Version of Record: https://www.sciencedirect.com/science/article/pii/S0306261921000209
Manuscript_1be3e2c681bc036b793d028b375cacbe

A Review of Machine Learning in Building Load Prediction

Liang Zhang[1*], Jin Wen[2], Yanfei Li[1], Jianli Chen[3], Yunyang Ye[4], Yangyang Fu[5], William Livingood[1]
1. National Renewable Energy Laboratory, Golden, Colorado, U.S.
2. Drexel University, Philadelphia, Pennsylvania, U.S.
3. The University of Utah, Salt Lake City, Utah, U.S.
4. Pacific Northwest National Laboratory, Richland, Washington, U.S.
5. Texas A&M University, College Station, Texas, U.S.
*Corresponding author, Liang.Zhang@nrel.gov

HIGHLIGHTS

- This paper reviews building load prediction with machine learning techniques.
- Review and technical papers are searched by Sub-keyword Synonym Searching method.
- Technical papers are reviewed in terms of application, algorithms, and data.
- Primary limitations and gaps are identified; future trends are predicted.
- A guidance for future technical paper on building load prediction is proposed.

ABSTRACT

The surge of machine learning and increasing data accessibility in buildings provide great opportunities for applying machine learning to building energy system modeling and analysis. Building load prediction is one of the most critical components for many building control and analytics activities, as well as grid-interactive and energy efficiency building operation. While a large number of research papers exist on the topic of machine-learning-based building load prediction, a comprehensive review from the perspective of machine learning is missing. In this paper, we review the application of machine learning techniques in building load prediction under the organization and logic of the machine learning, which is to perform tasks T using Performance measure P and based on learning from Experience E.

Firstly, we review the applications of building load prediction model (task **T**). Then, we review the modeling algorithms that improve machine learning performance and accuracy (performance **P**). Throughout the papers, we also review the literature from the data perspective for modeling (experience **E**), including data engineering from the sensor level to data level, pre-processing, feature extraction and selection. Finally, we conclude with a discussion of well-studied and relatively unexplored fields for future research reference. We also identify the gaps in current machine learning application and predict for future trends and development.

KEYWORDS

building energy system; building load prediction; building energy forecasting; machine learning; feature engineering; data engineering

NOMENCLATURE

ACO: Ant Colony Optimization
AI: Artificial Intelligence
ANFIS: Adaptive Neuro Fuzzy Inference System
ANN: Artificial Neural Network
ARIMA: AutoRegressive Integrated Moving Average model
ARIMAX: AutoRegressive Integrated Moving Average with eXternal input
ARMA: AutoRegressive Moving Average model
ARX: AutoRegressive eXogenous model
BAS: Building Automation System
BOT: Building Topology Ontology
CBR: Case-Based Reasoning
CHAID: Chi-squared Automatic Interaction Detector

CRBM: Conditional Restricted Boltzmann Machine
DE: Differential Evolution
ELM: Extreme Learning Machine
EWKM: Entropy-Weighted K-Means
FCRBM: Factored Conditional Restricted Boltzmann Machine
FFNN: Feed Forward Neural Network
GA: Genetic Algorithm
GESD: Generalized Extreme Studentized Deviate
GMDH: Group Method of Data Handling
HVAC: Heating Ventilating and Air-Conditioning
IBDA: IoT Big Data Analytics
IFC: Industry Foundation Classes
IoT: Internet of Things
kNN: k-Nearest Neighbor
LSSVM: Least Squares Support Vector Machine
MARS: Multivariate Adaptive Regression Splines
MLR: Multivariate Linear Regression
MPC: Model Predictive Control
NN: Neural Network
PCA: Principle Component Analysis
PSO: Particle Swarm Optimization
RBFN: Radial Basis Function Network
RC: Resistor-Capacitor network
RNN: Recurrent Neural Networks
SAREF: Smart Appliances REFerence ontology
SARIMA: Seasonal AutoRegressive Integrated Moving Average
SSS: Sub-keyword Synonym Searching
SVM: Support Vector Machine
SVR: Support Vector Regression
WSN: Wireless Sensor Network

1. Introduction

In the United States, the building sector consumed 32% of primary energy consumption in 2019, which consumes 11% more primary energy than that in 2010 and global energy consumption in buildings will grow by 1.3% per year on average from 2018 to 2050 [1]. Space heating, space cooling, and lighting were the dominant end uses in 2010, accounting for close to half of consumed energy in the buildings sector.

Building energy system modeling is essential to design, predict, optimize, control, and diagnose the operation of building energy systems. "Building energy system" in this paper is a general concept that covers a specific component in a building system, such as a chiller, to a building system, such as the heating ventilating, and air conditioning (HVAC) system, even to the whole building as one system.

The complexity, dynamics, and nonlinearity of building energy systems in a single building and building clusters (multiple buildings) have a high demand for modeling techniques. Existing building energy system modeling approaches can be categorized into physics-based (white-box) modeling approach, data-driven (black-box) modeling approach, and those in between (gray-box model) [2]. With the wide adoption of building automation system (BAS) and Internet of Things (IoT) in buildings, massive measurements reflecting on equipment and building operations are continuously collected by sensors and other sources. This provides ample opportunities to develop data-driven models for building operation and control. Furthermore, it is also reported that data-driven models often outperform other approaches, especially in terms of simplicity, automation, and development engineering cost. Unlike physics-based models that are demanding on domain knowledge and computational resources, data-driven modeling process is fully automated, which can be more quickly and widely applied, thus casting a greater impact on the industry.

If we consider data to be the fuel of data-driven models, machine learning is the powerful engine. Machine learning is one of the most rapidly growing data-driven technical fields, lying at the intersection of computer science and statistics, and serving the core of artificial intelligence (AI) and data science[3]. The application of data-intensive machine-learning methods spread across science, technology, and commerce fields, leading to more evidence-based decision-making across many walks of life, including engineering, healthcare, manufacturing, education, financial modeling, policing, and marketing [3]. Machine learning can "learn" without being explicitly programmed. In other words, the mechanism in machine learning algorithms enables models to learn by themselves once the learning algorithm is determined. This makes the model become more adaptable to the uncertain and fast-changing real-world environment. Machine learning addresses the question of how to build computational programs that improve automatically through experience[3]. Computer scientist Tom M. Mitchell provided a widely quoted and formal definition of machine learning in his 1997 textbook *Machine Learning*: "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**" [4].

Though the surge of machine learning applications in building energy systems started late in the 1990s, with thirty years' development, machine learning techniques have been widely used in the building energy system modeling. In the modeling of building energy systems, machine learning techniques have been found in the following fields, but not limited to: (1) load prediction [5], (2) control and operation [6], (3) fault detection and diagnostics [7], and (4) urban-scale modeling and analysis [8].

Building load prediction has drawn special attention as it is often required in developing various strategies for improving building energy performance, e.g., fault detection and diagnosis, demand side management, energy management, deployment of distributed and renewable generation resources, microgrids, and control optimization[9]. Load prediction is hence a critical research topic in building energy system, especially grid-interactive and energy efficient building modeling. First, load prediction is a key component in advanced control and operation strategies to achieve improved building energy efficiency. It is a critical component in the cost function evaluation of model predictive control (MPC) for advanced building operation control. Second, load prediction is essential in building-grid integration, for example, demand response and transactive load control. It is the key to facilitate the interaction between building-side or demand-side and the grid, which is beneficial to generation scheduling of power system, secure and reliable operation of power plants, economic dispatch, and reliability. Third, it is also important in the early-stage building parameter design and building energy/efficiency performance analysis. Since the data-driven load prediction is essentially a regression problem, machine learning techniques have been widely applied in this field.

Machine-learning-based load prediction research has experienced explosive growth since 2010. Although many review papers covered this topic, this review paper is still in great need because: (1) despite existing review papers, there lacks a systematic paper-searching methodology to exhaust relevant papers on this topic; (2) most existing review works do not summarize the current machine learning applications in building load modeling on the data side (experience **E**); (3) most review papers are not written with respect to the logic/organization (Task **T**, Performance **P**, and Experience **E**) and language/terminologies from the machine learning perspective.

In this paper, building load prediction based on machine learning techniques is reviewed under the organization and logic of the machine learning definition by Tom M. Mitchell. Task **T** represents the application field of machine learning models. The specific application of the building load prediction model can be categorized in terms of energy load type to be predicted, application scenarios, building type, building number, and forecasting horizon. Performance measure **P** measures how well the task is performed, namely how accurate the model fits the data and how accurately it can predict. To ensure accuracy, a large number of advanced machine learning algorithms have been developed to improve performance **P**. This topic is widely researched in the building load prediction. In the definition, machine learning learns from experience **E**, namely the data, making it a data-driven, black-box, or empirical method. In building load prediction, data play an essential role throughout the whole model development process, including data collection via sensors and network, data storage and query via database, data pre-processing, feature engineering, and so forth.

