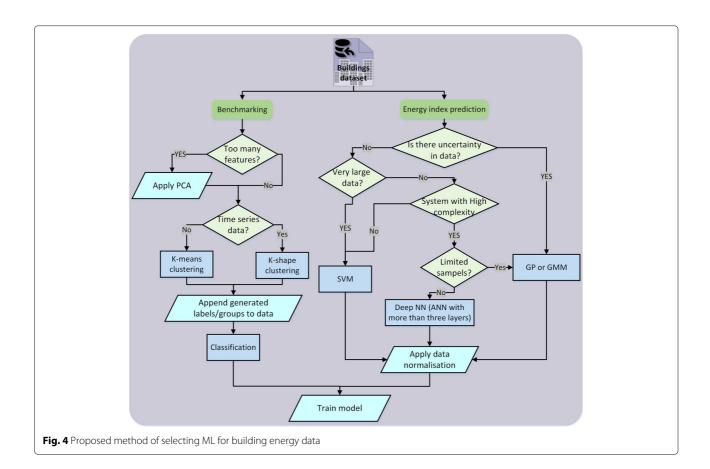| Target | ML | Prediction term | Building case and data | Features | Ref |
|---|---|---|---|---|---|
| | SVM, ANN | Year | 59 residential buildings in China | Mean heat transfer coefficient of building walls, Mean thermal inert index of building walls, Roof heat transfer coefficient, Building size coefficient, Absorption coefficient for solar radiation of exterior walls, Window to wall ratio in four directions, Mean window to wall ratio, Shading coefficient of window in four directions, Integrated shading coefficient | Li et al. (2010) |
| | SVM, ANN | Hour | A university office building (electrical load data is collected with a power meter) | Outdoor/indoor temperature and humidity, Indoor illumination, Solar radiation, Calendar nominal attributes | Massana et al. (2015) Li et al. (2010) |
| | GMM | Day | DoE super market reference model (climate data from Chicago) | Outside dry-bulb air temperature and humidity ratio, Direct solar radiation | Srivastav et al. (2013) |

forecasting models, and so increases the performance of BMS.

## Summary of ML models

A summary of ML approaches based on the application is given in Table 1. The table provides information on prediction duration, the building study cases and data or energy usage collection and features used in model training.

Based on the results from seminal works and proposed methods for different applications and considering some ML factors, we propose a framework for selecting the right method for building energy prediction and benchmarking as demonstrated in Fig. 4.



**Fig. 4** Proposed method of selecting ML for building energy data

## Conclusion

In recent years, optimisation of construction and building energy usage have been received considerable attention as this sector is known as the main contributor to air pollution and fossil energy consumption. The regulations and rising fuel prices have forced owners to reduce energy use using smart controls, sensors or retrofitting. This concern has become more critical in the non-domestic sector as a massive amount of energy is wasted due to inefficient management. As a result, various smart technologies have been applied for the purpose of energy saving. The rapid development of the modern technologies including sensors, information, wireless transmission, network communication, cloud computing, and smart devices has been led to an enormous amount of data accumulation. The traditional modelling of building energy using software and statistical approaches does not satisfy the demand for fast and accurate forecasting, which is essential for decision-making systems. ML models have shown great potential as an alternative solution for energy modelling and assessment for different types of buildings. This paper presented a review of ML models utilised for building energy forecasting and benchmarking indicating the advantages and drawbacks of each model. Moreover, several pre-processing techniques applied to models to enhance prediction accuracy were well discussed.

ANN has been broadly used in building energy forecasting since the first introduction in the sector at 1990's. ANNs provide a powerful tool for modelling building energy modelling and reliable prediction. However, they require a proper choice of network structure and precise adjustment of its several hyper-parameters for training. The performance of the models is not guaranteed as ANN suffer from a local minimum problem. Results from different researchers indicate that ANN should be fed with adequate number of samples in order to obtain acceptable accuracy. Otherwise it might be outperformed with simple MLR models. It could be concluded that ANN is much appropriate for engineers having a strong knowledge of deep learning and statistical modelling.

In contrast with ANN, SVM and GP are supervised using few parameters and provide satisfactory performance. It has been shown that SVM surpasses ANN in load forecasting and has the potential to build models from limited samples. Nevertheless, the ANNs used for comparison in the studies as mentioned earlier, exploit simple structure, and the hyper-parameters might not be well optimised due to the complexity. Among ML techniques and other black box methods, only GP is used for model training with uncertainty assessment. Nevertheless, it is not the solely capable technique. Recently, uncertainty and sensitivity analysis for other ML models has been introduced and utilised. Hence, it worth to

devote research attention to deploy these approaches for modelling building under uncertain data.

In general, it is challenging to conclude that which ML model is the best, as from literature it can be induced that all models provide reasonable accuracy by supplying large samples and optimising the hyper-parameters. Thereby, it is imperative to thoroughly analyse the nature of available or collectable data and the application, to choose the most suitable model. For example, ANN provides a fast and precise short-term load forecasting for EMSs where temperature and humidity data is collected using sensors, while GP is more beneficial for long-term energy estimation when there is uncertainty in input variables. In fact, feature selection itself require an extensive investigation for each application as it is the preliminary requisite for implementation of any ML model.

