

Qian SHI, Chenyu LIU, Chao XIAO

Machine learning in building energy management: A critical review and future directions

© Higher Education Press 2021

Abstract Over the past two decades, machine learning (ML) has elicited increasing attention in building energy management (BEM) research. However, the boundary of the ML-BEM research has not been clearly defined, and no thorough review of ML applications in BEM during the whole building life-cycle has been published. This study aims to address this gap by reviewing the ML-BEM papers to ascertain the status of this research area and identify future research directions. An integrated framework of ML-BEM, composed of four layers and a series of driving factors, is proposed. Then, based on the hype cycle model, this paper analyzes the current development status of ML-BEM and tries to predict its future development trend. Finally, five research directions are discussed: (1) the behavioral impact on BEM, (2) the integration management of renewable energy, (3) security concerns of ML-BEM, (4) extension to other building life-cycle phases, and (5) the focus on fault detection and diagnosis. The findings of this study are believed to provide useful references for future research on ML-BEM.

Keywords building energy management, machine learning, integrated framework, knowledge evolution

1 Introduction

Global warming and climate change remain unavoidable problems facing the 21st century, primarily due to energy consumption and carbon emissions from human activities

Received May 10, 2021; accepted August 26, 2021

Qian SHI, Chenyu LIU, Chao XIAO (✉)
School of Economics and Management, Tongji University, Shanghai 200092, China
E-mail: chaoxiao@tongji.edu.cn

This work was supported by the National Key R&D Program of China (Grant No. 2020YFD1100604) and the Science and Technology Commission of Shanghai Municipality (Grant No. 19DZ1202800).

(McMichael et al., 2006). Building energy accounts for 20%–40% of the primary energy consumption (Castleton et al., 2010) and even 45% in some developed regions (Butler, 2008). However, current energy use in buildings is not efficient (Molina-Solana et al., 2017). Nearly one-third of the energy consumption of buildings can be reduced through efficient energy management (Li et al., 2019). Therefore, it is crucial to clearly define the scope of the management activities related to building energy management (BEM), that is, the boundary of BEM.

Numerous scholars have conducted research on BEM from different perspectives, including building energy demand and consumption prediction (Dong et al., 2005; Tso and Yau, 2007; Neto and Fiorelli, 2008), building energy performance evaluation (Scheuer et al., 2003; Crawley et al., 2008; Schlueter and Thesseling, 2009), energy operation diagnosis (Jaggs and Palmer, 2000; Ascione et al., 2013; Magoulès et al., 2013; Ficco et al., 2015), and energy system control and optimization (Lozano et al., 2009; Dagdougui et al., 2012; Lu et al., 2015; Wu et al., 2017). However, few studies have directly defined the concept of BEM, partly due to its broad knowledge scope and vague edge. de Wilde et al. (2013) regarded BEM as monitoring and analyzing the energy use of buildings with the aim of controlling and reducing energy expenditure. Wu (2009) divided the BEM into macro and micro levels, with the former indicating policies, regulations and standards, while the latter focusing on the daily operation and maintenance of buildings and occupancy energy-consuming behavior.

With the rise of artificial intelligence (AI) in recent years, AI techniques have already been adopted in several fields, including facial recognition for smartphone cameras (Zhao et al., 2003), fraud detection in credit card payments (Ngai et al., 2011), vehicle and pedestrian identification in self-driving cars (Goodfellow et al., 2016), online product recommendations and advertisements (Turban et al., 2018), etc., and BEM research is no exception. An increasing number of scholars apply the AI method to the study of BEM, among which machine learning is the

most widely used. Machine learning (ML) is a popular AI domain because of its ability to extract and generate knowledge from raw data that can be used to solve problems intuitively, in a manner similar to that of human beings. The most well-known definition of ML was articulated by Tom Mitchell in 1997: “A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.”

As one of the recent technological developments, ML has been widely used to solve BEM-related problems, exerting a considerable impact on the BEM research method and therefore changing the boundaries of BEM. Many scholars have conducted literature reviews on the application of ML in the specific field of BEM. Fouquier et al. (2013) reviewed the application of ML to building modeling and energy performance prediction, focusing on the resolution of equations simulating building thermal behavior and other mathematics-related equations. Deb et al. (2017) focused on time series forecasting techniques for building energy consumption, presenting a comprehensive review of the existing ML techniques for forecasting time series energy consumption. There are also many reviews of energy use forecasting (Kuster et al., 2017; Amasyali and El-Gohary, 2018), specific ML technology applications, e.g., artificial neural network (ANN) applications (Yildiz et al., 2017; Rodrigues et al., 2018), and energy efficiency evaluations (Goudarzi and Mostafaeipour, 2017).

Most of the current reviews in this field are merely focused on the application of specific ML techniques in one specific aspect of BEM, while lacking of clear depiction of the ML-BEM boundary and a comprehensive view of ML-BEM. With the increase in building energy demand, higher requirements are put forward for BEM (Danish et al., 2019), calling for more systematic and efficient energy management (Tronchin et al., 2018), which also highlights the need for a full-scale review. A good understanding of ML-BEM is also crucial for improving the efficiency of BEM and making comprehensive and effective policies. Moreover, the current review focuses more on the description of the status of the ML-BEM-related research, without interpreting the development process and current stage of the application of ML in BEM. Based on the hype cycle model, this paper analyzes the current development of ML-BEM and tries to predict its future development trend, which can make the management of building energy more forward-looking.

Therefore, the main objectives of this study are (1) to understand how ML techniques refine the boundary of BEM during the whole building life-cycle; (2) to develop an integrated framework for ML-BEM and analyze the holistic research status; and (3) to depict the knowledge evolution of ML-BEM and to predict its future development trend based on the hype cycle model. The rest of this

paper is structured as follows. Section 2 elaborates the research method used in this study, and Section 3 states the results of the quantitative analysis. A comprehensive framework is established and discussed in Section 4, and Section 5 shows the ML-BEM knowledge evolution. Future directions are discussed in Section 6, and Section 7 concludes this study.

2 Methodology

The core research methodology is illustrated in Fig. 1. The first step is to identify the keywords. According to the topic of the review, the keywords to be searched can be divided into two categories: Related to ML or to BEM. The ML-related keywords consist of core paradigms of ML, including supervised learning, unsupervised learning and reinforcement learning (Jordan and Mitchell, 2015). The BEM-related keywords are deduced from the definition of BEM, and after going through a number of highly cited publications in BEM (Dounis and Caraiscos, 2009; Guan et al., 2010; Masoso and Grobler, 2010; Palensky and Dietrich, 2011; Menezes et al., 2012; Ahmad et al., 2014; Gu et al., 2014; Shaikh et al., 2014; de Wilde, 2014), in addition to “building energy management”, some frequent terms, including “building energy consumption”, “building energy performance”, “building energy monitor”, and “building energy diagnosis”, are identified. Overall, the key phrases selected are shown in Table 1.

The next step is to search in the selected database. Regarding the database, Web of Science (WoS), Scopus and Google Scholar were considered. Taking record quality and coverage depth into account, WoS and Scopus are the main sources with the largest number of papers used for citation data (Norris and Oppenheim, 2007; Mongeon and Paul-Hus, 2016). In terms of coverage, WoS focuses on Science, Technology, Social Science, and Arts and Humanities, while Scopus concentrates more on Physical Sciences, Health Sciences and Life Sciences. In addition, WoS is designed to conduct citation analyses, providing more detailed citation analyses of high quality (Falagas et al., 2008), which are more suitable for the analysis method in this paper. Considering the coverage field and citation analysis quality, WoS was selected as the database to search in.

After being put forward in 1956, AI experienced a trough after a short hot period until the early 21st century, when AI saw a significant rise in a variety of fields. Therefore, this study conducted a two-decade literature review, with the time span set as 1998–2020. Only peer-reviewed English articles were included for further analysis to ensure the representativeness of the papers. Book reviews, editorials and proceeding papers were excluded to maintain the consistency of the analytical structure of the research aims and methods. To ensure that the topics are restricted to the managerial issues of

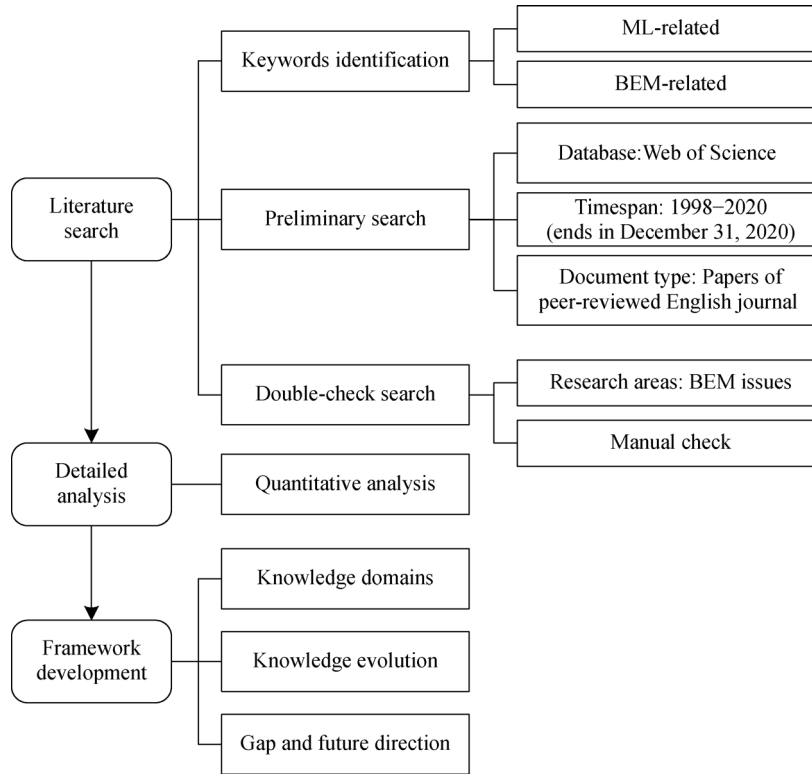


Fig. 1 Research processes of this study.