Notice that we understand that there have been discussions on whether machine-learning-based system identification should belong to a broader category of machine learning, as both develop models based on data. For dynamic system modeling, trends exist to integrate techniques from these two categories together. In this paper, those studies using system identification techniques and a combination of system identification and machine learning techniques for building energy forecasting, such as [10-13], are not discussed in details.

The target readers of this paper are not only researchers from the building industry who would be exposed to cutting-edge machine learning tools, but also those from the machine learning industry who could gain the understanding of potentials and challenges in applying machine learning to building load prediction.

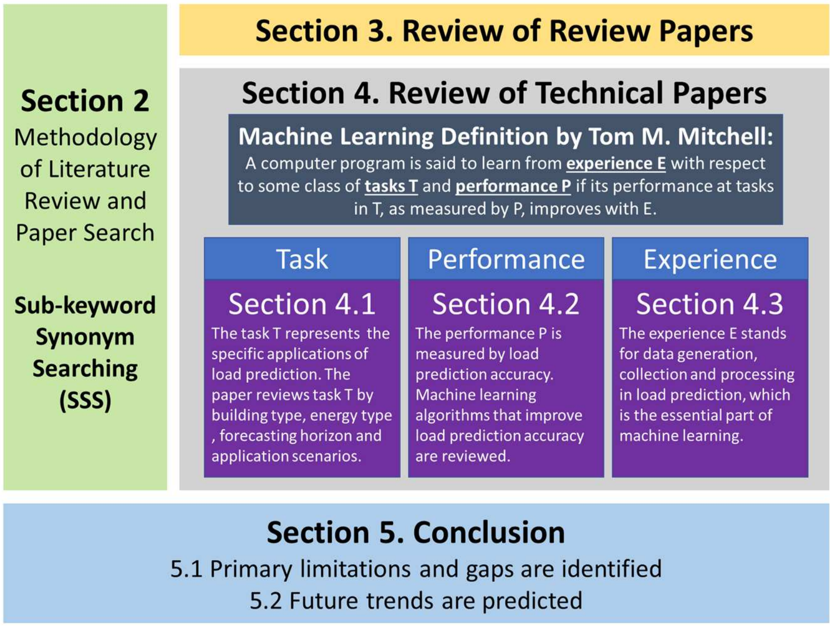

Figure 1. Content organization diagram of this review paper

More specifically, the remaining contents of this paper are organized as demonstrated in Figure 1. First, the methodology of literature review and paper search is introduced in Section 2. Then, existing review papers about machine learning load prediction and their shortcomings are summarized and discussed in Section 3. In Section 4, existing technical papers are reviewed, with a focus on papers published after year 2010, following the above definition of machine learning: building load prediction applications (Task **T**), machine learning algorithms (performance **P**), and data-related topics (experience **E**). Finally, in Section 5, we conclude with reemphasizing which topics are well-studied and which are lacking but with great potentials; we also identify the gaps in current utilization of machine learning techniques and predict future trends and development.

2. Methodology of Literature Review and Paper Search

Methodology of paper searching is critical to "exhaust" papers that are relevant to this topic. Most of review papers in this research topic do not introduce how they search for papers at all [2, 14-24]. Some paper first sets several (less than 10) keywords, reads the searched papers, and then includes additional papers cited in the searched papers [25]. Some paper uses a review protocol to down select papers step by step by using keywords to narrow the search [26]. There lacks a systematic searching methodology. To conduct a comprehensive review that captures the most important literature, we developed a searching methodology and applied it in this review paper. The methodology is called Sub-keyword Synonym Searching (SSS). The purpose of this methodology is to exhaust relevant papers by multiple searches with synonym sub-keywords. Taking this literature review of building load prediction as an example, some researchers use "energy forecast in buildings" instead of "building load prediction" in their research, but basically, these search terms are the same research field. If we only search "building load prediction" in databases such as Google Scholar, we will miss papers that use the term "energy forecast in buildings". There are many synonyms for one research field. Hence, it is necessary to develop a mechanism to exhaust the searching keywords and finally to exhaust the relevant papers. In SSS methodology, each searching keyword consists of multiple sub-keywords. There are multiple synonyms for each sub-keyword. The set of searching keywords consists of the full combination of every possible value of the sub-keyword.

In this paper, we use keywords that consist of three sub-keywords. The first sub-keyword narrows the paper to focus on building energy systems. The full list of the first sub-keyword is: "building", "HVAC", "in buildings", "cooling", and "heating". The second sub-keyword defines the energy type to be predicted. The full list of the second sub-keyword is: "energy", "load", "electricity", "consumption", "demand", "gas", and "steam". The third sub-keyword defines the action of prediction. The full list of the third sub-keyword is: "forecast", "predict", and "estimate". In this paper, Google Scholar is the main search engine of the methodology, and the full list of searching keywords in Google Scholar is the full combination of each sub-keyword. The example searching keywords are "building electricity predict", "cooling demand forecast", "electricity estimate in buildings", and so forth. The total number of searching keywords in this paper is 5 * 7 * 3 = 105 keywords. For each keyword searched in Google Scholar, the citation threshold, number of papers per search, and year range are defined. For example, in this paper, the first 10 papers per search are considered; the vintage of paper ranges from 2010 to 2020; and the citation threshold is 5. Multiple searches guided by SSS are realized by a Python module which can be found in the following GitHub repository: https://github.com/lz356/SSSS.

Table 1 summarizes the parameters of SSS methodology used in this paper. The total searched papers are: 105 keywords * 10 papers/keywords = 1,050 papers, but with duplicated papers in the search, 702 is the final number of nonduplicated papers searched in the field of "building load prediction."

Table 1. Parameters of Sub-keyword Synonym Searching (SSS) for building load prediction in this paper

Parameter	**Values**
Sub-keyword 1	building, HVAC, in buildings, cooling, heating
Sub-keyword 2	energy, load, electricity, consumption, demand, gas, steam
Sub-keyword 3	forecast, predict, estimate
Citation threshold	5
Number of papers per search	10
Year from	2010
Year to	2020

To sum up, the SSS methodology uses sub-keywords and synonyms to conduct multiple searches to comprehensively capture the most important papers in the same field. SSS makes sense because (1) different authors use different terms for the same concept and using synonyms can avoid missing papers with different terms (e.g. "building", "HVAC", and "in buildings"); and (2) SSS can cover various sub-topics (e.g. energy, load, cooling, heating, steam).

With the extracted 702 papers, we manually reviewed, categorized, and organized them according to the structure of this paper, shown in Figure 1. The Review of Review Papers in Section 3 and Review of Technical Papers in Section 4 are mostly based on these 702 papers. Many papers were removed due to irrelevancy or poor quality after the manual check. We use the following criteria to remove irrelevant papers: (a) whether the paper discuss the topic in buildings (some paper discuss the general prediction problem and some discuss other prediction problems such as grid electricity prediction) (b) whether machine learning techniques are actually applied in the paper, and (c) whether there are duplicated papers or papers discussing very similar contents, and the ones with lower citations are removed. The selected paper in this review meet all these three criteria. For some special topics like data-related topics that are not well-studied among the 702 papers, we further manually searched these topics in Google Scholar as supplementary. More searching keywords are added including "data generation/ data collection/ building automation system/ metadata schema/ semantic schema/ sensor network/ data pre-processing/ data engineering/ data science/ feature engineering/ feature extraction/ feature selection" in building load forecasting. In total, more than 200 papers are reviewed thoroughly in this paper.

3. Review of Review Papers

In this section, we first review papers on building load prediction with machine learning techniques. We used the searching method introduced in Section 2 to find the review papers from 2010 to 2019. Table 2 summarizes 15 review papers [2, 14-22, 24-28] in terms of covered model types, focuses, and summarized highlights. We categorized the review papers into two groups. The first group reviewed building load prediction models that only focus on machine learning/AI/data science [17, 18, 20, 21, 24-27]. The second group of review papers focused on building load prediction model using all kinds of modeling methods, including white-box, gray-box, and black-box. Machine learning for load prediction is only one section in these review papers. The papers in the second group are [2, 14, 16, 19, 22, 28].

All review papers listed in Table 2 focused on the algorithm side of machine learning in the building load prediction: they listed and categorized papers according to machine learning algorithms (the full algorithm list is shown in Table 6, and algorithms will be further discussed in Section 4.2); these papers also compared different algorithms and pointed out future trends of modeling algorithms. Among these papers, the papers [15, 19, 20] mentioned that hybrid methods will be a future trend of algorithm improvement. The paper written by Ahmad et al. [15] only focused on Artificial Neural Network (ANN) and Support Vector Machine (SVM). The paper written by Wang and Srinivasan [23] focused on ensemble algorithms. The paper written by Raza and Khosravi [17] focused on the architecture of ANN. Deb et al. [20] mentioned that there has not been a fair comparison of different machine leaning models.