Another issue with seminal literature is that there has not been a fair comparison of different ML models. As discusses before, there are several researches that compare the proposed ML method with conventional regression models or another simple ML model without providing sufficient detail of the structure. Hence, a thorough investigation of these techniques by tuning models is recommended, which will ease decision making for expert selecting MLs for energy forecasting.

Apart from modelling building energy, clustering buildings based on various input parameters remarkably facilitates and enhances energy benchmarking procedure. Smartly determination of reference buildings leads to more precise energy labelling, comparing with a traditional definition of notional buildings. Moreover, a combination of clustering with classification allows estimating the reference building for future cases. This area has not been studied thoroughly and seems to be a trending topic in the near future as the global concern about energy is increases and many countries put efforts to regulate the energy consumer industries especially buildings and construction.

The global warming issue raised by greenhouse gasses emission is getting more attention every year. Modern technologies such as Big Data and Internet of Things find their place in building energy applications where large data from sensors and energy meters need highly efficient data processing systems. It is clear that traditional methods of energy modelling and forecasting won't be able to accompany the novel data mining development. Consequently, intelligent models are required in industry to answer this demand, and further investigation of AI application in building sector focusing on industrial data seems to be essential.

## Abbreviations

AFDD: Automated fault detection and diagnostics; ANN: Artificial neural network; *b*: Kernel bias; BEER: Building energy efficiency retrofit; BPS: Building performance simulation; CBR: Case-based reasoning; CO2: Carbon dioxide; CV:

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 17 of 20

Coefficient of variance; DEC: Display energy certificate; EMS: Energy management system; FFN: Feed forward network; GMM: Gaussian Mixture Model; GP: Gaussian process; HVAC: Heating, ventilating, and air conditioning; *K*: Covariance function; MAE: Mean absolute error; MAPE: Mean absolute percentage error; MBE: Mean bias error; ML: Machine learning; MLR: Multivariate regression; MOO: Multi-objective optimisation; MSE: Mean squared error; MSPE: Mean squared percentage error; *O*: Algorithm complexity order; PCA: Principal component analysis; PDF: Probability distribution function; PSO: Partial swarm optimisation; RBFN: Radial basis function network; RMSE: Root MSE; RNN: Recurrent network; SVM: Support vector machine; *W*: Weight vector; ZEB: Zero energy building; $\gamma$: SVM rejection threshold; $\xi$: SVM slack variable; $\varepsilon$: Domain of epsilon insensitivity; $\alpha$: Lagrange multiplier for kernel function; $\sigma$: White noise; $\mu$: Mean function

### Availability of data and materials
We have not used any data in our study.

### Authors' contributions
All authors contributed extensively to the work presented in this paper. SS led the entire process of this study. All authors read and approved the final manuscript.

### Ethics approval and consent to participate
It is confirmed that there has not been any human participation or data involved in our study.

### Consent for publication
We have not used any personal data in any form in preparing the manuscript.

### Competing interests
The authors declare that they have no competing interests.

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

### Author details
[1]Faculty of Engineering, University of Strathclyde, Glasgow G1 1XW, UK. [2]Faculty of Engineering & Environment, Northumbria University, Newcastle, UK. [3]Faculty of Computer and Information Sciences, University of Strathclyde, Glasgow, UK.

### References
Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology*, *27*(4), 265–276. https://doi.org/10.1016/j.jenvp.2007.08.002.

Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F, Abdullah, H., Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, *33*, 102–109. https://doi.org/10.1016/j.rser.2014.01.069.

Ahn, J., Cho, S., Chung, D.H. (2017). Analysis of energy and control efficiencies of fuzzy logic and artificial neural network technologies in the heating energy supply system responding to the changes of user demands. *Applied Energy*, *190*, 222–231. https://doi.org/10.1016/j.apenergy.2016.12.155.

Alam, A.G., Baek, C.I., Han, H. (2016). Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments. *Applied Mechanics and Materials*, *819*, 541–545. https://doi.org/10.4028/www.scientific.net/AMM.819.541.

Antanasijević, D., Pocajt, V., Ristić, M., Perić-Grujić, A. (2015). Modeling of energy consumption and related GHG (greenhouse gas) intensity and emissions in Europe using general regression neural networks. *Energy*, *84*, 816–824. https://doi.org/10.1016/j.energy.2015.03.060.

Arambula Lara, R., Pernigotto, G., Cappelletti, F., Romagnoni, P., Gasparella, A. (2015). Energy audit of schools by means of cluster analysis. *Energy and Buildings*, *95*, 160–171. https://doi.org/10.1016/j.enbuild.2015.03.036.

Ascione, F., Bianco, N., De Stasio, C., Mauro, G.M., Vanoli, G.P. (2014). A new methodology for cost-optimal analysis by means of the multi-objective optimization of building energy performance. *Energy and Buildings*, *88*, 78–90. https://doi.org/10.1016/j.enbuild.2014.11.058.

Ascione, F., Bianco, N., De Stasio, C., Mauro, G.M., Vanoli, G.P. (2016). Multi-stage and multi-objective optimization for energy retrofitting a developed hospital reference building: A new approach to assess cost-optimality. *Applied Energy*, *174*, 37–68. https://doi.org/10.1016/j.apenergy.2016.04.078.