Table 1 Keywords for searching

	BEM	ML
Search clouds	building energy management; building energy consumption; building energy demand; building energy performance; building energy monitor; building energy diagnosis	machine learning; supervised learning; unsupervised learning; reinforcement learning
Exemplary search terms	('building energy management' or 'building energy consumption' or 'building energy demand' or 'building energy performance' or 'building energy monitor' or 'building energy diagnosis') AND ('machine learning' or 'supervised learning' or 'unsupervised learning' or 'reinforcement learning')	

life-cycle building energy, the search was further refined by limiting the research areas to “engineering”, “construction building technology”, “environmental sciences ecology”, “operations research management science” and “business economics”. As a result, 581 items were retrieved. Despite the specified criteria, there were still some papers that are outside the scope. Following the criteria that the approach must be ML and that the theme must be relevant to BEM, after screening and removing extraneous papers manually, a total of 387 papers were selected for further analysis.

3 Quantitative analysis

3.1 Paper co-citation analysis

Co-citation refers to a paper cited by two papers

simultaneously. The paper co-citation network of ML-BEM generated by CiteSpace is shown in Fig. 2. The nodes represent cited references, and lines between two nodes represent co-citation relationships. The thicker the line is, the more frequent the co-citation is, and the larger the node is, the more critical the study is. Detailed information of the publications co-cited more than 10 times is illustrated in Table 2. These highly co-cited papers can be divided into three categories: Literature reviews (Zhao and Magoulès, 2012; Fouquier et al., 2013; Ahmad et al., 2014; Robinson et al., 2017; Amasyali and El-Gohary, 2018; Wei et al., 2018), basic ML technology (Pedregosa et al., 2011), and predictive models (Li et al., 2009; Edwards et al., 2012; Tsanas and Xifara, 2012; Fan et al., 2014; 2017; Jain et al., 2014; Chae et al., 2016). It is normal for the literature reviews to have high citations because of their comprehensive and systematic analysis. The ML technology is a key tool for studying the BEM



Fig. 2 Paper co-citation network for the ML-BEM: 1998–2020.

problems, and energy prediction is one of the most concerning topic among scholars. More specifically, in energy prediction, scholars focus more on short-term

forecasting (hourly, daily) than on long-term forecasting (weekly, monthly).

3.2 Keyword co-occurrence analysis

When different terms are used in the same article, there is a relationship between the terms the authors adopted. The keywords of a paper are the reliable terms that the authors choose from the paper, reflecting the main content of the article (Whittaker et al., 1989). The papers were analyzed in terms of keywords, and the years per slice was set to 2. The keywords “machine learning”, “machine learning method”, and “machine learning model” were defined as the domain stop words, because these three words are directly related to ML-BEM. Substantial value was not added to the present analysis. In addition, the high percentage they occupied in the domain can influence the other keywords’ analysis. The keywords co-occurrence network is shown in Fig. 3, where the node represents keywords occurring more than twice, and the size of the keyword label is proportional to its frequency. The lines represent that the two keywords were used in one paper.

Table 2 Papers in the ML-BEM field with co-citation frequency more than 10

First author	Title	Year	Freq.	Journal
Zhao HX	A review on the prediction of building energy consumption	2012	44	<i>Renewable & Sustainable Energy Reviews</i>
Fan C	Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques	2014	23	<i>Applied Energy</i>
Jain RK	Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy	2014	21	<i>Applied Energy</i>
Fan C	A short-term building cooling load prediction method using deep learning algorithms	2017	20	<i>Applied Energy</i>
Pedregosa F	Scikit-learn: Machine learning in Python	2011	17	<i>Journal of Machine Learning Research</i>
Ahmad AS	A review on applications of ANN and SVM for building electrical energy consumption forecasting	2014	16	<i>Renewable & Sustainable Energy Reviews</i>
Edwards RE	Predicting future hourly residential electrical consumption: A machine learning case study	2012	16	<i>Energy and Buildings</i>
Amasyali K	A review of data-driven building energy consumption prediction studies	2018	15	<i>Renewable & Sustainable Energy Reviews</i>
Fouquier A	State of the art in building modelling and energy performances prediction: A review	2013	13	<i>Renewable & Sustainable Energy Reviews</i>
Wei YX	A review of data-driven approaches for prediction and classification of building energy consumption	2018	12	<i>Renewable & Sustainable Energy Reviews</i>
Robinson C	Machine learning approaches for estimating commercial building energy consumption	2017	12	<i>Applied Energy</i>
Tsanas A	Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools	2012	11	<i>Energy and Buildings</i>
Li Q	Applying support vector machine to predict hourly cooling load in the building	2009	11	<i>Applied Energy</i>
Chae YT	Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings	2016	10	<i>Energy and Buildings</i>



Fig. 4 Cluster analysis in the ML-BEM field: 1998–2020.

Regarding further interpretation, Cluster #0, namely, support vector machine (SVM), has the largest size, including 20 papers. Its top terms show that when studying energy demand and consumption prediction, SVM is often used, with more attention given to short-term (hourly) prediction and residential buildings. Cluster #1 illustrates that the research on the energy of cooling load generally uses the unsupervised ML method under the context of smart grids, and during the building maintenance stage, the failure mode is a common concern problem. Data mining is an important auxiliary means of ML because it is driving the vital developments of ML

in the 21st century. Cluster #2 shows that in the field of ML-BEM, data mining is intensively used in occupancy and indoor environmental data, and the linear mixed effect model is often adopted. Cluster #3, urban sustainability, focuses on energy consumption identification, and with the assistance of an adaptive algorithm, novel platforms, such as cloud forecasting systems and home automation systems, are used to solve urban sustainability problems. Cluster #4 reveals that air conditioning and mechanical research commonly concentrate on Assist-Controlled Mechanical Ventilation (ACMV) systems and that sparse swarm algorithms occur frequently. Cluster #5 illustrates that when studying the energy management of commercial buildings, especially office buildings, multiple buildings and passive commercial buildings, the surrogate model is introduced to reduce computation and to guide informed design (Westermann and Evins, 2019). Cluster #6 refers to the deep learning method applied in ML-BEM, including random forest (RF), gradient boosting (GB) machine, and deep recurrent neural network methods, and fault diagnosis is the most common problem solved (Zhao et al., 2019).

Previous research has reviewed several clusters, mainly concerning the ML applications (Cluster #0), data-driven methods (Cluster #2), as well as deep learning methods (Cluster #6) in BEM, indicating that attention has been given to technology improvement and innovation. In addition, research on cooling loads (Cluster #1) and air-conditioning systems (Cluster #4) has also gained much emphasis. From the results of keywords co-occurrence and cluster analysis, the main research topics are revealed to be accordant, and their complementarity has enriched the study of ML-BEM.

Table 4 Seven research clusters in the field of ML-BEM

Cluster	Size	Silhouette	Top terms	Typical papers
#0	20	0.793	Energy demand prediction, forecasting energy consumption, multi-family residential building, short-term prediction	Jain et al. (2014); Idowu et al. (2016); Chen and Tan (2017); Chen et al. (2017); Chou and Tran (2018); Guo et al. (2018b); Oneto et al. (2018)
#1	16	0.736	Unsupervised energy prediction, smart grid context, building maintenance, failure mode	Deb et al. (2016); Grolinger et al. (2016); Mocanu et al. (2016b); Fan et al. (2018); Yang et al. (2018)
#2	15	0.793	Occupancy data analytics, linear mixed effect model, using indoor environmental data, occupancy prediction model	Domahidi et al. (2014); Capozzoli et al. (2016); Liang et al. (2016); Naganathan et al. (2016); Zhang et al. (2016); Smarra et al. (2018)
#3	14	0.621	Identifying building energy consumption, adaptive algorithm, cloud forecasting system, home automation system	Benedetti et al. (2016); Chou and Ngo (2016); Marasco and Kontokosta (2016); Kuroha et al. (2018); Chou and Truong (2019); Gajewski et al. (2019)
#4	13	0.904	Assist-controlled mechanical ventilation system, sparse swarm algorithm, balancing indoor thermal comfort	Khosrowpour et al. (2016); Schmidt et al. (2017); Zhai et al. (2017); Zhai and Soh (2017); Cirigliano et al. (2018)
#5	11	0.946	Informed design guidance, passive commercial building, multiple building operation scenario, office building	Rackes et al. (2016); Chen and Yang (2017); Edwards et al. (2017); Rahman and Smith (2017); Tian et al. (2017)
#6	10	0.949	Random forest, gradient boosting machine, using deep recurrent neural network, deep learning-based fault diagnosis	Guo et al. (2018a); Rahman and Smith (2018); Rahman et al. (2018); Touzani et al. (2018); Wang et al. (2018b)

4 Integrated framework of ML-BEM

An integrated framework is developed to describe the current research status of the ML applications in BEM, as shown in Fig. 5. Four layers and a series of driving factors are identified to form the main body of the ML-BEM knowledge domain.