A few papers mentioned the data side of machine learning applications: Li and Wen [2] mentioned that the key obstacle of machine learning application lies on the data side; the papers [20, 25] reviewed the data side of machine learning application, arguing that a reliable testing platform and high-quality, representative, and real-world-collected testing data to test the algorithms in different forecasting horizons are lacking. Additionally, Gerwig [28] discussed the dataset for model evaluation. The data side of machine learning in load prediction will be further reviewed in Section 4.3.

Feature/inputs/predictor is also a widely reviewed topic. Zhao and Magoulès categorized features of load prediction into four groups: ambient weather conditions, building structure and characteristics, the operation of sub-level components, occupancy, and their behavior [14]. Fumo categorized features of load prediction into five groups: building characteristics, equipment and systems, weather, occupants, and sociological influences [16]. Bilbao and Sproul concluded that dry-bulb temperature is the most pronounced climate variable used in previous regression studies; occupancy and temporal parameters are also important variables [18]. Deb et al. discussed the importance of time series variables like outdoor weather and indoor environmental conditions [20]. Lazos et al. established the importance of involving weather forecasting inputs in energy management systems by highlighting the dependencies of various building components on weather conditions [22]. Raza and Khosravi mentioned that feature selection is also one of the techniques to improve model accuracy: forecast accuracy of the model could be enhanced with better input selection of forecast models [17].

Table 2. Review papers of building load prediction using machine learning techniques

No	Author(s)	Ref	Year	Covered Model Type	Focus	Summarized Highlights
1	Zhao and Magoulès	[14]	2012	*White/black/gray-box* - Statistical regression model - Neural Networks (NN, or ANN) - Support Vector Machine (SVM)	• Algorithms • Model parameter tuning • Features/predictors	• The papers it reviewed concentrate on applying models to new predicting problems, optimizing model parameters or input samples for better performance, simplifying the problems or model development, and comparing different models under certain conditions. • AI and machine learning are developing rapidly. Many new and more powerful technologies developed in this field may bring alternatives or even breakthroughs in the prediction of building energy consumption. • The paper categorizes features of load prediction into four groups: ambient weather conditions, building structure and characteristics, the operation of sub-level components, occupancy, and their behavior
2	Ahmad et al.	[15]	2014	*ANN and SVM* - ANN - SVM - Hybrid	• Hybrid algorithms • ANN and SVM	• Among the AI methods, the ANN and SVM are the widely-used models. • Hybrid models of ANN and SVM have great potential for better model accuracy.
3	Fumo	[16]	2014	*White/black/gray-box* - Regression - ANN - SVM	• Algorithms • Features/predictors	• This paper aims to provide an up-to-date review on the basics of building energy estimation. The paper focuses on models developed with whole building energy simulation software and their validation. • The paper categorizes features of load prediction into five groups: building characteristics, equipment and systems, weather, occupants, and sociological influences
4	Raza and Khosravi	[17]	2015	*AI* - Parametric techniques (statistical techniques) - Non-parametric techniques (AI techniques)	• Algorithms • Feature selection • Model architecture • Focus on grid	• This paper shows the potential of AI techniques for effective short-term load forecasting to achieve the concept of smart grid and buildings. • Forecast accuracy of the model may be enhanced with better training of ANN, better input selection of forecast model and optimized neural network architecture
5	Bilbao and Sproul	[18]	2017	*Regression and machine learning* - Autoregressive model - ANN - SVM - Regression tree	• Algorithms • Features/predictors	• This paper reviews regression models for commercial building electricity short-term load forecasting. • Dry-bulb temperature was the most pronounced climate variable used in previous regression studies and their analysis also confirmed this finding. Occupancy and temporal parameters are also important variables.
6	Amasyali, El-Gohary	[25]	2018	*Data-driven models* - SVM - ANN - Decision trees - Other statistical algorithms	• Algorithms • Data properties • Data pre-processing • Performance measure • Temporal granularities/forecasting horizon	• This paper reviews the scopes of prediction, data properties and data pre-processing methods used, machine learning algorithms utilized for prediction, and performance measures used for evaluation. • The results of this review indicate some research areas that may require more attention: long-term building energy consumption prediction, residential building energy consumption prediction, and lighting building energy consumption prediction.
7	Daut et al.	[19]	2017	*Conventional and AI* - Stochastic time series - Regression-based approach - ANN - SVM - Swarm intelligence with AI	• Algorithms • Hybrid algorithms	• This paper states that AI is the most suitable method to manage non-linear factors as it can provide better forecasting performance. • Compared to using a single method of forecasting, the hybrid of two forecasting methods can possibly be applied for more precise results.

No	Author(s)	Ref	Year	Covered Model Type	Focus	Summarized Highlights
				method - Hybrid artificial neural network - Hybrid SVM		
8	Deb et al.	[20]	2017	*Machine learning* - ANN - Autoregressive Integrated Moving Average (ARIMA) - SVM - Case-based reasoning - Fuzzy time series - Moving average and exponential smoothing - k-Nearest Neighbor (kNN) - Hybrid	• Algorithms • Hybrid algorithms • Features/predictors	• This study presents a comprehensive review of the existing machine learning techniques for forecasting time series energy consumption. This study also discusses models that are co-analyzed with other time series variables like outdoor weather and indoor environmental conditions. The nine most popular forecasting techniques based on the machine learning platform are analyzed. • This paper presents an in-depth review and analysis of the "hybrid model", which combines two or more forecasting techniques.
9	Seyedzadeh et al.	[21]	2018	*Machine learning* - ANN - SVM - Gaussian process and mixture models	• Algorithms • Comparison among algorithms • Clustering techniques • Very detailed review on ANN	• This paper provides a substantial review on the four main machine learning 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. • In general, it is challenging to conclude which machine learning model is the best, as from literature it can be induced that all models provide reasonable accuracy by supplying large samples and optimizing the hyper-parameters. Another issue with seminal literature is that there has not been a fair comparison of different machine learning models. • Apart from modeling building energy, clustering buildings based on various input parameters remarkably facilitates and enhances energy benchmarking procedure.
10	Lazos et al.	[22]	2014	*White/black/gray-box* - Statistical forecasting approaches - Stochastic forecasts - ANN - SVM	• Algorithms • Temporal granularities/forecastin g horizon • Weather forecasting • Features/predictors	• This review investigates the multiple dimensions and value of forecasting and energy optimization algorithms for commercial building energy systems, emphasizing those with weather variable inputs. • While annual or longer horizon forecasting is certainly valuable for planning upgrades and policies, most building management systems are focusing on short term load forecasting and high-resolution dynamic optimization. • Furthermore, the importance of considering weather forecasting inputs in energy management systems is established by highlighting the dependencies of various building components on weather conditions. Difficulties in implementing integrated weather forecasts at commercial building level and the potential added value through energy management optimization are also addressed. • Finally, a novel framework is proposed that utilizes a range of weather variable predictions to optimize certain commercial building systems.
11	Li and Wen	[2]	2014	*White/black/gray-box* - ANN - Regression	• Algorithms • Black-box models lack enough training data	• This paper states that the calculation speed for statistics (black-box) can be fast enough for on-line building operation; however, they need large number training data, and the training data should cover the forecasting building operation range, which is bounded by the building operation schemes.
12	Wang and Srinivasan	[23]	2017	*AI* - Multivariate Linear	• Algorithms • Ensemble algorithms	• In this paper, the authors provide a review of AI-based building energy prediction methods with a special focus on ensemble models. The principles and applications of

No	Author(s)	Ref	Year	Covered Model Type	Focus	Summarized Highlights
				Regression (MLR) - ANN - SVM - Ensemble model - Others	• Building type • Energy type • Feature type	four main types of AI-based prediction models including multiple linear regression, artificial neural networks, support vector regression, and ensemble model have been reviewed. The advantages and limitations of each type of model are also discussed in this paper. • An intensive discussion of advantages and disadvantages of the AI-based prediction model has been carried out. Each AI-based prediction technique has its own advantages and limitations. • The ensemble model which integrates different techniques together can take over the advantages and cancel out the limitations. However, problems such as which technique should be selected as the base model and how many base models should be used to maintain the diversity of the ensemble model need to be solved in the future.
13	Gerwig	[28]	2017	*White/black/gray-box*	• Algorithms • Dataset for evaluation	• This paper provides an overview of the applied methods and points out comparable results. A structured literature review is conducted with an analysis of 375 papers categorized via a concept matrix. The paper points out which methods achieve good results for which purpose and which publicly available datasets can be used for evaluation.
14	Kuster	[26]	2017	*Machine learning* - ANN - Timeseries analysis (ARIMA, etc.) - SVM - Regression	• Algorithms • Comparison among algorithms • Model selection • Data Pre-processing	• This paper provides a systematic review protocol that provides unbiased and meaningful meta-information. • The paper argues that a direct model accuracy comparison across studies is meaningless. • A taxonomy for an informed forecasting model's selection is proposed. • Recommendations on writing electrical load forecasting paper are given
15	Molina-Solana et al.	[24]	2017	*Data science*	• Algorithms • Analysis of building operations • Detection of energy consumption patterns	• This paper summarizes that, in the area of building energy management, Data Science is now used to address problems such as the following: (1) the prediction of energy demand to adapt production and distribution; (2) the analysis of building operations as well as of equipment status and failures to optimize operation and maintenance costs; (3) the detection of energy consumption patterns to create customized commercial offers and detect fraud. • The paper reviews research with different algorithms for prediction of building energy load

Furthermore, some other supporting techniques are discussed. Seyedzadeh et al. [21] discussed the benefit of applying clustering techniques to improve model performance. Zhao and Magoulès [14] emphasized the importance of model parameter tuning. Lazos et al. [22] emphasized the importance of weather forecasting on load prediction.