Ascione, F., Bianco, N., De Stasio, C., Mauro, G.M., Vanoli, G.P. (2017). Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy*, *118*, 999–1017. https://doi.org/10.1016/j.energy.2016.10.126.

Aydinalp, M., Ugursal, V.I., Fung, A.S. (2002). Modeling of the appliance, lighting and space-cooling energy consumption in the residential sector using neural networks. *Applied Energy*, *71*(2), 87–110. https://doi.org/10.1016/j.apenergy.2003.12.006.

Aydinalp, M., Ugursal, V.I., Fung, A.S. (2004). Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, *79*(2), 159–178. https://doi.org/10.1016/j.apenergy.2003.12.006.

Azadeh, A., Ghaderi, S.F., Sohrabkhani, S. (2008). Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management*, *49*(8), 2272–2278. https://doi.org/10.1016/j.enconman.2008.01.035.

Azadeh, M.A., & Sohrabkhani, S. (2006). Annual electricity consumption forecasting with Neural Network in high energy consuming industrial sectors of Iran, vol. 49, In *Proceedings of the ieee international conference on industrial technology*. https://doi.org/10.1109/ICIT.2006.372572 (pp. 2166–2171). Pergamon: IEEE.

Beccali, M., Ciulla, G., Lo Brano, V., Galatioto, A., Bonomolo, M. (2017). Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the nonresidential building stock in Southern Italy. *Energy*, *137*, 1201–1218. https://doi.org/10.1016/j.energy.2017.05.200.

Bell, M. (2004). Energy Efficiency in Existing Buildings: the Role of Building Regulations, In *Cobra 2004 proc. of the rics foundation construction and building research conference*. Retrieved from http://www.leedsbeckett.ac.uk/as/cebe/projects/cobra04-1.pdf (p. 16).

Benedetti, M., Cesarotti, V., Introna, V., Serranti, J. (2016). Energy consumption control automation using Artificial Neural Networks and adaptive algorithms: Proposal of a new methodology and case study. *Applied Energy*, *165*, 60–71. https://doi.org/10.1016/j.apenergy.2015.12.066.

Ben-Nakhi, A.E., & Mahmoud, M.A. (2004). Cooling load prediction for buildings using general regression neural networks. *Energy Conversion and Management*, *45*(13–14), 2127–2141. https://doi.org/10.1016/j.enconman.2003.10.009.

Biswas, M.R., Robinson, M.D., Fumo, M.D. (2016). Prediction of residential building energy consumption: A neural network approach. *Energy*, *117*, 84–92. https://doi.org/10.1016/j.energy.2016.10.066.

Bukkapatnam, S.T., & Cheng, C. (2010). Forecasting the evolution of nonlinear and nonstationary systems using recurrencebased local Gaussian process models. *Physical Review E Statistical, Nonlinear, and Soft Matter Physics*, *82*(5), 56206. https://doi.org/10.1103/PhysRevE.82.056206.

Buratti, C., Barbanera, M., Palladino, D. (2014). An original tool for checking energy performance and certification of buildings by means of Artificial Neural Networks. *Applied Energy*, *120*, 125–132. https://doi.org/10.1016/j.apenergy.2014.01.053.

Burkhart, M.C., Heo, Y., Zavala, V.M. (2014). Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. *Energy and Buildings*, *75*, 189–198. https://doi.org/10.1016/j.enbuild.2014.01.048.

Cheng, M.-Y., & Cao, M.-T. (2014). Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Applied Soft Computing*, *22*, 178–188. https://doi.org/10.1016/j.asoc.2014.05.015.

Chung, W (2011). Review of building energy-use performance benchmarking methodologies. *Applied Energy*, *88*(5), 1470–1479. https://doi.org/10.1016/j.apenergy.2010.11.022.

Crawley, D.B., Lawrie, L.K., Winkelmann, F.C., Buhl, W.F., Huang, Y.J., Pedersen, C.O., Strand, R.K., Liesen, R.J., Fisher, D.E., Witte, M.J., Glazer, J. (2001). EnergyPlus: Creating a newgeneration building energy simulation

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 18 of 20

program. *Energy and Buildings*, *33*(4), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6.

Deb, C., Eang, L.S., Yang, J., Santamouris, M. (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*, *121*, 284–297. https://doi.org/10.1016/j.enbuild.2015.12.050.

Dombayci, Ö.A. (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. https://doi.org/10.1016/j.advengsoft.2009.09.012.

Dong, B., Cao, C., Lee, S.E. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, *37*(5), 545–553. https://doi.org/10.1016/j.enbuild.2004.09.009.

Dounis, A.I., & Caraiscos, C. (2009). Advanced control systems engineering for energy and comfort management in a building environment A review. *Renewable and Sustainable Energy Reviews*, *13*(6), 1246–1261. https://doi.org/10.1016/j.rser.2008.09.015.

Du, Z., Fan, B., Jin, X., Chi, J. (2013). Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Building and Environment*, *73*, 1–11. https://doi.org/10.1016/j.buildenv.2013.11.021.