4.1 Perception layer

The technology layers are composed of perception layer, data layer and algorithm layer. The bottom layer, the perception layer, is used to identify and collect information, which feeds basic information regarding various building and environment parameters.

The most crucial elements in the perception layer are various types of sensors located in the interior of building equipment or the immediate environment. A huge amount of data are generated by sensors, which can be exploited to

increase energy efficiency (Molina-Solana et al., 2017). These sensors collect different kinds of information about the outside. For example, thermistors, thermocouples and IC sensors can collect temperature information, capacitive sensors and resistive sensors can collect humidity information, and light information can be collected by photodiodes and ultraviolet radiation sensors. What should be emphasized is the relevant sensor that collects occupant information. On the one hand, occupant behaviors and occupancy can affect the efficiency and performance of building energy. Dong and Lam (2014), Liang et al. (2016), Aftab et al. (2017) and Wei et al. (2019b) all stressed occupant behaviors and occupancy information when studying BEM. On the other hand, research on occupant thermal comfort (Zhai and Soh, 2017; Ghahramani et al., 2018; Du et al., 2019) is also of great concern in BEM. To collect occupant-related data, infrared sensor cameras are adopted to obtain the heat information of the human body to optimize the building energy performance,

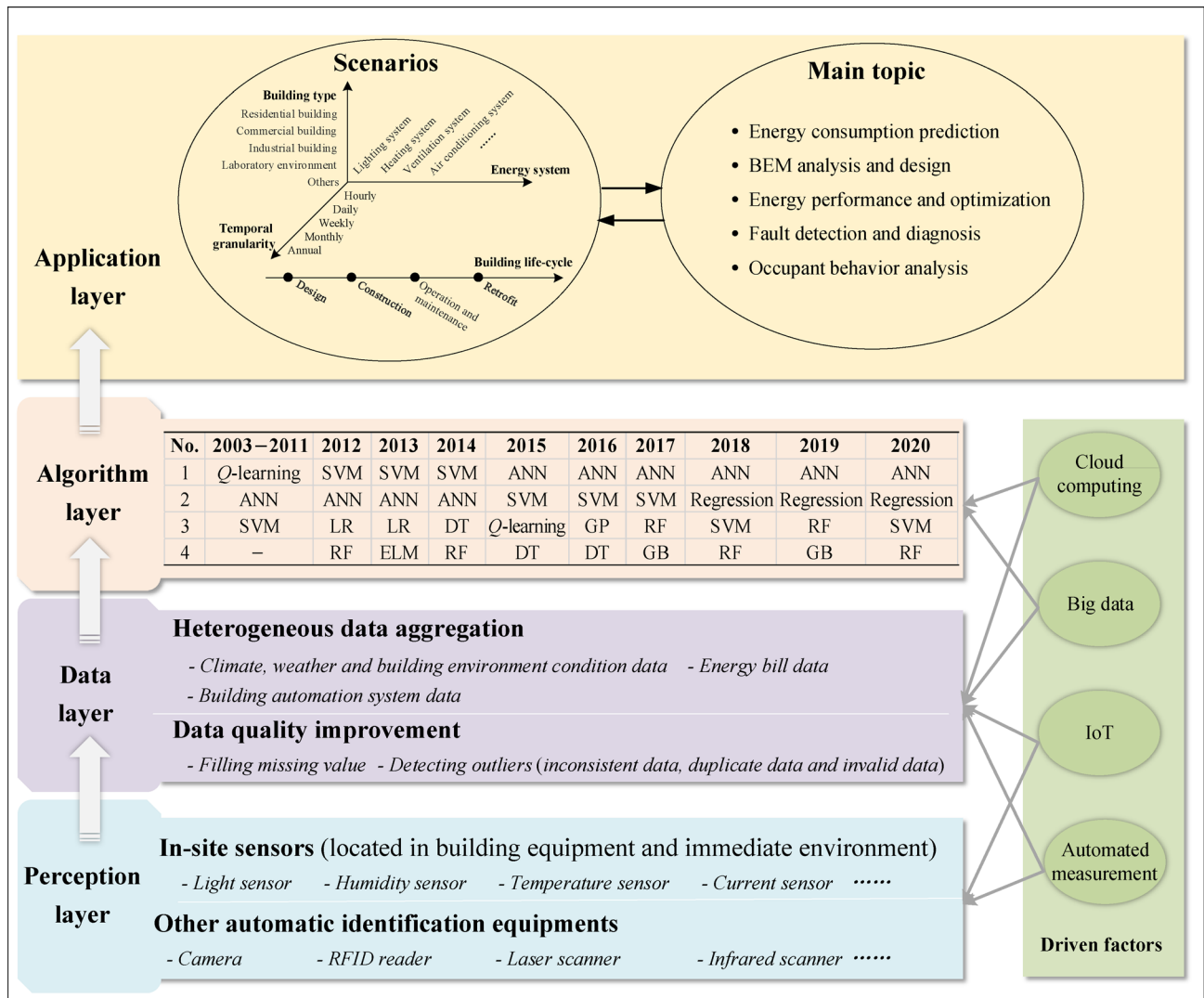


Fig. 5 The comprehensive framework.

while the Radio Frequency Identification (RFID) device can gain occupancy information, helping determine the heating and cooling loads. In addition, a laser scanner and infrared scanner can collect building shape and appearance information.

4.2 Data layer

The second layer is the data layer, where the data already collected are preprocessed before entering the algorithm layer. The sources of the data are varied. Apart from different kinds of sensors and other automatic identification equipment, climate, weather records and energy bills are data sources. Environmental factors have been influencing the development of BEM. Increasing climate change and environmental degradation, as well as people's requirements for environmental sustainability, continue to promote and influence research in BEM. Climate change (Cenek et al., 2018), weather conditions (Deb et al., 2017) and building environment conditions (Yun and Won, 2012) all exert impacts on building energy-related problems, especially on building energy efficiency and building energy modeling. Various data sources lead to heterogeneous data with diverse structures. As a consequence, it is necessary to aggregate the data properly, according to the research purpose and algorithm requirements. Data cleaning is another crucial part of the data layer because it is common to face inconsistent data, missing-value data, duplicate data and data with invalid values.

4.3 Algorithm layer

Before the introduction of ML into BEM, engineering and statistical methods were the most commonly used methods. Engineering methods are comprehensive but time-consuming, and statistical methods require a large amount of data. Through exploring many studies, it can be found that ML has been widely used in many fields, and BEM is no exception.

A large number of ML methods have been used to study BEM, including ANN (de Wilde, 2014; Benedetti et al., 2016; Deb et al., 2016; Zhang et al., 2016), SVM (Eisenhower et al., 2012; Jain et al., 2014; Cui et al., 2016; Grolinger et al., 2016; Candanedo et al., 2017), regression (Edwards et al., 2012; Tsanas and Xifara, 2012; Aftab et al., 2017; Fan et al., 2017; Kontokosta and Tull, 2017), and decision tree (DT) methods (Zhao et al., 2014; Basu et al., 2015; Idowu et al., 2016; Liang et al., 2016; Ryu and Moon, 2016). As seen in Table 5, ANN, SVM, regression, especially linear regression (LR), random forest and decision tree methods are the most popular methods. Additionally, many articles use several ML methods to study BEM-related problems (Edwards et al., 2012; Dong and Lam, 2014; Castelli et al., 2015; Mocanu et al., 2016a; Kim et al., 2018). Edwards et al. (2012) reported on the evaluation of seven different ML

algorithms applied to predict next hour residential building consumption and showed that the least squares SVM performs the best. Dong and Lam (2014) illustrated a methodology for integrated building heating and cooling control to reduce energy consumption, employing an adaptive Gaussian process (GP) and hidden Markov model.

In the comprehensive framework, the most frequently used ML algorithms every year from 2003 to 2020 are shown. ANN is obviously the most commonly used ML method due to its strong robustness (Liu et al., 2019), which has the capability to be immune to faults and noise (Tso and Yau, 2007). Benedetti et al. (2016) compared three different structures of neural networks to create the most suitable energy consumption control tool with an automated mechanism. Rahman et al. (2018) developed a recurrent neural network model to make medium- to long-term electricity consumption predictions for commercial and residential buildings with the purpose of supporting decision making concerning operation and demand response strategies. Typically, SVM, as a generalized classifier, can settle nonlinear and complicated problems effectively and is also frequently used in BEM. Eisenhower et al. (2012) established an analytical meta-model to optimize building energy models without repeating time-consuming energy simulations. Paudel et al. (2017) adopted the SVM method to predict the heating energy consumption of low-energy residential buildings, utilizing two kinds of modeling approaches, namely, "all data" and "relevant data", where the "relevant data" approach has lower complexity and higher accuracy. In addition, there are many other ML methods adopted in BEM, including random forest, extreme learning machine (ELM), decision tree methods, etc.

Table 5 Most frequently used ML-related method

No.	ML-related method	Number of article	%
1	ANN	145	37.5
2	SVM	83	21.4
3	Regression	76	19.6
4	Random forest	43	11.1
5	Decision tree	41	10.6
6	Extreme learning machine	25	6.5
7	Gaussian process	22	5.7
8	<i>K</i> -nearest neighbor	20	5.2
9	Markov model	20	5.2
10	<i>K</i> -means	17	4.4
11	<i>Q</i> -learning	15	3.9
12	Gradient boosting	14	3.6
13	Fuzzy method	12	3.1
14	Bayesian method	11	2.8

Note: Only shows method used more than 10th.