To generally summarize the above-discussed review papers, here are the key take-aways: (1) the algorithm side of machine learning is the most attractive topic in review, i.e., most review paper is organized in terms of algorithms; (2) data side of machine learning is receiving increasing attentions; (3) supporting techniques, like feature selection, extraction, clustering, and weather forecasting are mentioned as well. Although many review papers can be found on this topic, this review paper is still in great need because: (1) among existing review papers, there lacks a systematic paper-searching methodology to exhaust relevant papers on this topic; (2) most existing review works do not summarize the current machine learning applications in building load modeling on the data side (experience **E**); (3) most review papers are not written with respect to the logic/organization (Task **T**, Performance **P**, and Experience **E**) and language/terminologies from the machine learning perspective.

4. Review of Technical Papers

Reviewing the review papers in the previous section provides a general overview of the current development of machine learning techniques for building load prediction. In this section, technical papers are reviewed. Section 4.1 summarizes the major applications of load prediction model (Task **T**). Section 4.2 summarizes the machine learning algorithms with associated approaches to improve their performance and accuracy (performance **P**). Section 4.3 focuses on the data perspective (experience **E**), including data sources and generation, data pre-processing, and feature extraction and selection. Section 4.4 discusses other related topics.

4.1 Applications (Task **T**)

The applications can be further categorized into four aspects: (1) the energy type of load to predict; (2) the application scenarios of load prediction model; (3) building type; and (4) forecasting horizon. Section 4.1 is organized according to the categorization of applications above.

In terms of different types of load, as shown in Table 3, the applications of load prediction model can be classified as (1) cooling load/energy prediction, (2) heating load/energy prediction, (3) HVAC load/energy prediction, (4) district heating load/energy prediction, (5) primary energy load/consumption prediction, (6) natural gas consumption prediction, (7) electricity load/consumption prediction, (8) steam load prediction. Wang and Srinivasan [23] also reviewed the papers in terms of energy type: building energy, heating, and cooling, heating, cooling, and others.

Table 3. Energy types to predict in machine-learning-based building load prediction

Applications	**Example Literatures**
Cooling load/energy prediction	[5, 29, 30]
Heating load/energy prediction	[29, 31]
HVAC load/energy prediction	[32, 33]
District heating load/energy prediction	[34, 35]
Primary energy load/consumption prediction	[36, 37]
Natural gas consumption prediction	[38, 39]
Electricity load/consumption prediction	[11, 38-43]
Steam load prediction	[44]

In terms of application scenarios of load prediction model, there are three major categories: (1) model predictive control/demand response/control optimization, (2) building parameter design/retrofit, (3) building energy planning/climate change impact. The forecasting horizon has three categories: (1) hour(s); (2) day(s); (3) month(s) or longer. Wang and Srinivasan [23] also categorizes forecasting horizon to hour, day, year, and other. This categorization is better than categorizing forecasting horizon into short-term, medium-term, and long-term categorization like what Ahmad and Chen did [45], because very few load prediction papers falling into the long-term category using that categorization. The object of load prediction has two categories in terms number of buildings: single building and multiple buildings. This paper also categorizes target building types as commercial and residential buildings, although some review paper have categorized building types to commercial, residential, education, and other [27]. Table 4 and Table 5 present the commercial and residential building load prediction application summary table, respectively.

Overall, facilitating the decision-making process on model predictive control/demand response/control optimization is the major purpose for most of load prediction research. In this field, research covers both single and multiple buildings. Most of the papers predict load in the forecasting horizon of hours and days. The load forecasting model is critical in the building-grid-integration by guiding a more efficient the power plant production scheduling to meet building needs and more accurate building reactions to the electricity price. Under this premise, load prediction is widely used in the model predictive control, where building load prediction model serves as the cornerstone. The application of model predictive control here is used not only for optimization of building operations, but also for demand response and building-grid integration purpose.

Regarding two other applications, building parameter design/retrofit and building energy planning/climate change impact, on the other hand, load predictions are in the forecasting horizon of months and years with the focus of the long-term energy impact evaluation. The inputs used for these two applications usually include building characteristics. Specifically, Building parameter design/retrofit" is the second most widely studied topics using machine learning-based building load prediction since the data-driven methods, by building the direct relationship between design parameters and building energy consumption, help alleviate the heavy computational burden and engineering efforts required in the design evaluation, especially in the early building design phase. Finally, the application of machine learning based building prediction on "Building energy planning/climate change impact" is a more macro study of long-term impact from energy planning and climate change on building energy consumption.

The key difference between these three applications are their building load prediction resolution, i.e., "Model predictive control/demand response/control optimization" focused on the short-term load prediction that reflect the dynamics of the energy transfer inside a building and between building and the grid since real-time control signals require to be in a high time resolution . "Building parameter design/retrofit" requires a relatively longer prediction horizon because in most of the time, the design is to optimize the building design parameters on annual, seasonal, or monthly energy consumption pattern. And "Building energy planning/climate change impact" has the longest prediction horizon, which is normally years and even decades of time to study and predict the energy change in a long term perspective: it is more related to policy making and long-term energy planning.

Table 4. Summary of commercial building machine-learning-based load prediction applications

Applications	**Single / Multiple Building(s)**	**Forecasting Horizon**	**Example Reference**
Model predictive control/demand response/control optimization	Single building	Hour(s)	[5, 30, 32, 33, 46-49]
		Day(s)	[37, 43, 45, 49-54]
		Month(s)	-
	Multiple buildings	Hour(s)	[27, 55, 56]
		Day(s)	[34, 35, 44, 56-58]
		Month(s)	-
Building parameter design/retrofit	Single building	Month(s)	[29, 59-61]
	Multiple buildings		NA
Building energy planning/climate change impact	Single building	Month(s)	[39, 62]
	Multiple buildings		[63]

Table 5. Summary of residential building machine-learning-based load prediction applications

Applications	**Single / Multiple Building(s)**	**Forecasting Horizon**	**Example Reference**
Model predictive control/demand response/control optimization	Single building	Hour(s)	[11, 31, 38, 41, 64]
		Day(s)	[11, 42, 45, 54]
		Month(s)	[11]
	Multiple buildings	Hour(s)	[56, 65]
		Day(s)	[35, 56]
		Month(s)	-
Building parameter design/retrofit	Single building	Month(s)	[36, 61]
	Multiple buildings		[66]
Building energy planning/climate change impact	Single building	Month(s)	[40, 62]
	Multiple buildings		-

4.2 Algorithms (Performance **P**)

The topic of machine learning algorithms is widely studied and reviewed in the existing literature. Table 6 summarizes most of machine learning algorithms used for building load prediction from both review and technical papers. There are 9 major types of machine learning algorithms used for building load prediction. They are: (1) Linear Regression, (2) Support Vector Machine (SVM), (3) Neural Network (NN), (4) Deep Learning, (5) Tree-based Algorithms, (6) Hybrid, (7) AutoRegressive methods, (8) fuzzy timeseries model, and (9) Others. It is worth mentioning that some methods applied for grid load prediction are also included in this table due to the similarity between grid load prediction and building load prediction. No duplicated algorithm is listed in Table 6, and a total of 128 different algorithms are reviewed in the table.

4.2.1 General Review of Algorithms

Traditional machine learning algorithms, such as artificial neural network (ANN, or neural network, NN) and support vector machine (SVM, or support vector regression, SVR), have been extensively studied in the short-term building load prediction. Ahmad et al. [15] reviewed only ANN and SVM algorithms in building load prediction. Nonetheless, many researchers have utilized new algorithms to further improve the performance. Seyedzadeh et al. [21] provided a very detailed review on ANN. The development of the application of Neural Network on load prediction has a tendency toward two directions: (1) parameter and structure optimization, and (2) hybridization with other machine learning algorithms. Raza and Khosravi [17] summarized the trends of ANN, noting that it is increasingly applied along with other algorithms, such as expert system and regression technique, wavelet and time series, support vector machine and artificial immune system. ANN is also focused more on the optimization techniques of model parameter and structure, using genetic algorithm, particle swarm optimization, and so on. SVM is also applicable to the two directions mentioned above as Neural Network.