Edwards, R.E., New, J., Parker, L.E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, *49*, 591–603. https://doi.org/10.1016/j.enbuild.2012.03.010.

Ekici, B.B., & Aksoy, U.T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, *40*(5), 356–362. https://doi.org/10.1016/j.advengsoft.2008.05.003.

Ferlito, S., Atrigna, M., Graditi, G., De Vito, S., Salvato, M., Buonanno, A., Di Francia, G. (2015). Predictive models for building's energy consumption: An Artificial Neural Network (ANN) approach, In *2015 xviii aisem annual conference*. https://doi.org/10.1109/AISEM.2015.7066836 (pp. 1–4).

Filippin, C. (2000). Benchmarking the energy efficiency and greenhouse gases emissions of school buildings in central Argentina. *Building and Environment*, *35*(5), 407–414. https://doi.org/10.1016/S0360-1323(99)00035-9.

Gaitani, N., Lehmann, C., Santamouris, M., Mihalakakou, M., Patargias, P. (2010). Using principal component and cluster analysis in the heating evaluation of the school building sector. *Applied Energy*, *87*(6), 2079–2086. https://doi.org/10.1016/j.apenergy.2009.12.007.

Gao, X., & Malkawi, A. (2014). A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. *Energy and Buildings*, *84*, 607–616. https://doi.org/10.1016/j.enbuild.2014.08.030.

Gath, I., & Geva, A. (1989). Unsupervised optimal fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, *11*(7), 773–780. https://doi.org/10.1109/34.192473.

Gers, F., & Schmidhuber, J. (2000). Recurrent nets that time and count, vol. 3, In *Ieee-inns-enns international joint conference on neural networks*. https://doi.org/10.1109/IJCNN.2000.861302 (pp. 189–194): IEEE.

González, P.A., & Zamarreño, J.M. (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, *37*(6), 595–601. https://doi.org/10.1016/j.enbuild.2004.09.006.

Grosicki, E., Abed-Meraim, E., Hua, Y. (2005). A weighted linear prediction method for near-field source localization. *IEEE Transactions on Signal Processing*, *53*(10 I), 3651–3660. https://doi.org/10.1109/TSP.2005.855100.

Harpham, C., & Dawson, C.W. (2006). The effect of different basis functions on a radial basis function network for time series prediction: a comparative study. *Neurocomputing*, *69*(16), 2161–2170. https://doi.org/10.1016/j.neucom.2005.07.010.

He, H, Menicucci, D., Caudell, T., Mammoli, A. (2011). Real-time fault detection for solar hot water systems using adaptive resonance theory neural networks, In *Asme 2011 5th international conference on energy sustainability, volume es2011, Washington, DC*. Retrieved from http://proceedings.asmedigitalcollection.asme.org/data/conferences/es2011/70415/1059_1.pdf. Washington.

Heo, Y., Choudhary, R., Augenbroe, G.A. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, *47*, 550–560. https://doi.org/10.1016/j.enbuild.2011.12.029.

Heo, Y., & Zavala, V.M. (2012). Gaussian process modeling for measurement and verification of building energy savings. *Energy and Buildings*, *53*, 7–18. https://doi.org/10.1016/j.enbuild.2012.06.024.

Hong, S.M., Paterson, G., Burman, E., Steadman, P., Mumovic, D. (2014). A comparative study of benchmarking approaches for non-domestic buildings: Part 1 Top-down approach. *International Journal of Sustainable Built Environment*, *2*(2), 119–130. https://doi.org/10.1016/j.ijsbe.2014.04.001.

Hong, S.-M., Paterson, G., Mumovic, D., Steadman, P. (2014a). Improved benchmarking comparability for energy consumption in schools. *Building Research & Information*, *42*(1), 47–61. https://doi.org/10.1080/09613218.2013.814746.

Hong, S.M., Paterson, G., Mumovic, D., Steadman, P. (2014b). Improved benchmarking comparability for energy consumption in schools. *Building Research and Information*, *42*(1), 47–61. https://doi.org/10.1080/09613218.2013.814746.

Hong, T., Koo, C., Kim, J., Lee, M., Jeong, K. (2015). A review on sustainable construction management strategies for monitoring, diagnosing and retrofitting the building's dynamic energy performance: Focused on the operation and maintenance phase. *Applied Energy*, *155*, 671–707. https://doi.org/10.1016/j.apenergy.2015.06.043.

Hou, Z., & Lian, Z. (2009). An application of support vector machines in cooling load prediction, In *Intelligent systems and applications, 2009. isa, vol. 2*. https://doi.org/10.1109/IWISA.2009.5072707 (pp. 1–4): IEEE.

Hou, Z., Lian, Z., Yao, Y., Yuan, X. (2006). Cooling-load prediction by the combination of rough set theory and an artiticial neural-network based on data-fusion technique. *Applied Energy*, *83*(9), 1033–1046. https://doi.org/10.1016/j.apenergy.2005.08.006.

Huang, H., Chen, L., Hu, E. (2015). A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. *Energy and Buildings*, *97*, 86–97. https://doi.org/10.1016/j.enbuild.2015.03.045.