4.4 Application layer

Scholars generally focus on specific research scenarios. The research scenarios can be divided into four dimensions: Building type, building life-cycle phase, temporal granularity and energy type. The characteristics of energy use and management of different building types vary greatly, and the focal points of the papers are also different. Among the papers reviewed, the following building types were studied: Residential buildings, commercial buildings, industrial buildings, laboratory rooms and others. For a larger scale, multibuilding (Chen and Yang, 2017; Xu et al., 2019) or urban city energy (Kontokosta and Tull, 2017; Nutkiewicz et al., 2018) are also studied. If a paper studies two or more types of buildings, the counts are done separately. As Table 6 shows, the vast majority of the studies focused on commercial buildings (Liu and Henze, 2006; Najafi et al., 2012; Gao and Malkawi, 2014; Mulumba et al., 2015; Wang et al., 2018b), research on residential buildings takes second place (Wen et al., 2015; Henao et al., 2017; Jin et al., 2017; Paudel et al., 2017; Papadopoulos and Kontokosta, 2019), and research on industrial buildings accounts for only 12.4% (Eisenhower et al., 2012; de Wilde et al., 2013; Gong et al., 2017; Gallagher et al., 2018). Some scholars have studied problems with building types other than those above. Deb et al. (2016) focused on institutional buildings, employing the ANN method to forecast diurnal cooling load energy consumption, and through examination of three institutional buildings, the method was shown to be applicable to other institutional buildings.

Table 6 Proportion of papers studying different building types

No.	Building type	Number of article	%
1	Commercial	177	45.7
2	Residential	151	39.0
3	Laboratory rooms	67	17.3
4	Industrial	48	12.4
5	Others	20	5.2

In terms of the building life-cycle phase, most of the research focuses on the design phase and operation phase of buildings. While vast amounts of energy are consumed during the building operation phase, energy consumption during the construction phase, particularly in the production of construction materials, is also substantial. To estimate energy consumption during material production, an ANN model, taking into account the influence of the production and composition of asphalt, was developed by Androjić and Dolacek-Alduk (2018) with the purpose of successfully forecasting natural gas consumption in the process of hot mix asphalt production.

In regard to temporal granularity, hourly, daily, weekly,

monthly and even annual energy use are studied. The granularity of the collected data and the research topic lead to different selection of temporal granularity. For instance, until now, most residential modeling research only has access to monthly electrical consumption data. It is difficult to predict hourly or daily energy consumption based on monthly unit data alone, but it is possible to analyze time units with larger granularity by accumulating energy consumption over a period of time. Meanwhile, the research issues also affect the choice of data granularity. As to office building energy consumption prediction, for example, due to the difference between rest and working days, daily energy use data are required. While analyzing the fluctuation of household electricity consumption in one day, it is necessary to select hourly data at least. Many scholars have concentrated on hourly energy consumption prediction (Edwards et al., 2012; Chen et al., 2017; Wang et al., 2018b), while others have focused on daily energy consumption prediction (Manjarres et al., 2017; Wang et al., 2018a). In addition, many models have been developed for multiple granularities. Li et al. (2018) established a deep belief network-based hybrid model for daily and weekly prediction, while Nutkiewicz et al. (2018) proposed a data-driven urban energy simulation framework for hourly, daily and monthly use.

With regard to energy type, topics ranging from different energy loads, including cooling, heating, and electricity, to various energy systems, such as high voltage alternating current (HVAC) systems or a single air-conditioning system, are fully discussed. Among various energy types, the most studied one is electricity. Ahmad et al. (2019), Dong and Lam (2014), Lin et al. (2019) and Ngo (2019) studied cooling and/or heating loads, and Zhu et al. (2013), Chen and Sun (2017), Ghahramani et al. (2017), and Tian et al. (2019) considered different energy systems. Different combinations of building type, building life-cycle phase, temporal granularity and energy type constitute the research scenarios of ML-BEM.

With respect to the management of energy buildings, scholars have discussed different aspects of ML-BEM, and five main topics can be identified: a) energy demand/use/consumption prediction; b) energy system analysis and design; c) energy performance and optimization; d) fault detection and diagnosis; and e) occupant behavior analysis. The topic of energy demand/use/consumption prediction exerts a vital impact on reducing energy consumption and improving building energy demand and supply management (Amasyali and El-Gohary, 2018). Mocanu et al. (2016b) investigated two reinforcement learning algorithms based on a deep belief network to model building energy consumption in a smart grid context. The topic, fault detection and diagnosis, is gaining increasing attention. Li et al. (2016) proposed a fault detection and diagnosis method based on tree-structured learning for building cooling systems that improved the performance

with higher-accuracy fault severity levels. Karami and Wang (2018), Guo et al. (2018c) and Hu et al. (2019) also studied the fault detection and diagnosis of building energy. Occupant behavior analysis is another hot topic, playing an important role in energy consumption in buildings (Zhu et al., 2017; Razavi et al., 2019; Wei et al., 2019a). The topic of building energy modeling is more likely to solve building energy-related problems. Many scholars develop various building energy models to solve various energy problems for predicting (Tooke et al., 2014; Dong et al., 2016; Candanedo et al., 2017), optimizing (Eisenhower et al., 2012; Zhai et al., 2017; Smarra et al., 2018) and other purposes.

4.5 Driving factors

With the development of society, a number of new technologies, such as the Internet of Things (IoT) and big data, have gradually come into view. Four driving factors are identified, including cloud computing, big data, the IoT and automated measurement, which exert considerable impacts on all layers of the ML-BEM knowledge domain.

Big data are regarded as an effective tool for solving many social problems, providing new perspectives for many fields (Boyd and Crawford, 2012), and the ML-BEM area is in the same way. On the one hand, big data expand the data capacity of the data layer; on the other hand, it also promotes the data-driven development of algorithms in the algorithm layer. Edwards et al. (2017) presented a large-scale surrogate model using feed forward neural networks and Lasso regression in the context of big data. Grolinger et al. (2016) used ANN and SVM to show the significance of big data, especially the influence that the temporal data granularity exerts on the accuracy of energy consumption forecasting in the background of an event-organizing venue. Luo et al. (2019) developed an IoT-based big data platform based on the hybrids of *K*-means clustering and ANN for the prediction of building heating and cooling demands and demonstrated that due to the involvement of IoT sensors and big data, the overall prediction accuracy was improved.

The IoT provides a new opportunity for improving the intelligence of BEM by integrating various sensors, smart meters and actuators in a cost-saving way (Pan et al., 2015). For the perception layer, the IoT enables many new types of sensors to be installed and diversified sensor networks to be established. Tushar et al. (2018) illustrated the process of gaining building occupancy data via IoT sensors with low price and explained how occupancy behaviors influenced building energy use. Considering a transfer learning-based technique and an unsupervised learning technique, this research can help to design more efficient energy conservation measures and reduce electricity costs.

Cloud computing and the IoT cannot be separated from

each other. Mengistu et al. (2019), Yu et al. (2016), Chou and Truong (2019), and Gallagher et al. (2019) adopted ML methods, together with cloud-based technology and the IoT, to explore building energy. Yu et al. (2016) proposed a building management system based on the IoT and cloud-based technology that helps reduce the energy consumption of buildings. However, although cloud computing reduces the computing cost and expands the computing power and flexibility, it also increases the security risk and latency issues, which should be of importance.

An increasing number of automatic and intelligent equipment have put forward higher requirements for the management of building energy and automated measurement, therefore gaining much attention. Automatic measurement increases the complexity of sensors for data acquisition in the perception layer because increasing automation leads to the need for a more complex intelligence program. Moreover, the automated measurement enriches the data set and refines the data granularity. Along with moving toward smart buildings and smart cities, as one important component of automated measurement, the smart grid appears. Fenza et al. (2019) proposed a methodology for anomaly detection based on sensor data coming from a smart grid, monitoring the difference between the predicted and actual consumption to help identify energy consumption anomalies. Actually, the greatest challenge of automated measurement is determining which part of the data should be measured to support research and decision making.

The four layers, i.e., perception layer, data layer, algorithm layer and application layer, together with a series of driving factors, constitute the main knowledge domains of ML-BEM. The evolutionary trend of knowledge is discussed as follows.

5 Knowledge evolution

The ML technology has gone through many different stages since it was applied to BEM. Based on the hype cycle model introduced by Gartner nearly two decades ago, the knowledge evolution of ML-BEM can be concluded. It is very difficult to predict the development of a certain technology in one field because its numerous influencing factors increase its uncertainty. One useful model to analyze and predict technology development is the hype cycle model, which has gained substantial attention from many scholars in recent years (Jarvenpaa and Makinen, 2008; van Lente et al., 2013; Dedehayir and Steinert, 2016; Khodayari and Aslani, 2018). The hype cycle model proposes that technologies go through several stages: First culminating in peaks, then decreasing through disappointments, and finally reverting to expectations and achieving a smooth upward trend. The model explains a general path that a technology tracks over time. As shown

in Fig. 6, the evolution of the ML-BEM knowledge is concluded and divided into four phases: The ML-BEM concept formation phase, the booming phase, the concept extension and strategy phase, and the new technology and context phase. In light of the number of papers published per year, the four stages correspond to the ascent phase of the hype cycle curve before reaching the peak.