Tree-based method has a flow-chart-like structure, where each internal node denotes a test on an attribute, and each branch represents the outcome of a test. The topmost node in a tree is the root node, and the downmost node is the leaf node. Yu et al. [36] proposed a decision tree method (C4.5) for building energy modeling. The results have demonstrated that the use of C4.5 algorithm can classify and predict building energy demand levels accurately (93% for training data and 92% for test data), identify and rank significant factors of building energy use intensity levels automatically, and provide the combination of significant factors as well as the threshold values that contribute to high building energy prediction performance. Chou and Bui [67] applied a regression tree algorithm (Chi-squared Automatic Interaction Detector, or CHAID) to forecast the short-term cooling and heating load.

Deep learning is a class of machine learning algorithms that use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation [68]. As an evolution of artificial neural network (ANN)-based prediction methods, deep learning is expected to increase the prediction accuracy by higher levels of abstraction, better scalability, and automatic hierarchical feature learning. Mocanu et al. investigated two newly developed stochastic models for time-series short-term prediction of energy consumption, namely Conditional Restricted Boltzmann Machine (CRBM) and Factored Conditional Restricted Boltzmann Machine (FCRBM). In the comparison study, the results showed that FCRBM outperformed Artificial Neural Network, Support Vector Machine (SVM), Recurrent Neural Networks (RNN) and CRBM [11]. Mocanu et al. also incorporated the Deep Belief Network with automated feature extraction for the short-term building energy modeling process [69].

A hybrid algorithm combines two or more other algorithms in solving the same problem, either choosing one (depending on the data), switching between them over the course of the algorithm, or combining predictions from different algorithms using weights. Jovanović et al. integrated FFNN, RBFN, and ANFIS in one model to predict short-term heating consumption. Results showed that the hybrid algorithm, by combining the outputs of member networks, achieved better prediction results [34]. It is worth mentioning that some papers mistook ensemble method as hybrid method. Here, we use a definition introduced by Kazienko et.al, which denotes the difference: ensemble classifiers combine multiple but homogeneous weak models, while hybrid methods combine completely different heterogeneous machine learning approaches [70].

Ensemble learning in this paper refers to the integration of homogeneous learning algorithms. The ensemble model is defined as an approach using multiple homogeneous learning algorithms/models to obtain better predictive performance than that could be obtained from any of the constituent learning algorithms/models [71]. The advantage of the ensemble prediction method lies in its remarkably improved prediction accuracy and stability [27]. Wang and Srinivasan [23] comprehensively reviewed ensemble methods in building energy prediction.

Autoregressive algorithms are traditional methods for load prediction. The autoregressive model's output depends linearly on its previous values and a stochastic term. Some traditional algorithms such as Autoregressive Model (AR) [72], Autoregressive Exogenous Model (ARX) [73], AutoRegressive–Moving-Average (ARMA) [74], AutoRegressive Integrated Moving Average (ARIMA) [75], AutoRegressive Integrated Moving Average with eXternal input (ARIMAX) [75] are applied to load prediction. New autoregressive algorithms are also applied to load prediction, including Double Seasonal ARIMA [76], Analytical Hierarchy Process [77], Novel Box Jenkins method [78], Periodic Autoregression [79], Seasonal Autoregressive Integrated Moving Average with Exogenous Variables model [80], ARMAX with PSO [81], and Fuzzy AutoRegressive Moving Average with eXogenous input variables [82]. Most of machine learning algorithms for load prediction are time-invariant with decreased accuracy in severe dynamic situations, especially for building with heavy envelope and thermal inertia. Obviously, the best way to treat dynamics is to use time-serious models like autoregressive exogenous (ARX), autoregressive integrated moving average (ARIMA), autoregressive–moving-average model (ARMA), etc. Li and Huang evaluate four data-driven models and found autoregressive Moving Average with Exogenous inputs (ARMAX) is superior to ANN, MLR, and RC network (resistor-capacitor network) in prediction accuracy when predicting short-term building cooling load [83]. Chou and Ngo proposed a system that integrates a seasonal autoregressive integrated moving average (SARIMA) model and metaheuristic firefly algorithm-based least squares support vector regression (MetaFA-LSSVR) model. The prediction system yielded high and reliable accuracy rates in 1-day-ahead predictions of building energy consumption with a total error rate of 1.18%[84].

Other sophisticated predictive algorithms such as Bayesian Networks (BNs) [85], Extreme Learning Machine (ELM) [61] , Case-based Reasoning (CBR) [86] , Meta learning[87], k-Nearest Neighbors (kNN) [88-90], Gaussian process and mixture models [91, 92], and fuzzy timeseries algorithm [80, 93-98], are found to be used in the literatures on building load prediction.

Table 6. List of machine learning algorithms in review and technical papers of building load prediction

No	Type	Algorithm	Abbreviation	Example Reference
1	Linear Regression	Multiple Linear Regression	MLR	[41]
2		General Linear Regression	GLR	[67]
3		Polynomial Regression	PR	[72]
4		Exponential Regression	ER	[72]
5		Hierarchical Mixture of Experts with Linear Regression	HME-REG	[41]
6		Multiple Linear Regression Self-Regression	MRL-SR	[99]
7		Conditional Demand Analysis	CDA	[100]
8		Forward Stepwise Regression	FSR	[101]
9		Fuzzy Regression	FR	[102]
10		Multivariate Adaptive Regression Spline	MARS	[103]
11	Support Vector Machine (SVM)	Support Vector Machine with Gaussian Radial Basis Function	SVM-RBF	[40]
12		Simple Support Vector Machine	SVM	[104]
13		Parallel Support Vector Machine	PSVM	[105]
14		Support Vector Machine with Pearson VII Universal Kernel	PUK-SVM	[47]
15		Least Squares Support Vector Machine	LSSVM	[41]
16		Weighted Support Vector Machine	WSVM	[106]
17		Fuzzy Support Vector Machines	FSVM	[107]
18		Support Vector Regression with Differential Evolution	DE-SVM	[108]
19		Self-Organized Map and Support Vector Machine	SOM-SVM	[109]
20		Support Vector Machine with Particle Swarm Optimization	PSO-SVM	[110]
21		Support Vector Machine with Simulated Annealing Particle Swarm Optimization	SAPSO-SVM	[111]
22		Support Vector Machine with Ant Colony Optimization	ACO-SVM	[112]
23		Support Vector Machine with Optimal Training Subset and Adaptive Particle Swarm Optimization	OTS-APSO-SVR	[113]
24		Particle Swarm Optimization and Least Squares Support Vector Machine	PSO-LSSVM	[114]
25		Support Vector Regression with Chaotic Gravitational Search Algorithm	CGSA-SVR	[83]
26		Seasonal Recurrent Support Vector Regression with Chaotic Artificial Bee Colony	CABC-SRSVR	[115]
27		Hybrid Particle Swarm Optimization with Genetic Algorithm Mutation to optimize Support Vector Machine	HPSO-GA-SVM	[116]
28		Recurrent Support Vector Machines with Genetic Algorithms	GA-SVM	[117]
29		Support Vector Regression with Chaotic Genetic Algorithm	CGA-SVR	[118]
30		Support Vector Regression with Modified Firefly Algorithm	MFA-SVR	[119]
31		Support Vector Regression with Krill Herd Algorithm	KH-SVR	[120]
32		Support Vector Regression and Chaotic Particle Swarm Optimization	CPSO-SVR	[121]
33		Least Squares Support Vector Machine with Group Method of Data Handling	GMDH-LSSVM	[122]
34		Support Vector Machine and Stimulated Annealing Particle Swarm Optimization	SAPSO-SVM	[123]
35	Neural Network (NN)	Feed Forward Neural Network/Back propagation neural network	FFNN/BPNN	[32, 41]
36		Multilayer Perceptron Neural Network	MLP-NN	[47]
37		Simple Artificial Neural Network	NN	[124]
38		General Regression Neural Network	GRNN	[125]
39		Levenberg Marquardt Artificial Neural Network	LM-NN	[126]
40		Self-Recurrent Wavelet Neural Network	SRWNN	[56]
41		Wavelet Neural Networks	WNN	[127]
42		Weighted Evolving Fuzzy Neural Network	WEFuNN	[128]
43		Radial Basis Function Neural Network	RBF-NN	[129]
44		Multi-RSAN	MRSAN	[130]
45		Fuzzy C-Means with Feed Forward Neural Network	FCM-FFNN	[41]
46		Hierarchical Mixture of Experts with Feed Forward Neural Network	HME-FFNN	[41]
47		Probabilistic Entropy-Based Neural Network	PE-NN	[30]