Hygh, J.S., DeCarolis, J.F., Hill, D.B., Ranjithan, S.R. (2012). Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, *57*, 165–175. https://doi.org/10.1016/j.buildenv.2012.04.021.

Jain, R.K., Smith, K.M., Culligan, P.J., Taylor, J.E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, *123*, 168–178. https://doi.org/10.1016/j.apenergy.2014.02.057.

Jalori, S., & Reddy, T.A. (2015). A new clustering method to identify outliers and diurnal schedules from building energy interval data. *ASHRAE Transactions*, *121*, 33–44. Retrieved from http://auroenergy.com/wp-content/uploads/2016/05/2015SaurabhASHRaE-TransClustering.pdf.

Jiang, X., Dong, B., Xie, L., Sweeney, L. (2010). Adaptive Gaussian Process for Short-Term Wind Speed Forecasting, In *ECAI*. Retrieved from http://www.ece.tamu.edu/le.xie/papers/Xie-AdaptiveGaussian-2010.pdf (pp. 661–666).

Jinhu, L., Xuemei, L., Lixing, D., Liangzhong, J. (2010). Applying principal component analysis and weighted support vector machine in building cooling load forecasting, In *International conference on computer and communication technologies in agriculture engineering, vol. 1*. https://doi.org/10.1109/CCTAE.2010.5543476 (pp. 434–437): IEEE.

Jung, H.C., Kim, J.S., Heo, H. (2015). Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach. *Energy and buildings*, *90*, 76–84. Elsevier B.V. https://doi.org/10.1016/j.enbuild.2014.12.029.

Kalogirou, S., & Bojic, M. (2000). Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy*, *25*(5), 479–491. https://doi.org/10.1016/S0360-5442(99)00086-9.

Kalogirou, S., Florides, G., Neocleous, C., Schizas, C. (2001). *Estimation of Daily Heating and Cooling Loads Using Artificial Neural Networks*. Naples. Retrieved from http://ktisis.cut.ac.cy/bitstream/10488/883/3/C41-CLIMA2001.pdf.

Kalogirou, S., Lalot, S., Florides, G., Desmet, B. (2008). Development of a neural network-based fault diagnostic system for solar thermal applications. *Solar Energy*, *82*(2), 164–172. https://doi.org/10.1016/j.solener.2007.06.010.

Kalogirou, S.A. (2000). Applications of artificial neural-networks for energy systems. *Applied Energy*, *67*(1–2), 17–35. https://doi.org/10.1016/S0306-2619(00)00005-2.

Karatasou, S., Santamouris, M., Geros, V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, *38*(8), 949–958. https://doi.org/10.1016/j.enbuild.2005.11.005.

Kavousian, A., & Rajagopal, R. (2014). Data-Driven Benchmarking of Building Energy Efficiency Utilizing Statistical Frontier Models. *Journal of Computing in Civil Engineering*, *28*(1), 79–88. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000327.

Seyedzadeh *et al. Visualization in Engineering*   (2018) 6:5

Page 19 of 20

Kelly, S., Crawford-Brown, D., Pollitt, M.G. (2012). Building performance evaluation and certification in the UK: Is SAP fit for purpose? *Renewable and Sustainable Energy Reviews*, *16*(9), 6861–6878. https://doi.org/10.1016/j.rser.2012.07.018.

Khayatian, F., Sarto, L., Dall'O', G. (2016). Application of neural networks for evaluating energy performance certificates of residential buildings. *Energy and Buildings*, *125*, 45–54. https://doi.org/10.1016/j.enbuild.2016.04.067.

Kialashaki, A., & Reisel, J.R. (2013). Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks. *Applied Energy*, *108*, 271–280. https://doi.org/10.1016/j.apenergy.2013.03.034.

Kialashaki, A., & Reisel, J.R. (2014). Development and validation of artificial neural network models of the energy demand in the industrial sector of the United States. *Energy*, *76*, 749–760. https://doi.org/10.1016/j.energy.2014.08.072.

Kumar, R., Aggarwal, R.K., Sharma, J.D. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, *65*, 352. https://doi.org/10.1016/j.enbuild.2013.06.007.

Lai, F., Magoulès, F., Lherminier, F. (2008). Vapnik's learning theory applied to energy consumption forecasts in residential buildings. *International Journal of Computer Mathematics*, *85*(10), 1563–1588. https://doi.org/10.1080/00207160802033582.

Leung, H., Lo, T., Wang, S. (2001). Prediction of Noisy Chaotic Time Series Using an Optimal Radial Basis Function Neural Network. *IEEE Transactions on Neural Networks*, *12*(5), 1163–1172. https://doi.org/10.1109/72.950144.

Li, K., Hu, C., Liu, G., Xue, W. (2015). Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. *Energy and Buildings*, *108*, 106–113. https://doi.org/10.1016/j.enbuild.2015.09.002.

Li, Q., Meng, Q., Cai, J., Yoshino, H., Mochida, A. (2009a). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, *86*(10), 2249–2256. https://doi.org/10.1016/j.apenergy.2008.11.035.

Li, Q., Meng, Q., Cai, J., Yoshino, H., Mochida, A. (2009b). Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*, *50*(1), 90–96. https://doi.org/10.1016/j.enconman.2008.08.033.