From 2003 to 2013, the basic conceptual problem was discussed. Frequent terms, such as “system”, “reinforcement learning”, “consumption”, “model”, and “thermal storage”, reflect the range of study: The system and energy use as the initial research object, the design and operation as the main life-cycle stages and the modeling as the dominating tool. This can be considered as the incipient phase of ML-BEM. Since 2014, ML has been explosively adopted in BEM, and ML-BEM has entered a booming phase. A range of ML methods have come into view, including ANN, SVM, and regression, accounting for a large frequency of occurrence. From 2016 to 2017, the research entered a period of accelerated development, during which the scope of ML-BEM extended to a wider range. Instead of concerning solely the system or the building itself, the research scope has gradually included the demand side, incorporating human behavior and comfort into the field. The keywords or terms “demand”, “behavior”, “thermal comfort”, and “consumer” reflect this tendency. From 2018 to 2020, new technologies and a new context led ML-BEM into a new phase. With the emergence of the IoT, cloud computing, big data, and smart sensors/grids/buildings/cities, ML-BEM faces more challenges and opportunities. In this phase, new technologies are introduced in conjunction with the ML technology, generating various integrated and innovative approaches for every application scenario of BEM. In

summary, the trend of ML-BEM is toward a more extensive research scope with more comprehensive research methods. According to the general pattern of technology development, as ML-BEM flourishes, it is worth considering that bottlenecks and downturns may be coming.

6 Future directions

6.1 Behavioral impact on BEM

It is widely accepted that sustainability analysis consists of environmental, social and economic impact evaluations. In the field of ML-BEM, occupant presence and behavior in buildings have been shown to have large impacts on space heating, cooling and ventilation demand, energy consumption of lighting and space appliances, and building controls (Page et al., 2008). Pan et al. (2019) used the Gauss distribution model to predict window behavior in office buildings, which significantly impacts the indoor environment and energy consumption. This study also compared the prediction performance between the Gaussian distribution modeling approach and the logistic regression modeling approach, and the results showed that the Gaussian distribution models could provide higher prediction accuracy. With the development of ML and other intelligent technologies, data collection methods, data types, and data magnitude also change. To reach the purpose of forecasting building energy and performance, it is essential to study occupant-related factors in BEM. For instance, in the context of ML and big data, the number of people in the building can be acquired and processed through the mobile terminal to realize the fine operation

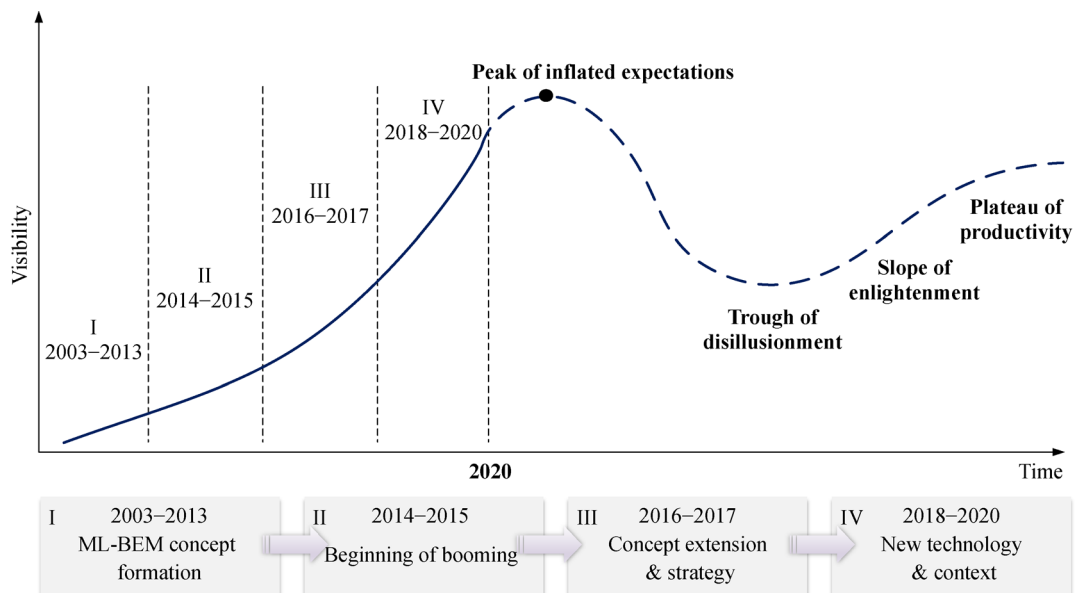


Fig. 6 The knowledge evolution of ML-BEM.

and maintenance management of the building. In addition to the impact of human behavior on building energy use, balancing occupant comfort and the energy consumption is also worth studying.

6.2 Integrated management of renewable energy

Having access to clean, affordable and reliable energy has been a cornerstone of the world's increasing prosperity and economic growth since the beginning of the industrial revolution (Chu and Majumdar, 2012). An increasing amount of renewable energy is used in buildings. Some scholars have studied BEM related to renewable energy, such as solar energy. Mohajeri et al. (2018) used SVM to classify 10085 building roofs in relation to their received solar energy in the city of Geneva in Switzerland, providing basic information for designing clean buildings and efficient solar integration on the roofs of buildings. The intermittent nature of renewable energy also brings new problems to energy management. With the gradual promotion of renewable energy in buildings, the integration of renewable energy in the existing building energy system and the joint management of traditional energy and renewable energy are topics worthy of study.

6.3 Security concerns of ML-BEM

As big data and the IoT become more popular, an increasing number of sensors and IoT devices are installed in buildings during the initial design phase or later retrofitting phase to collect many kinds of data. The data spreading over wireless networks and the energy management database both aggravate the security issue of BEM. Rathinavel et al. (2017) analyzed the security threats raised from IoT devices in the context of a specific building automation system and came up with corresponding countermeasures. There are many topics worth studying in terms of security in ML-BEM to ensure security and confidentiality when data are transmitted, stored and read, such as developing more robust security technologies and designing more secure approaches for information transmission and storage. In addition, different security scenarios can be developed to test different models of energy management.

6.4 Extension to other building life-cycle phases

Most of the research on building energy demand and sustainability focuses on the design phase and operation phase of buildings. Paudel et al. (2014) evaluated the transition- and time-dependent effects of operational power level characteristics of heating plant systems and employed neural networks to predict building heating energy demand over a short time horizon, which was essential for energy service companies to control the heat

plant system dynamically. While a majority of energy consumption occurs in the operation stage, the consumption during the construction phase is not to be underestimated. Although a small number of scholars have begun to focus on building energy consumption during the construction phase, research that focuses on the construction phase is too sparse compared to research that focuses on the design and operation phase. It is suggested that the construction phase also needs to be considered.

6.5 Focus on automatic fault detection and diagnosis

Often, ML serves three purposes: To predict, to diagnose, or to summarize (Jordan and Mitchell, 2015). Almost all studies on the application of ML in BEM are predictive studies, and relatively few studies are for diagnostic purposes. Karami and Wang (2018) developed an adaptive Gaussian mixture modeling approach for automatic fault detection and diagnosis in nonlinear systems. Najafi et al. (2012) proposed diagnostic algorithms based on a Bayesian network for air handling units that can solve the modeling limitations and measurement constraints. As more energy systems are used over longer periods of time, the likelihood of accidents inevitably increases. Combined with an image recognition algorithm, the running state of a building energy system can be analyzed and diagnosed based on infrared images of equipment. The fault detection and diagnosis of energy systems, especially early detection and diagnosis, has become more important and needs more attention.

7 Conclusions

Defining the scope of energy management in buildings can help improve energy management and thus improve energy efficiency in buildings (Guan et al., 2010; Li et al., 2019). The increasing use of ML technology brings about many profound changes in BEM and redefines the boundary of BEM. This study analyzes publications related to ML-BEM with the purpose of providing a comprehensive knowledge summary and development tendency of ML-BEM.

First, an integrated framework of ML-BEM was proposed. Four layers, namely, the perception layer, data layer, algorithm layer and application layer, composed of a series of research scenarios and main topics, together with four driving factors, constitute the knowledge domains. The main topics include 5 parts: a) energy consumption prediction; b) BEM analysis and design; c) energy performance and optimization; d) fault detection and diagnosis; and e) occupant behavior analysis. The framework proposed in the paper can be regarded as one of the best practices of ML-BEM. The increasingly renewing driving factors may bring breakthroughs to the application

of ML in the field of BEM, which can promote the further development of ML-BEM and alleviate the downward trend to some extent.

Based on the hype cycle model, the knowledge evolution of ML-BEM can be divided into 4 phases: The ML-BEM concept formation phase, the booming phase, the concept extension and strategy phase, and the new technology and context phase. It can be seen from the ML-BEM knowledge evolution that the ML-BEM research is fully operational, developing toward more extensive applications. It is also necessary to pay attention to the arrival of downturns.

Then, based on the discussion above, five future research directions were discussed. (1) The behavioral impact on BEM, such as occupancy, occupant behaviors and occupant comfort, can be taken into consideration when considering the factors affecting building energy and performance. (2) With the wide use of renewable energy, research on the integrated management of renewable energy in BEM can be strengthened. (3) As more information is transferred and stored, to ensure the reliability of BEM, more research is needed to explore the security of BEM. (4) Due to the vast amounts of energy consumed when buildings are constructed, more attention should be given to the construction phase. And (5) as more energy systems are used over longer periods of time, future studies are recommended to focus on building energy detection and diagnosis.