No	Type	Algorithm	Abbreviation	Example Reference
48		Artificial Neural Network with Bayesian Regularization	BR-NN	[43]
49		Elman Neural Network	IE-NN	[27]
50		Genetic Algorithm and ANN	GA-ANN	[131]
51		Hybrid Genetic Algorithm-Hierarchical Adaptive Network-Based Fuzzy Inference System	GA-HANFIS	[54]
52		ANN and Hybrid Genetic Algorithm-Adaptive Network-Based Fuzzy Inference System	GA-ANFIS	[132]
53		Neural Fuzzy Network with improved Genetic Algorithm	GA-NFN	[133]
54		Genetic Algorithm and Particle Swarm Optimization Multi-Layer Perceptron	GA-PSO-MLP	[134]
55		Group Method of Data Handling Neural Network	GMDH-NN	[135]
56		Adaptive Network Based Inference System	ANFIS	[29]
57		Improved Elman Neural Network and Novel Shark Smell Optimization	NSSO-IENN	[55]
58		Neural Network with Particle Swarm Optimization	PSO-NN	[136]
59		Neural Network with Chaotic Particle Swarm Optimization	CPSO-NN	[137]
60		Radial Basis Function Neural Network with Particle Swarm Optimization	PSO-RBF-NN	[138]
61		Radial Basis Function Neural Network with Ant Colony Optimization	ACO-RBF-NN	[139]
62		Fuzzy Neural Network with Particle Swarm Optimization	PSO-FNN	[140]
63		Back Propagation Neural Network and Self-Adapting Particle Swarm Optimization	SPSO - BPNN	[141]
64		Forward Feedback Neural Network with Adaptive Particle Swarm Optimization	APSO - FFNN	[142]
65		Radial Basis Neural Network with Nonlinear Time-Varying Evolution Particle Swarm Optimization	NTVEPSO-RBF-NN	[143]
66		Self-Organizing Fuzzy Neural Network with a Bilevel Optimization	BO- SOFNN	[144]
67		Radial Basis Neural Network with Nonlinear Time-Varying Evolution Particle Swarm Optimization	NTVE-PSO-RBF-NN	[143]
68		Neural Network with Estimation of Distribution Algorithms	EDA-NN	[145]
		Generative Adversarial Network	GAN	[146]
69	Deep Learning	Extreme Gradient Boosting	XGB	[5]
70		Long Short-Term Memory	LSTM	[42]
71		Conditional Restricted Boltzmann Machine	CRBM	[11]
72		Factored Conditional Restricted Boltzmann Machine	FCRBM	[11]
73		Long Short-Term Memory based Deep Recurrent Neural Network	LSTM-DRNN	[62]
74		Deep Belief Network	DBN	[69].
75	Tree-based Algorithms	Random Forest	RF	[32]
76		Classification and Regression Tree	CART	[67]
77		Bagged Tree	BaggedT	[45]
78		Boosted Tree	BoostedT	[45]
79		Chi-Squared Automatic Interaction Detector	CHAID	[67]
80		C4.5	C4.5	[36]
81	Hybrid	Neural Network with Support Vector Machine	NN-SVM	[147]
82		AutoRegressive Integrated Moving Average with Support Vector Machine	ARIMA-SVM	[148]
83		Neural Network and Fuzzy Expert System	FES-NN	[149]
84		Back Propagation Neural Network with Artificial Immune System	AIS-BPNN	[150]
85		Kernel Regression with k-Nearest Neighbors	KR-kNN	[151]
86		Feed Forward Neural Network, Radial Basis Function Network and Adaptive Network Based Inference System	FFNN-RBFN-ANFIS	[34]
87		Polynomial and Exponential Regression	PR-ER	[72]
88		AutoRegressive Integrated Moving Average and Back Propagation Neural Network	ARIMA-BPNN	[152]

No	Type	Algorithm	Abbreviation	Example Reference
89		Support Vector Regression with Autoregressive Model	AR-SVM	[153]
90		Real-Valued Genetic Algorithm-Based Neural Network with Support Vector Machine	RGA-NN-SVM	[147]
91		Hybrid Quantized Elman Neural Network	HQENN	[154]
92		Wavelet Transform, Adaptive Genetic Algorithm and Fuzzy System with Generalized Neural Network	WT-GA-FS-GNN	[155]
93		Self-Organizing Map, Support Vector Regression and Fuzzy Inference System	SOM-SVR-FIS	[127]
94		Wavelet Transform-Neural Network	WT-BNN	[156]
95		Wavelet Decomposition, Neural Network, and Genetic Algorithm	WD-GA-NN	[157]
96		Neural Network with Nonlinear Autoregressive with eXogenous input structure	NARX-NN	[158]
97		Neural Network and AutoRegressive Integrated Moving Average	ARIMA-ANN	[159]
98		MLR, ARIMA, SVR, RF, MLP, BT, MARS, kNN	MLR-ARIMA-SVR-RF-MLP-BT-MARS-kNN	[37]
99	Autoregressive Methods	AutoRegressive Integrated Moving Average	ARIMA	[75]
100		Double Seasonal ARIMA	DS-ARIMA	[76]
101		AutoRegressive–Moving-Average	ARMA	[74]
102		AutoRegressive Integrated Moving Average with eXternal input	ARIMAX	[75]
103		Analytical Hierarchy Process	AHP	[77]
104		Novel Box Jenkins Method	Box-Jenkins	[78]
105		Periodic Autoregression	PAR	[79]
106		Autoregressive Model	AR	[72]
107		Autoregressive Exogenous Model	ARX	[73]
108		Seasonal Autoregressive Integrated Moving Average with Exogenous Variables model	SARIMAX	[80]
109		ARMAX with PSO	PSO-ARMAX	[81]
110		Fuzzy AutoRegressive Moving Average with eXogenous Variables	FARMAX	[82]
111	Extreme Learning	Online Sequential Extreme Learning Machine	OSELM	[61]
112	Bayesian Networks	Bayesian Networks	BNs	[85]
113	Case-Based Reasoning	Case-Based Reasoning	CBR	[86]
114	Meta Learning	Meta Learning	ML	[87]
115	k-Nearest Neighbors (kNN)	Multivariate kNN Regression	kNN	[88]
116		Maximum Length Weighted Nearest Neighbor	MLWNN	[89]
117		Hybrid method based on Wavelet Transform, Triple Exponential Smoothing model, and Weighted Nearest Neighbor	WT-TES-WNN	[90]
118	Gaussian Process and Mixture Models	Gaussian Process	GP	[91]
119		Gaussian Mixture Model	GMM	[92]
120	Ensemble	Neural Network Ensemble	NN ensemble	[34, 51]
121		MPL Ensemble	MPL ensemble	[160]
122	Fuzzy Timeseries Algorithms	Fuzzy Time Series with Singh's method	SFTS	[93]
123		Linguistic Out-Sample Approach for Fuzzy Time Series	LOA-FTS	[94]
124		Slide Window Fuzzy Time Series	SW-FTS	[95]
125		Fuzzy Interface System	FIS	[80]
126		Fuzzy Polynomial Regression	FPR	[96]
127		Fuzzy Linear Regression	FLR	[97]

No	Type	Algorithm	Abbreviation	Example Reference
128		Refined High-Order Weighted Fuzzy Algorithm and Imperialist Competitive Algorithm	ICA-RHWFTS	[98]

4.2.2 Parameter Tuning/Optimization

As mentioned in Section 4.2.1, an increasing number of researchers focus on improving parameter identification/optimization process utilizing parametric algorithms with increased accuracy. Yokoyama et al. predicted short-term hourly building energy demand using a combination of back propagation neural network and system identification. A global optimization method called "Modal Trimming Method" was proposed for non-linear programming problems and adopted to identify neural network parameters [161]. Zhang et al. developed a weighted multi-SVR model to forecast the energy consumption of an institutional building. Differential evolution (DE) algorithm was found to be effective to optimize the model parameters and their weights in the experiment [49]. Li et al. developed a novel methodology that hybridized genetic algorithms (GAs) and support vector regression (SVR) and implemented this model in forecasting hourly cooling load. To build an effective SVR model with accurate prediction and generalization ability, real value GAs were adopted to automatically determine the optimal hyper-parameters for SVR [162]. As shown in Table 6, the most widely used optimization algorithms are: Differential Evolution (DE), Self-organized Map (SOM), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Ant Colony Optimization (ACO).

4.2.3 Clustering of Building Operation Data

Clustering building operation data before applying machine learning algorithms can reduce the data variation associated with complicated building operation, thus improving performance of each sub-model and finally overall model performance [51, 107, 160, 163]. Clustering techniques are also used for recognizing the pattern of building energy use [64, 164-167], outlier detection [37, 168], identification of occupancy behavior [169] and classification of lighting system [170]. Table 7 summarizes the papers utilizing clustering to support machine learning algorithms for load prediction. Clustering technique is becoming widely applied in the existing studies as an important supporting technique to aid machine learning in solving real-world problems.