Li, Q., Ren, P., Meng, Q. (2010). Prediction model of annual energy consumption of residential buildings, In *2010 international conference on advances in energy engineering*. Retrieved from https://doi.org/10.1109/ICAEE.2010.5557576 (pp. 223–226): IEEE.

Li, X., Bowers, C.P., Schnier, T. (2010). Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics*, *57*(11), 3639–3644. https://doi.org/10.1109/TIE.2009.2027926.

Li, X., Ding, L., L, J., Xu, G., Li, J. (2010). A novel hybrid approach of KPCA and SVM for building cooling load prediction, In *3rd international conference on knowledge discovery and data mining, wkdd 2010*. https://doi.org/10.1109/WKDD.2010.137 (pp. 522–526).

Li, X., Ding, L., Li, L. (2010). A novel building cooling load prediction based on SVR and SAPSO, In *3ca 2010 - 2010 international symposium on computer, communication, control and automation, vol. 1*. https://doi.org/10.1109/3CA.2010.5533863 (pp. 528–532).

Li, Z., Han, Y., Xu, P. (2014). *Methods for benchmarking building energy consumption against its past or intended performance: An overview, vol. 124*. https://doi.org/10.1016/j.apenergy.2014.03.020.

Liang, J., & Du, R. (2007). Model-based Fault Detection and Diagnosis of HVAC systems using Support Vector Machine method. *International Journal of Refrigeration*, *30*(6), 1104–1114. https://doi.org/10.1016/j.ijrefrig.2006.12.012.

Lundin, M., Andersson, S., Åstin, R. (2004). Development and validation of a method aimed at estimating building performance parameters. *Energy and Buildings*, *36*(9), 905–914. https://doi.org/10.1016/j.enbuild.2004.02.005.

Ma, Z., Cooper, P., Daly, D., Ledo, L. (2012). Existing Building Retrofits : Methodology and State - of - the - Art. *Energy and buildings*, *55*(12), 889–902. https://doi.org/10.1016/j.enbuild.2012.08.018.

MacArthur, J.W., Mathur, A., Zhao, J. (1989). On-line recursive estimation for load profile prediction. *ASHRAE transactions*, *95*, 621–628. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=6935287.

Magoulès, F., Zhao, H.x., Elizondo, D. (2013). Development of an RDP neural network for building energy consumption fault detection and diagnosis. *Energy and Buildings*, *62*, 133–138. https://doi.org/10.1016/j.enbuild.2013.02.050.

Manfren, M., Aste, N., Moshksar, R. (2013). Calibration and uncertainty analysis for computer models - A meta-model based approach for integrated building energy simulation. *Applied Energy*, *103*, 627–641. https://doi.org/10.1016/j.apenergy.2012.10.031.

Marszal, A.J., Heiselberg, P., Bourrelle, J.S., Musall, E., Voss, K., Sartori, I., Napolitano, A. (2011). Author's personal copy Zero Energy Building A review of definitions and calculation methodologies Author's personal copy. *Energy and buildings*, *43*(4), 971–979. https://doi.org/10.1016/j.enbuild.2010.12.022.

Massana, J., Pous, C., Burgas, L., Melendez, J., Colomer, J. (2015). Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings*, *92*, 322–330. https://doi.org/10.1016/j.enbuild.2015.02.007.

Mena, R., Rodríguez, F., Castilla, M., Arahal, M.R. (2014). A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy and Buildings*, *82*, 142–155. https://doi.org/10.1016/j.enbuild.2014.06.052.

Mihalakakou, G., Santamouris, M., Tsangrassoulis, A. (2002). On the energy consumption in residential buildings. *Energy and Buildings*, *34*(7), 727–736. https://doi.org/10.1016/S0378-7788(01)00137-2.

Mousavi-Avval, S.H., Rafiee, S., Jafari, A., Mohammadi, A. (2011). Optimization of energy consumption for soybean production using Data Envelopment Analysis (DEA) approach. *Applied Energy*, *88*(11), 3765–3772. https://doi.org/10.1016/j.apenergy.2011.04.021.

Neto, A.H., & Fiorelli, F.A.S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, *40*(12), 2169–2176. https://doi.org/10.1016/j.enbuild.2008.06.013.

Nghiem, T.X., & Jones, C.N. (2017). Data-driven Demand Response Modeling and Control of Buildings with Gaussian Processes, In *2017 American control conference*. https://doi.org/10.1145/1235.

Nikolaou, T., Kolokotsa, D., Stavrakakis, G., Apostolou, A., Munteanu, C. (2015). Review and State of the Art on Methodologies of Buildings' Energy-Efficiency Classification, In *Managing indoor environments and energy in buildings with integrated intelligent systems*. https://doi.org/10.2174/9781608052851112010 (pp. 13–31): Springer International Publishing.

Noh, G., & Rajagopal, R. (2013). Data-driven forecasting algorithms for building energy consumption, In *Sensors and smart structures technologies for civil, mechanical, and aerospace systems, vol. 8692*. https://doi.org/10.1117/12.2009894 (p. 86920T). San Diego: SPIE.