Although the use of ML in traditional areas such as facial recognition, fraud detection, and product recommendations is highly mature, the research on application of ML in BEM still needs to be developed. This study provides an integrated framework and development tendency of ML-BEM, contributing to the knowledge body of ML-BEM. This paper reviews the ML-BEM research based on the papers retrieved from WoS database before December 2020. To ensure the representativeness of papers and the consistency of the analytical structure of the research aims and methods, only peer-reviewed articles were included, which may miss some relevant proceeding papers. Despite the rapid development of ML, the most advanced development is in the theoretical stage and has not been applied in practice in the BEM area. Hence, although the papers published after 2020 are not covered, the framework of the paper would not be changed. With the development of the ML technology and the emergence of other new technologies, much exploration is required to further promote the development of ML-BEM research.

References

Aftab M, Chen C, Chau C K, Rahwan T (2017). Automatic HVAC control with real-time occupancy recognition and simulation-guided model predictive control in low-cost embedded system. *Energy and*

- Buildings*, 154: 141–156
- 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 & Sustainable Energy Reviews*, 33: 102–109
- Ahmad T, Chen H X, Shair J, Xu C L (2019). Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management. *International Journal of Refrigeration*, 98: 399–409
- Amasyali K, El-Gohary N M (2018). A review of data-driven building energy consumption prediction studies. *Renewable & Sustainable Energy Reviews*, 81: 1192–1205
- Androjić I, Dolacek-Alduk Z (2018). Artificial neural network model for forecasting energy consumption in hot mix asphalt (HMA) production. *Construction & Building Materials*, 170: 424–432
- Ascione F, de Masi R F, de Rossi F, Fistola R, Sasso M, Vanoli G P (2013). Analysis and diagnosis of the energy performance of buildings and districts: Methodology, validation and development of Urban Energy Maps. *Cities*, 35: 270–283
- Basu K, Debusschere V, Bacha S, Maulik U, Bondyopadhyay S (2015). Nonintrusive load monitoring: A temporal multilabel classification approach. *IEEE Transactions on Industrial Informatics*, 11(1): 262–270
- 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
- Boyd D, Crawford K (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, 15(5): 662–679
- Butler D (2008). Architecture: Architects of a low-energy future. *Nature*, 452(7187): 520–523
- Candanedo L M, Feldheim V, Deramaix D (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, 140: 81–97
- Capozzoli A, Piscitelli M S, Neri F, Grassi D, Serale G (2016). A novel methodology for energy performance benchmarking of buildings by means of Linear Mixed Effect Model: The case of space and DHW heating of out-patient Healthcare Centres. *Applied Energy*, 171: 592–607
- Castelli M, Trujillo L, Vanneschi L, Popovic A (2015). Prediction of energy performance of residential buildings: A genetic programming approach. *Energy and Buildings*, 102: 67–74
- Castleton H F, Stovin V, Beck S B M, Davison J B (2010). Green roofs: Building energy savings and the potential for retrofit. *Energy and Buildings*, 42(10): 1582–1591
- Cenek M, Haro R, Sayers B, Peng J (2018). Climate change and power security: Power load prediction for rural electrical microgrids using long short term memory and artificial neural networks. *Applied Sciences*, 8(5): 749
- Chae Y T, Horesh R, Hwang Y, Lee Y M (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings*, 111: 184–194
- Chen C M (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the*

- American Society for Information Science and Technology, 57(3): 359–377
- Chen J Y, Sun Y J (2017). A new multiplexed optimization with enhanced performance for complex air conditioning systems. *Energy and Buildings*, 156: 85–95
- Chen X, Yang H (2017). A multi-stage optimization of passively designed high-rise residential buildings in multiple building operation scenarios. *Applied Energy*, 206: 541–557
- Chen Y, Tan H (2017). Short-term prediction of electric demand in building sector via hybrid support vector regression. *Applied Energy*, 204: 1363–1374
- Chen Y B, Tan H W, Berardi U (2017). Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: A clustering-based hybrid approach. *Energy and Buildings*, 148: 228–237
- Chou J S, Ngo N T (2016). Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Applied Energy*, 177: 751–770
- Chou J S, Tran D S (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, 165(Part B): 709–726
- Chou J S, Truong N S (2019). Cloud forecasting system for monitoring and alerting of energy use by home appliances. *Applied Energy*, 249: 166–177
- Chu S, Majumdar A (2012). Opportunities and challenges for a sustainable energy future. *Nature*, 488(7411): 294–303
- Cirigliano A, Cordone R, Nacci A A, Santambrogio M D (2018). Toward smart building design automation: Extensible CAD framework for indoor localization systems deployment. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(1): 133–145
- Crawley D B, Hand J W, Kummert M, Griffith B T (2008). Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, 43(4): 661–673
- Cui C, Wu T, Hu M Q, Weir J D, Li X W (2016). Short-term building energy model recommendation system: A meta-learning approach. *Applied Energy*, 172: 251–263
- Dagdougui H, Minciardi R, Ouammi A, Robba M, Sacile R (2012). Modeling and optimization of a hybrid system for the energy supply of a “Green” building. *Energy Conversion and Management*, 64: 351–363
- Danish M S S, Senjyu T, Ibrahim A M, Ahmadi M, Howlader A M (2019). A managed framework for energy-efficient building. *Journal of Building Engineering*, 21: 120–128
- Deb C, Eang L S, Yang J J, Santamouris M (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*, 121: 284–297
- Deb C, Zhang F, Yang J J, Lee S E, Shah K W (2017). A review on time series forecasting techniques for building energy consumption. *Renewable & Sustainable Energy Reviews*, 74: 902–924
- Dedehayir O, Steinert M (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, 108: 28–41
- de Wilde P (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction*, 41: 40–49
- de Wilde P, Martinez-Ortiz C, Pearson D, Beynon I, Beck M, Barlow N (2013). Building simulation approaches for the training of automated data analysis tools in building energy management. *Advanced Engineering Informatics*, 27(4): 457–465
- Domahidi A, Ullmann F, Morari M, Jones C N (2014). Learning decision rules for energy efficient building control. *Journal of Process Control*, 24(6): 763–772
- 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
- Dong B, Lam K P (2014). A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation*, 7(1): 89–106
- Dong B, Li Z X, Rahman S M M, Vega R (2016). A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings*, 117: 341–351
- Dounis A I, Caraiscos C (2009). Advanced control systems engineering for energy and comfort management in a building environment: A review. *Renewable & Sustainable Energy Reviews*, 13(6–7): 1246–1261
- Du C Q, Li B Z, Liu H, Ji Y, Yao R M, Yu W (2019). Quantification of personal thermal comfort with localized airflow system based on sensitivity analysis and classification tree model. *Energy and Buildings*, 194: 1–11
- 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
- Edwards R E, New J, Parker L E, Cui B, Dong J (2017). Constructing large scale surrogate models from big data and artificial intelligence. *Applied Energy*, 202: 685–699
- Eisenhower B, O’Neill Z, Narayanan S, Fonoberov V A, Mezic I (2012). A methodology for meta-model based optimization in building energy models. *Energy and Buildings*, 47: 292–301
- Falagas M E, Pitsouni E I, Malietzis G A, Pappas G (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal*, 22(2): 338–342
- Fan C, Xiao F, Wang S (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, 127: 1–10
- Fan C, Xiao F, Zhao Y (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy*, 195: 222–233
- Fan C, Xiao F, Zhao Y, Wang J (2018). Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied Energy*, 211: 1123–1135
- Fenza G, Gallo M, Loia V (2019). Drift-aware methodology for anomaly detection in smart grid. *IEEE Access*, 7: 9645–9657
- Ficco G, Iannetta F, Ianniello E, d’Ambrosio Alfano F R, Dell’Isola M (2015). U-value in situ measurement for energy diagnosis of existing buildings. *Energy and Buildings*, 104: 108–121
- Fouquier A, Robert S, Suard F, Stephan L, Jay A (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable & Sustainable Energy Reviews*, 23: 272–288
- Gajewski M, Mongay Batalla J, Levi A, Togay C, Mavromoustakis C X,