Table 7. Summary of clustering techniques applied to support machine learning for load prediction

No	Clustering Algorithm	Contents	Example Reference
1	k-means	Cluster data and build sub-models for each data cluster to improve prediction accuracy	[160]
2	k-means	Cluster load and temperature to improve the performance of ensemble neural network	[51]
3	Ward's linkage	A clustering technique was used to determine various daily schedule pattern types for an individual household	[64]
4	k-means	Cluster data to improve neural network performance	[163]
5	Fuzzy C-mean	Cluster data to improve fuzzy support vector machines	[107]
6	Entropy-weighted k-means (EWKM)	Detect abnormal building energy consumption	[37]
7	Hierarchical, k-means, fuzzy k-means	Improved performance in the classification of customers and generation of electricity load profiles	[164]
8	k-Shape	Building energy usage patterns analysis	[165]
9	Density-based spatial clustering of applications with noise (DBSCAN)	Identify outliers and diurnal schedules	[168]
10	Fuzzy clustering	Building energy classification	[166]
11	K-means	Classify occupancy behavior	[169]
12	Expectation–Maximization (EM) algorithm	Classify lighting system	[170]
13	Entropy-weighted k-means (EWKM)	Cluster analysis identifies three daily power consumption patterns.	[167]

4.2.4 Extract Interpretation from Data-Driven Models

One of the major weaknesses of data-driven and machine learning algorithm is that the model itself is difficult to interpret because it is built purely on data without any physical meaning and explanation. To solve this problem, some researchers are trying to endow data-driven model with comprehensibility and interpretation. Che et al. found that fuzzy rule-based forecast systems introduced both accuracy and comprehensibility of the forecast result at the same time. They combined SOM with fuzzy membership function concepts. The key idea behind the combination is to build a human-understandable knowledge base by constructing a fuzzy membership function for each homogeneous sub-population. The comparison of different mathematical models and effectiveness of the presented model were shown by the real data of New South Wales electricity market, which confirmed the validity of the developed model [127]. Che introduced a multiple linear regression model that treated all seasonal cycles as input attributes. The result helps managers to interpret the series structure with multiple seasonal cycles [113].

4.2.5 Conclusion of Machine Learning Algorithm Development in Building Load Prediction

Our extensive literature review focusing on the latest research of short-term data-driven building load prediction presents the following trends:

(a) With the development of machine learning and information science, more novel algorithms and modeling approaches are applied for building load prediction. According to the literature review, Table 6 shows that more than one hundred machine learning algorithms have already been applied in building energy prediction/analysis.

(b) Traditional algorithms like ANN and SVM are combined with clustering algorithms, sophisticated parameter optimization, rule-based modeling, and other data-driven algorithms (hybrid algorithms) to increase prediction accuracy.

(c) Improving comprehensibility and interpretability of modeling process is another important topic of short-term building load prediction modeling. Many approaches are brought up to reflect dynamic, periodic, and non-linear characteristics of building energy consumption.

What was well-studied: (1) increasingly sophisticated machine learning algorithms were applied for building load prediction; and (2) more domain knowledge was applied in machine learning models, making the design of machine learning algorithm become "physics-based" and robust to data quality.

What is lacking: (1) Dataset and testbed to evaluate algorithms. High-quality and diversified testbeds are lacking for algorithm performance evaluation. Although ASHRAE - Great Energy Predictor III [171], the Kaggle competition: https://www.kaggle.com/c/ashrae-energy-prediction, has provided the training/testing dataset, researchers still need more real-world-collected and high-quality testbeds to evaluate algorithm performance. Tested under standardized experimental environment (for example, Flexible Research Platform at Oak Ridge National Lab), the conclusion of the algorithm performance will be more convincing to be applied in practice. (2) Most of papers demonstrated the effectiveness of their algorithms by comparison with other algorithms. However, although some papers introduce the tuning process of their developed algorithms in detail, most papers do not explicitly mention how other algorithms for comparison were tuned. Certainly, not introducing the tuning of other comparison algorithms in the paper cannot prove that authors didn't do the tuning for the comparison algorithms, but we can see from the literature review that most authors do not spend as much effort in introducing the tuning part of work as they did for the proposed algorithm, making the comparison less strict and the results less convincing to readers..

4.3 Data-Related Topics (Experience **E**)

If machine learning is a powerful engine for modeling, then data are the fuel. The topic of applying machine learning to building load prediction is widely studied as discussed in Section 4.2; however, while most of the studies carefully reviewed load prediction on the algorithm side, discussion from the data perspective is often neglected.

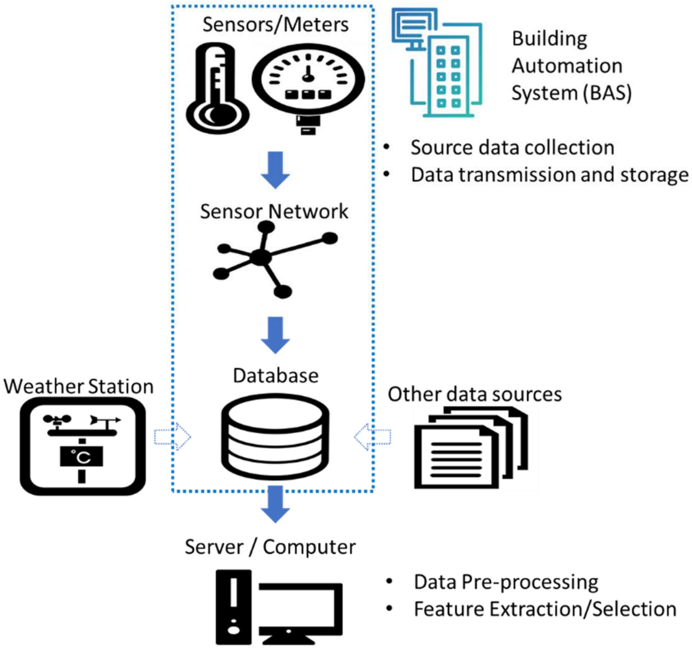

Figure 2. Diagram of building data engineering procedures for machine learning applications

Advanced machine learning algorithms need data support with better quality and larger quantity. The procedures from data generation by hardware to the final inputs of machine learning model have 4 steps as shown in Figure 2: (1) data collection from sensors/meters, (2) data transmission and storage through the sensor network and database, (3) data querying and pre-processing, and (4) feature extraction and selection to for machine learning models. The data side and algorithm side are equally important, in spite that the data side is more time consuming and expensive, in terms of engineering efforts. In most of machine learning applications, works on the data side account for about 80% of the modeling time ("80-20 rule" [172]). In existing literatures, academic researchers often focus more on the algorithm side, while industrial researchers and field engineers actually focus more on the data side. The general diagram of the data engineering process can be summarized in the following diagram.

Most of data used for load prediction are collected from BAS or Building Management System (BMS), which is a computer-based control system installed in buildings to control and monitor the building mechanical and electrical equipment, such as ventilation, lighting, power systems, fire systems, and security systems within a local area network [173]. Despite common unavailability in the BAS, the weather data need to be recorded due to the significant impact of weather on building energy consumption. Other data sources, like building parameters, economic indicators, Geographic Information System, and other independent sensors (e.g., Wi-Fi connection to reflect occupancy) are also useful in the building load prediction.

Collected data are stored in one or multiple databases and can be queried via computers or servers. The enormous amount of sensor data can be exploited to increase energy efficiency. However, the sheer quantity of the data poses a challenge at various levels for traditional data analysis approaches. New infrastructure and tools need to be developed to accommodate what is now known as Big Data [24].

Data pre-processing is the next step after data are queried from database. It deals with raw sensor data that are incomplete, faulty, and unstable. Data cleaning is essential before sensor data can be used. The final steps before the machine learning training are feature (or input) extraction and feature selection. Feature extraction is a technique that further digs important information from raw sensors or measurements. Feature selection is another important technique to select useful information from large quantities of sensor data to avoid curse of dimensionality.

In Section 4.3, data generation and collection are firstly reviewed. Secondly, data pre-processing topics are discussed. Finally, feature extraction and selection techniques are reviewed.

4.3.1. Data Generation and Collection

Data generation and data collection are essential components because data are the backbone of machine learning models. In this section, data generation and data collection are reviewed in the sensor level, system level, and data level. Throughout the entire process of data generation and collection, sensor level is the first stage where data are logged, generated, and collected by hardware, including sensors and meters. System level is the middle stage that transfers, integrates, formalizes, and stores collected sensor data from the sensor level via sensor networks, BAS, IoT, and so forth. Data level is the final stage where data are queried, cleaned, processed, extracted, and readied to be used in machine learning modeling. Figure 3 shows a diagram of the review structure in this sub-section.