Olofsson, T., & Andersson, S. (2001). Long-term energy demand predictions based on short-term measured data. *Energy and Buildings*, *33*(2), 85–91. https://doi.org/10.1016/S0378-7788(00)00068-2.

Park, B., Messer, C.J., Urbanik II, T. (1998). Short-term freeway traffic volume forecasting using radial basis function neural network. *Transportation Research Record: Journal of the Transportation Research Board, 1651*, *1651*, 39–47. https://doi.org/10.3141/1651-06.

Park, Y.-S., & Lek, S. (2016). *Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. In Developments in environmental modelling*, (pp. 123–140): Wiley Online Library. https://doi.org/10.1016/B978-0-444-63623-2.00007-4.

Paudel, S., Elmtiri, M., Kling, W.L., Corre, O.L., Lacarrière, B. (2014). Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. *Energy and Buildings*, *70*, 81–93. https://doi.org/10.1016/j.enbuild.2013.11.051.

Pérez-Ortiz, J.A., Gers, F.A., Eck, D., Schmidhuber, J. (2003). Kalman filters improve LSTM network performance in problems unsolvable by traditional recurrent nets. *Neural Networks*, *16*(2), 241–250. https://doi.org/10.1016/S0893-6080(02)00219-8.

Petcharat, S., Chungpaibulpatana, S., Rakkwamsuk, P. (2012). Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings. *Energy and Buildings*, *52*, 145–152. https://doi.org/10.1016/j.enbuild.2012.06.006.

Pieri, S.P., Tzouvadakis, I., Santamouris, M. (2015). Identifying energy consumption patterns in the Attica hotel sector using cluster analysis techniques with the aim of reducing hotels' CO2 footprint. *Energy and Buildings*, *94*, 252–262. https://doi.org/10.1016/j.enbuild.2015.02.017.

Platon, R., Dehkordi, V.R., Martel, J. (2015). Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy and Buildings*, *92*, 10–18. https://doi.org/10.1016/j.enbuild.2015.01.047.

Seyedzadeh *et al. Visualization in Engineering* (2018) 6:5

Page 20 of 20

Popescu, D., Ungureanu, F., Hernández-Guerrero, A. (2009). Simulation models for the analysis of space heat consumption of buildings. *Energy*, *34*(10), 1447–1453. https://doi.org/10.1016/j.energy.2009.05.035.

Pour Rahimian, F., Arciszewski, T., Goulding, J.S. (2014). Successful education for AEC professionals: case study of applying immersive gamelike virtual reality interfaces. *Visualization in Engineering*, *2*(1), 4. https://doi.org/10.1186/2213-7459-2-4.

Rastogi, P., Polytechnique, E., Lausanne, F.D. (2017). Gaussian-Process-Based Emulators for Building Performance Simulation, In *Building simulation 2017: The 15th international conference of ibpsa*. Retrieved from https://infoscience.epfl.ch/record/252858/files/BS2017448.pdf. San Francisco: IBPSA.

Reynolds, D. (2015). Gaussian Mixture Models. *Encyclopedia of biometrics*, 827–832. https://doi.org/10.1007/978-1-4899-7488-4196.

Ruch, D., Chen, L., Haberl, J.S., Claridge, D.E. (1993). A Change-Point Principal Component Analysis (CP/PCA) Method for Predicting Energy Usage in Commercial Buildings: The PCA Model. *Journal of solar energy engineering*, *115*(2), 77. https://doi.org/10.1115/1.2930035.

Santamouris, M., Mihalakakou, G., Patargias, P., Gaitani, N., Sfakianaki, K., Papaglastra, M., Pavlou, C., Doukas, P., Primikiri, E., Geros, V., Assimakopoulos, M.N., Mitoula, R., Zerefos, S. (2007). Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy and Buildings*, *39*(1), 45–51. https://doi.org/10.1016/j.enbuild.2006.04.018.

Shaikh, P.H., Nor, N.B.M., Nallagownden, P., Elamvazuthi, I., Ibrahim, T. (2014). A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews*, *34*, 409–429. https://doi.org/10.1016/j.rser.2014.03.027.

Smarra, F., Jain, A., de Rubeis, T., Ambrosini, D., D'Innocenzo, A., Mangharam, R. (2018). *Data-driven model predictive control using random forests for building energy optimization and climate control*. https://doi.org/10.1016/J.APENERGY.2018.02.126.

Solomon, D.M., Winter, R.L., Boulanger, A.G., Anderson, R.N., Wu, L.L. (2011). *Forecasting energy demand in large commercial buildings using support vector machine regression (Tech. Rep.)* Retrieved from http://academiccommons.columbia.edu/catalog/ac:143154.

Srivastav, A., Tewari, A., Dong, B. (2013). Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. *Energy and Buildings*, *65*, 438–447. https://doi.org/10.1016/j.enbuild.2013.05.037.

The Energy Systems Research Unit (ESRU) (2011). ESP-r. Retrieved 2018-02-25, from http://www.esru.strath.ac.uk/Programs/ESP-r.htm.

Tso, G.K.F., & Yau, K.K.W. (2007). Predicting electricity energy consumption : A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010.

University of Wisconsin-Madison (2015). A Transient Systems Simulation Program. Retrieved 31/02/2018, from http://sel.me.wisc.edu/trnsys/.