- Mastorakis G (2019). Two-tier anomaly detection based on traffic profiling of the home automation system. *Computer Networks*, 158: 46–60
- Gallagher C V, Bruton K, Leahy K, O'Sullivan D T J (2018). The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings. *Energy and Buildings*, 158: 647–655
- Gallagher C V, Leahy K, O'Donovan P, Bruton K, O'Sullivan D T J (2019). IntelliMaV: A cloud computing measurement and verification 2.0 application for automated, near real-time energy savings quantification and performance deviation detection. *Energy and Buildings*, 185: 26–38
- Gao X F, Malkawi A (2014). A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. *Energy and Buildings*, 84: 607–616
- Ghahramani A, Castro G, Karvigh S A, Becerik-Gerber B (2018). Towards unsupervised learning of thermal comfort using infrared thermography. *Applied Energy*, 211: 41–49
- Ghahramani A, Karvigh S A, Becerik-Gerber B (2017). HVAC system energy optimization using an adaptive hybrid metaheuristic. *Energy and Buildings*, 152: 149–161
- Gong H F, Chen Z S, Zhu Q X, He Y L (2017). A Monte Carlo and PSO based virtual sample generation method for enhancing the energy prediction and energy optimization on small data problem: An empirical study of petrochemical industries. *Applied Energy*, 197: 405–415
- Goodfellow I, Bengio Y, Courville A (2016). *Deep Learning*. Cambridge, MA: MIT Press
- Goudarzi H, Mostafaiepour A (2017). Energy saving evaluation of passive systems for residential buildings in hot and dry regions. *Renewable & Sustainable Energy Reviews*, 68: 432–446
- Grolinger K, L'Heureux A, Capretz M A M, Seewald L (2016). Energy forecasting for event venues: Big data and prediction accuracy. *Energy and Buildings*, 112: 222–233
- Gu W, Wu Z, Bo R, Liu W, Zhou G, Chen W, Wu Z J (2014). Modeling, planning and optimal energy management of combined cooling, heating and power microgrid: A review. *International Journal of Electrical Power & Energy Systems*, 54: 26–37
- Guan X H, Xu Z B, Jia Q S (2010). Energy-efficient buildings facilitated by microgrid. *IEEE Transactions on Smart Grid*, 1(3): 243–252
- Guo Y, Tan Z, Chen H, Li G, Wang J, Huang R, Liu J, Ahmad T (2018a). Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. *Applied Energy*, 225: 732–745
- Guo Y, Wang J, Chen H, Li G, Liu J, Xu C, Huang R, Huang Y (2018b). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. *Applied Energy*, 221: 16–27
- Guo Y B, Tan Z H, Chen H X, Li G N, Wang J Y, Huang R G, Liu J Y, Ahmad T (2018c). Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. *Applied Energy*, 225: 732–745
- He Q, Wang G, Luo L, Shi Q, Xie J, Meng X (2017). Mapping the managerial areas of Building Information Modeling (BIM) using scientometric analysis. *International Journal of Project Management*, 35(4): 670–685
- Heno N, Agbossou K, Kelouwani S, Dubé Y, Fournier M (2017). Approach in nonintrusive type I load monitoring using subtractive clustering. *IEEE Transactions on Smart Grid*, 8(2): 812–821
- Hossain M M, Prybutok V, Evangelopoulos N (2011). Causal latent semantic analysis (cLSA): An illustration. *International Business Research*, 4(2): 38–50
- Hu R L, Granderson J, Auslander D M, Agogino A (2019). Design of machine learning models with domain experts for automated sensor selection for energy fault detection. *Applied Energy*, 235: 117–128
- Idowu S, Saguna S, Ahlund C, Schelen O (2016). Applied machine learning: Forecasting heat load in district heating system. *Energy and Buildings*, 133: 478–488
- Jaggs M, Palmer J (2000). Energy performance indoor environmental quality retrofit — A European diagnosis and decision making method for building refurbishment. *Energy and Buildings*, 31(2): 97–101
- 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
- Jarvenpaa H, Makinen S J (2008). Empirically detecting the Hype Cycle with the life cycle indicators: An exploratory analysis of three technologies. In: *IEEE International Conference on Industrial Engineering and Engineering Management*. Singapore, 12–16
- Jin X, Baker K, Christensen D, Isley S (2017). Foresee: A user-centric home energy management system for energy efficiency and demand response. *Applied Energy*, 205: 1583–1595
- Jordan M I, Mitchell T M (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245): 255–260
- Karami M, Wang L P (2018). Fault detection and diagnosis for nonlinear systems: A new adaptive Gaussian mixture modeling approach. *Energy and Buildings*, 166: 477–488
- Khodayari M, Aslani A (2018). Analysis of the energy storage technology using Hype Cycle approach. *Sustainable Energy Technologies and Assessments*, 25: 60–74
- Khosrowpour A, Gulbinas R, Taylor J E (2016). Occupant workstation level energy-use prediction in commercial buildings: Developing and assessing a new method to enable targeted energy efficiency programs. *Energy and Buildings*, 127: 1133–1145
- Kim J, Zhou Y X, Schiavon S, Raftery P, Brager G (2018). Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment*, 129: 96–106
- Kontokosta C E, Tull C (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied Energy*, 197: 303–317
- Kuroha R, Fujimoto Y, Hirohashi W, Amano Y, Tanabe S, Hayashi Y (2018). Operation planning method for home air-conditioners considering characteristics of installation environment. *Energy and Buildings*, 177: 351–362
- Kuster C, Rezgui Y, Mourshed M (2017). Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, 35: 257–270
- Li C, Ding Z, Yi J, Lv Y, Zhang G (2018). Deep Belief Network based hybrid model for building energy consumption prediction. *Energies*, 11(1): 242
- Li D, Zhou Y X, Hu G Q, Spanos C J (2016). Fault detection and diagnosis for building cooling system with a tree-structured learning

- method. *Energy and Buildings*, 127: 540–551
- Li Q, Meng Q, Cai J, Yoshino H, Mochida A (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10): 2249–2256
- Li Y, Garcia-Castro R, Mihindikulasooriya N, O'Donnell J, Vega-Sanchez S (2019). Enhancing energy management at district and building levels via an EM-KPI ontology. *Automation in Construction*, 99: 152–167
- Liang X, Hong T, Shen G Q (2016). Occupancy data analytics and prediction: A case study. *Building and Environment*, 102: 179–192
- Lin X Y, Tian Z, Lu Y K, Zhang H J, Niu J D (2019). Short-term forecast model of cooling load using load component disaggregation. *Applied Thermal Engineering*, 157: 113630
- Liu S M, Henze G P (2006). Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: Part I. Theoretical foundation. *Energy and Buildings*, 38(2): 142–147
- Liu Z, Wu D, Liu Y, Han Z, Lun L, Gao J, Jin G, Cao G (2019). Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. *Energy Exploration & Exploitation*, 37(4): 1426–1451
- Lozano M A, Ramos J C, Carvalho M, Serra L M (2009). Structure optimization of energy supply systems in tertiary sector buildings. *Energy and Buildings*, 41(10): 1063–1075
- Lu Y H, Wang S W, Zhao Y, Yan C C (2015). Renewable energy system optimization of low/zero energy buildings using single-objective and multi-objective optimization methods. *Energy and Buildings*, 89: 61–75
- Luo X J, Oyedele L O, Ajayi A O, Monyei C G, Akinade O O, Akanbi L A (2019). Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. *Advanced Engineering Informatics*, 41: 100926
- 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
- Manjarres D, Mera A, Perea E, Lejarazu A, Gil-Lopez S (2017). An energy-efficient predictive control for HVAC systems applied to tertiary buildings based on regression techniques. *Energy and Buildings*, 152: 409–417
- Marasco D E, Kontokosta C E (2016). Applications of machine learning methods to identifying and predicting building retrofit opportunities. *Energy and Buildings*, 128: 431–441
- Masoso O T, Grobler L J (2010). The dark side of occupants' behaviour on building energy use. *Energy and Buildings*, 42(2): 173–177
- McMichael A J, Woodruff R E, Hales S (2006). Climate change and human health: Present and future risks. *Lancet*, 367(9513): 859–869
- Menezes A C, Cripps A, Bouchlaghem D, Buswell R (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97: 355–364
- Mengistu M A, Girmay A A, Camarda C, Acquaviva A, Patti E (2019). A cloud-based on-line disaggregation algorithm for home appliance loads. *IEEE Transactions on Smart Grid*, 10(3): 3430–3439
- Mocanu E, Nguyen P H, Gibescu M, Kling W L (2016a). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids & Networks*, 6: 91–99
- Mocanu E, Nguyen P H, Kling W L, Gibescu M (2016b). Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning. *Energy and Buildings*, 116: 646–655
- Mohajeri N, Assouline D, Guiboud B, Bill A, Gudmundsson A, Scartezzini J L (2018). A city-scale roof shape classification using machine learning for solar energy applications. *Renewable Energy*, 121: 81–93
- Molina-Solana M, Ros M, Ruiz M D, Gomez-Romero J, Martin-Bautista M J (2017). Data science for building energy management: A review. *Renewable & Sustainable Energy Reviews*, 70: 598–609
- Mongeon P, Paul-Hus A (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1): 213–228
- Mulumba T, Afshari A, Yan R, Shen W, Norford L K (2015). Robust model-based fault diagnosis for air handling units. *Energy and Buildings*, 86: 698–707
- Naganathan H, Chong W O, Chen X (2016). Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches. *Automation in Construction*, 72: 187–194
- Najafi M, Auslander D M, Bartlett P L, Haves P, Sohn M D (2012). Application of machine learning in the fault diagnostics of air handling units. *Applied Energy*, 96: 347–358
- 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
- Newman M E J (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23): 8577–8582
- Ngai E W T, Hu Y, Wong Y H, Chen Y J, Sun X (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3): 559–569
- Ngo N T (2019). Early predicting cooling loads for energy-efficient design in office buildings by machine learning. *Energy and Buildings*, 182: 264–273
- Norris M, Oppenheim C (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2): 161–169
- Nutkiewicz A, Yang Z, Jain R K (2018). Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied Energy*, 225: 1176–1189
- Oneto L, Laureri F, Robba M, Delfino F, Anguita D (2018). Data-driven photovoltaic power production nowcasting and forecasting for polygeneration microgrids. *IEEE Systems Journal*, 12(3): 2842–2853
- Page J, Robinson D, Morel N, Scartezzini J L (2008). A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, 40(2): 83–98
- Palensky P, Dietrich D (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, 7(3): 381–388
- Pan J, Jain R, Paul S, Vu T, Saifullah A, Sha M (2015). An Internet of Things framework for smart energy in buildings: Designs, prototype, and experiments. *IEEE Internet of Things Journal*, 2(6): 527–537
- Pan S, Han Y Y, Wei S, Wei Y X, Xia L, Xie L, Kong X R, Yu W (2019).

- A model based on Gauss Distribution for predicting window behavior in building. *Building and Environment*, 149: 210–219
- Papadopoulos S, Kontokosta C E (2019). Grading buildings on energy performance using city benchmarking data. *Applied Energy*, 233–234: 244–253
- Paudel S, Elmitri M, Couturier S, Nguyen P H, Kamphuis R, Lacarriere B, Le Corre O (2017). A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. *Energy and Buildings*, 138: 240–256
- Paudel S, Elmtiri M, Kling W L, Le Corre O, Lacarriere B (2014). Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. *Energy and Buildings*, 70: 81–93
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830
- Rackes A, Melo A P, Lamberts R (2016). Naturally comfortable and sustainable: Informed design guidance and performance labeling for passive commercial buildings in hot climates. *Applied Energy*, 174: 256–274
- Rahman A, Smith A D (2017). Predicting fuel consumption for commercial buildings with machine learning algorithms. *Energy and Buildings*, 152: 341–358
- Rahman A, Smith A D (2018). Predicting heating demand and sizing a stratified thermal storage tank using deep learning algorithms. *Applied Energy*, 228: 108–121
- Rahman A, Srikumar V, Smith A D (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied Energy*, 212: 372–385
- Rathinavel K, Pipattanasomporn M, Kuzlu M, Rahman S (2017). Security concerns and countermeasures in IoT-integrated smart buildings. In: *IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. Washington, DC, 1–5
- Razavi R, Gharipour A, Fleury M, Akpan I J (2019). Occupancy detection of residential buildings using smart meter data: A large-scale study. *Energy and Buildings*, 183: 195–208
- Robinson C, Dilkina B, Hubbs J, Zhang W, Guhathakurta S, Brown M A, Pendyala R M (2017). Machine learning approaches for estimating commercial building energy consumption. *Applied Energy*, 208: 889–904
- Rodrigues E, Gomes Á, Gaspar A R, Henggeler Antunes C (2018). Estimation of renewable energy and built environment-related variables using neural networks: A review. *Renewable & Sustainable Energy Reviews*, 94: 959–988
- Ryu S H, Moon H J (2016). Development of an occupancy prediction model using indoor environmental data based on machine learning techniques. *Building and Environment*, 107: 1–9
- Scheuer C, Keoleian G A, Reppe P (2003). Life cycle energy and environmental performance of a new university building: Modeling challenges and design implications. *Energy and Buildings*, 35(10): 1049–1064
- Schlueter A, Thesseling F (2009). Building information model based energy/exergy performance assessment in early design stages. *Automation in Construction*, 18(2): 153–163
- Schmidt M, Moreno M V, Schülke A, Macek K, Mařík K, Pastor A G (2017). Optimizing legacy building operation: The evolution into data-driven predictive cyber-physical systems. *Energy and Buildings*, 148: 257–279
- Shaikh P H, Nor N B, Nallagownden P, Elamvazuthi I, Ibrahim T (2014). A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable & Sustainable Energy Reviews*, 34: 409–429
- 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. *Applied Energy*, 226: 1252–1272
- Tian W, Yang S, Zuo J, Li Z, Liu Y (2017). Relationship between built form and energy performance of office buildings in a severe cold Chinese region. *Building Simulation*, 10(1): 11–24
- Tian Z C, Si B H, Shi X, Fang Z G (2019). An application of Bayesian Network approach for selecting energy efficient HVAC systems. *Journal of Building Engineering*, 25: 100796
- Tooke T R, Coops N C, Webster J (2014). Predicting building ages from LiDAR data with random forests for building energy modeling. *Energy and Buildings*, 68: 603–610
- Touzani S, Granderson J, Fernandes S (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158: 1533–1543
- Tronchin L, Manfren M, Nastasi B (2018). Energy efficiency, demand side management and energy storage technologies — A critical analysis of possible paths of integration in the built environment. *Renewable & Sustainable Energy Reviews*, 95: 341–353
- Tsanas A, Xifara A (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49: 560–567
- 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
- Turban E, Outland J, King D, Lee J K, Liang T P, Turban D C (2018). *Electronic Commerce 2018: A Managerial and Social Networks Perspective*. Cham: Springer
- Tushar W, Wijerathne N, Li W T, Yuen C, Poor H V, Saha T K, Wood K L (2018). Internet of Things for Green Building Management disruptive innovations through low-cost sensor technology and artificial intelligence. *IEEE Signal Processing Magazine*, 35(5): 100–110
- van Lente H, Spitters C, Peine A (2013). Comparing technological hype cycles: Towards a theory. *Technological Forecasting and Social Change*, 80(8): 1615–1628
- Wang J Y, Li G N, Chen H X, Liu J Y, Guo Y B, Sun S B, Hu Y P (2018a). Energy consumption prediction for water-source heat pump system using pattern recognition-based algorithms. *Applied Thermal Engineering*, 136: 755–766
- Wang Z Y, Wang Y R, Zeng R C, Srinivasan R S, Ahrentzen S (2018b). Random Forest based hourly building energy prediction. *Energy and Buildings*, 171: 11–25
- Wei Y X, Zhang X X, Shi Y, Xia L, Pan S, Wu J S, Han M J, Zhao X Y (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable & Sustainable Energy Reviews*, 82(Part 1): 1027–1047

- Wei Y X, Xia L, Pan S, Wu J S, Zhang X X, Han M J, Zhang W Y, Xie J C, Li Q P (2019a). Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. *Applied Energy*, 240: 276–294
- Wei Y X, Yu H W, Pan S, Xia L, Xie J C, Wang X R, Wu J S, Zhang W J, Li Q P (2019b). Comparison of different window behavior modeling approaches during transition season in Beijing, China. *Building and Environment*, 157: 1–15
- Wen Z, O'Neill D, Maei H (2015). Optimal demand response using device-based reinforcement learning. *IEEE Transactions on Smart Grid*, 6(5): 2312–2324
- Westermann P, Evins R (2019). Surrogate modelling for sustainable building design: A review. *Energy and Buildings*, 198: 170–186
- Whittaker J, Courtial J P, Law J (1989). Creativity and conformity in science: Titles, keywords and co-word analysis. *Social Studies of Science*, 19(3): 473–496
- Wu R, Mavromatidis G, Orehoung K, Carmeliet J (2017). Multi-objective optimisation of energy systems and building envelope retrofit in a residential community. *Applied Energy*, 190: 634–649
- Wu Y (2009). Scientific management — The first step of building energy efficiency. In: *International Conference on Information Management, Innovation Management and Industrial Engineering*. Xi'an: IEEE, 619–622
- Xu X D, Wang W, Hong T Z, Chen J Y (2019). Incorporating machine learning with building network analysis to predict multi-building energy use. *Energy and Buildings*, 186: 80–97
- Yang C, Shen W, Chen Q, Gunay B (2018). A practical solution for HVAC prognostics: Failure mode and effects analysis in building maintenance. *Journal of Building Engineering*, 15: 26–32
- Yildiz B, Bilbao J I, Sproul A B (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable & Sustainable Energy Reviews*, 73: 1104–1122
- Yu J, Kim M, Bang H C, Bae S H, Kim S J (2016). IoT as an application: Cloud-based building management systems for the Internet of Things. *Multimedia Tools and Applications*, 75(22): 14583–14596
- Yun J, Won K H (2012). Building environment analysis based on temperature and humidity for smart energy systems. *Sensors*, 12(10): 13458–13470
- Zhai D Q, Chaudhuri T, Soh Y C (2017). Modeling and optimization of different sparse Augmented Firefly Algorithms for ACMV systems under two case studies. *Building and Environment*, 125: 129–142
- Zhai D Q, Soh Y C (2017). Balancing indoor thermal comfort and energy consumption of ACMV systems via sparse swarm algorithms in optimizations. *Energy and Buildings*, 149: 1–15
- Zhang D, Li S H, Sun M, O'Neill Z (2016). An optimal and learning-based demand response and home energy management system. *IEEE Transactions on Smart Grid*, 7(4): 1790–1801
- Zhao H X, Magoulès F (2012). A review on the prediction of building energy consumption. *Renewable & Sustainable Energy Reviews*, 16(6): 3586–3592
- Zhao J, Lasternas B, Lam K P, Yun R, Loftness V (2014). Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. *Energy and Buildings*, 82: 341–355
- Zhao W, Chellappa R, Phillips P J, Rosenfeld A (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4): 399–458
- Zhao Y, Li T, Zhang X, Zhang C (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable & Sustainable Energy Reviews*, 109: 85–101
- Zhu N, Shan K, Wang S W, Sun Y J (2013). An optimal control strategy with enhanced robustness for air-conditioning systems considering model and measurement uncertainties. *Energy and Buildings*, 67: 540–550
- Zhu Q C, Chen Z H, Masood M K, Soh Y C (2017). Occupancy estimation with environmental sensing via non-iterative LRF feature learning in time and frequency domains. *Energy and Buildings*, 141: 125–133