Wang, B., Xia, X., Zhang, J. (2014). A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. *Energy and Buildings*, *77*, 227–235. https://doi.org/10.1016/j.enbuild.2014.03.025.

Wong, S., Wan, K.K., Lam, T.N. (2010). Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, *87*(2), 551–557. https://doi.org/10.1016/j.apenergy.2009.06.028.

Xing-ping, Z., & Rui, G.U. (2007). Electrical Energy Consumption Forecasting Based on Cointegration and a Support Vector Machine in China, In *Wseas transactions on mathematics, vol. 6*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.533.9017&rep=rep1&type=pdf (pp. 878–883).

Xuemei, L., Yuyan, D., Lixing, D., Liangzhong, J. (2010). Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering, In *Computer and communication technologies in agriculture engineering (cctae), 2010 international conference on, vol. 1*. https://doi.org/10.1109/CCTAE.2010.5543577 (pp. 438–441).

Xuemei, L.X.L., Jin-hu, L.J.-h.L., Lixing, D.L.D., Gang, X.G.X., Jibin, L.J.L. (2009). Building Cooling Load Forecasting Model Based on LSSVM. *Asia-Pacific Conference on Information Processing*, *1*, 55–58. https://doi.org/10.1109/APCIP.2009.22.

Yalcintas, M. (2006). An energy benchmarking model based on artificial neural network method with a case example for tropical climates. *International Journal of Energy Research*, *31*(14), 1158–1174. https://doi.org/10.1002/er.1212.

Yalcintas, M., & Ozturk, U.A. (2007). An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. *International Journal of Energy Research*, *31*(4), 412–421. https://doi.org/10.1002/er.1232.

Yan, C.W., & Yao, J. (2010). Application of ANN for the prediction of building energy consumption at different climate zones with HDD and CDD, In *Proceedings of the 2010 2nd International Conference on Future Computer and Communication, ICFCC 2010, Vol. 3 (Cdd)*. https://doi.org/10.1109/ICFCC.2010.5497626 (pp. 286–289).

Yang, I.-H., Yeo, M.-S., Kim, K.-W. (2003). Application of artificial neural network to predict the optimal start time for heating system in building. *Energy Conversion and Management*, *44*(17), 2791–2809. https://doi.org/10.1016/S0196-8904(03)00044-X.

Yang, J., Ning, C., Deb, C., Zhang, F., Cheong, D., Lee, S.E., Sekhar, C., Tham, K.W. (2017). k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy and Buildings*, *146*, 27–37. https://doi.org/10.1016/j.enbuild.2017.03.071.

Yang, J., Rivard, H., Zmeureanu, R. (2005). On-line building energy prediction using adaptive artificial neural networks. *Energy and Buildings*, *37*(12), 1250–1259. https://doi.org/10.1016/j.enbuild.2005.02.005.

Yang, R., & Wang, L. (2013). Development of multi-agent system for building energy and comfort management based on occupant behaviors. *Energy and Buildings*, *56*, 1–7. https://doi.org/10.1016/j.enbuild.2012.10.025.

Yokoyama, R., Wakui, T., Satake, R. (2009). Prediction of energy demands using neural network with model identification by global optimization. *Energy Conversion and Management*, *50*(2), 319–327. https://doi.org/10.1016/j.enconman.2008.09.017.

Yu, Z., Fung, B.C., Haghighat, F., Yoshino, H., Morofsky, E. (2011). A systematic procedure to study the in uence of occupant behavior on building energy consumption. *Energy and Buildings*, *43*(6), 1409–1417. Retrieved from https://doi.org/10.1016/j.enbuild.2011.02.002.

Zhang, Y., O'Neill, Z., Dong, B., Augenbroe, G. (2015a). *Building and Environment*, *86*, 177. https://doi.org/10.1016/j.buildenv.2014.12.023.

Zhang, Y., O'Neill, Z., Dong, B., Augenbroe, G. (2015b). Comparisons of inverse modeling approaches for predicting building energy performance. *Building and Environment*, *86*, 177–190. https://doi.org/10.1016/j.buildenv.2014.12.023.

Zhang, Y., O'Neill, Z., Wagner, T., Augenbroe, G. (2013). An inverse model with uncertainty quantification to estimate the energy performance of an office building. *IBPSA Building Simulation*, 614–621. Retrieved from http://www.ibpsa.org/proceedings/BS2013/p1410.pdf.

Zhang, Y.-m., & Qi, W.-g. (2009). Interval Forecasting for Heating Load Using Support Vector Regression and Error Correcting Markov Chains, In *International conference on machine learning and cybernetics*. https://doi.org/10.1109/ICMLC.2009.5212405 (pp. 1106–1110). Hebei.

Zhao, H.-x., & Magoulès, F. (2010). Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption. *Journal of Algorithms & Computational Technology*, *4*(2), 231–249. https://doi.org/10.1260/1748-3018.4.2.231.

Zhao, H.-X., & Magoulès, F. (2012a). Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method. *Journal of Algorithms & Computational Technology*, *6*(1), 59–77. https://doi.org/10.1260/1748-3018.6.1.59.

Zhao, H.X., & Magoulès, F. (2012b). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049.