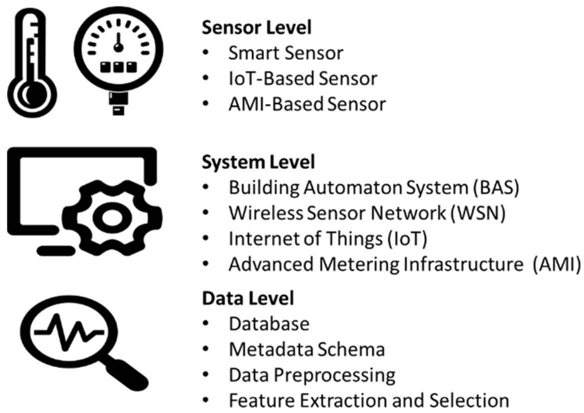

Figure 3. Diagram of review structure for data generation and collection

4.3.1.1 Sensor Level:

Ahmad et al. [174] reviewed energy metering and environmental monitoring technologies along with a discussion on their working principles, types, and cost comparison of the most widely available meters and sensors. The paper also reviewed social, economic, environmental, and legislative drivers for the installation of these technologies. Then, it presented a table to summarize energy legislation and regulations in different countries. Factors influencing the selection of meters and sensors were also discussed in detail. It reviewed the most widely used communication and network technologies. Finally, it discussed possible future research directions and challenges. Plageras et al. [175] demonstrated how IoT-based smart sensors can empower the data collection, processing, and analysis in smart buildings and smart cities.

Occupancy sensor is one of the most widely discussed sensor types because of the increasing importance of occupancy-based control in buildings. Yang et al. [176] reviewed the occupancy sensing systems and occupancy modeling methodologies in buildings and listed the pros and cons for the application of occupancy sensors in institutional buildings. Zhao et al. [177] developed a generic, feasible and low cost occupancy sensing solution to provide reliable real-time occupancy information in buildings. Zou et al. [178] presented the design and implementation of a novel and practical occupancy sensing system, WinOSS, which can provide fine-grained occupancy information by leveraging existing commodity Wi-Fi infrastructure along with the Wi-Fi-enabled mobile devices carried by occupants.

4.3.1.2 System Level:

Zhou et al. [179] discussed the sources and characteristics of energy big data. In smart grid, the main source of data is the advanced metering infrastructure (AMI), which is one of the underlying enabling technologies of smart grid. AMI deploys a large number of smart meters and other measuring terminals at the end-user side. Many other intelligent devices like BAS, sensors, and thermostats, deployed throughout the whole process of power generation, transmission, distribution, substation, also collect vast quantities of data. Weather data, such as solar radiation, wind speeds, and outdoor air dry-bulb temperature, play an important role in supporting smart energy management. In addition, the Geographic Information System data are also an integral part of energy big data.

On the system side, there is a trend of moving from traditional BAS system with a local area network, to wireless sensor network (WSN), and finally to the IoT. Traditional BAS uses wired technologies where sensors and actuators are connected to controllers. The first modern BAS deployed a point-to-point control that was later replaced by a centralized bus. Modern BAS uses the distributed bus control, also called field buses, and all system nodes have a built-in control unit that enables them to act without a centralized control [180]. The Wireless Sensor Networks (WSN) consists of tiny resource-limited devices called sensor nodes. Further, nodes have an on-board low-power radio and a micro-controller. Sensor nodes are usually battery driven. The measured sensor data are transported in a multi-hop fashion to a base station [180]. WSNs are playing an increasingly important role. Furthermore, WSNs are characterized by high heterogeneity because there are many different proprietary and non-proprietary solutions. The current trend, however, is to move away from proprietary and closed standards, to embrace IP-based sensor networks using the emerging standard 6LoWPAN/IPv6. This allows native connectivity between WSN and Internet, enabling smart objects to participate in the IoT [181]. With the development of IoT, cloud computing [182] originates in the concept of grid computing, whose goal is to reduce computation costs and increase the flexibility and reliability of systems by aggregating computing resources. A new issue of IoT and cloud computing is the security and privacy concern [24].

Domingues et al. [183] systematized fundamental aspects of BAS, contributing to a common understanding of fundamental building automation concepts aligned with the well-known standards ISO 16484-3 and EN 15232. Using these standards as guidelines, this work highlighted the scope of the functionality to be expected from the typical BAS. Another contribution of this work is the assessment of the industry's standard building automation technologies in terms of functional requirements employing a uniform terminology. Moreover, the study also provided a detailed mapping of features between existing technologies, according to their official specifications, and identified their functionality gaps. It becomes clear that no single technology can cover all functions expected from a BAS, thus requiring posterior, custom made, and feature developments in building automation.

With increasing awareness and responsibility toward sustainable living, there is a lot of ground to be broken in the fields of building energy monitoring and BAS. The collaboration between academic and industry experts to delve into issues pertaining to data collection, energy monitoring, and subsequent savings is thus of continuous importance. Mantha et al. [184] presented the taxonomy of data types that directly or indirectly affect building energy consumption and an overview of data collection methods along with their advantages and disadvantages. The challenges associated with data collection from the perspective of decision-makers (i.e. building managers or design/engineers) were discussed. In summary, the extensive review of literature showed that wired networks provided more reliable transmission but building managers and users might find them aesthetically displeasing and cost-prohibitive to install in certain type of facilities (e.g., old historic buildings). In addition, installing wiring is a critical challenge for building managers in existing buildings or buildings that need to be retrofitted. On the other hand, wireless networks are flexible, cost-effective, and easy to install compared to the wired network. They do not need to be connected to the building wiring and hence are immune to electrical damage existing inside buildings. However, because wireless sensor nodes are battery-powered, power consumption issues and subsequent maintenance (which comes at a cost) are major concerns for building managers. Future research should focus on addressing these issues and making the availability and accessibility of data more streamlined and easier to collect and analyze [184].

Yang et al. [185] reported the deployment of a ZigBee-based WSN inside an existing building duct system for intelligent waste collection in an industrial environment. The measured Received Signal Strength (RSS) and path losses reveal that the duct communication channel could act as a very effective waveguide, providing more reliable and consistent network performance than conventional free space channels.

Kaa et al. [186] investigated the important factors that influence the process and outcome of platform battles in the building automation industry. Firstly, they performed a literature review and developed a framework with relevant factors for platform dominance. Then, they analyzed the relative importance of these factors for three types of platforms (subsystem platforms, system platforms, and evolved subsystem platforms) by applying a fuzzy multi-criteria decision-making methodology. Finally, they provided an indication that the influence of factors for BAS platform dominance is modified by the type of platform used.

Kazmi et al. [187] surveyed the state-of-the-art in building energy management systems. A generic architecture was proposed along with a detailed taxonomy of existing documented systems. Gaps in the literature were highlighted while future research directions were identified.

4.3.1.3 Data Level:

Bashir and Gill [188] mentioned that there is a growing interest in IoT-enabled smart buildings. However, the storage and analysis of large amounts of high-speed real-time smart building data are challenging tasks. There are several contemporary Big Data management technologies and advanced analytics techniques that can be used to deal with this challenge. There is a need for an integrated IoT Big Data Analytics (IBDA) framework to fill the research gap in the Big Data Analytics domain. This paper presented one such IBDA framework for the storage and analysis of real-time data generated from IoT sensors deployed inside smart buildings. The initial version of the IBDA framework has been developed using Python and the Big Data Cloudera platform. The applicability of the framework was demonstrated in a scenario involving the analysis of real-time smart building data for automatically managing the oxygen level, luminosity, and smoke/hazardous gases in different parts of the smart building. The initial results indicated that the proposed framework was fit for the purpose and useful for IoT-enabled Big Data Analytics for smart buildings. The key contribution of this paper is the complex integration of Big Data Analytics and IoT for addressing the large volume and velocity challenge of streamlined real-time data in the smart building domain. This framework will be further evaluated and extended through its implementation in other domains.

Koseleva and Ropaite [189] mentioned three main problems with Big Data: (1) taking out the accumulated data in a short time, (2) information overwhelming when involving several dimensions, and (3) limited existing applications to process data in a large volume.

Metadata schema and semantic data for buildings are important topics relevant to data. Mathew et al. mentioned in [190] that when dealing with data from 750,000 buildings, most of the efforts lie in data cleaning and mapping to a common data schema. The common metadata schema in the industry include: Haystack, Brick, Industry Foundation Classes (IFC) and Building Information Models, Building Topology Ontology (BOT), Smart Appliances REFerence Ontology (SAREF), which are summarized in Table 8.

Table 8. Summary table of metadata and semantic schema

No.	**Name**	**Introduction**
1	Brick	Balaji et al. [191] described Brick, a uniform schema for representing metadata in buildings. The schema defines a concrete ontology for sensors, subsystems, and

		relationships among them, which enables portable applications. Their paper demonstrated the completeness and effectiveness of Brick by using it to represent the entire vendor-specific sensor metadata of six diverse buildings across different campuses, comprising 17,700 data points, and running eight complex unmodified applications on these buildings.
2	Project-Haystack	Project Haystack is an open-source initiative to streamline working with data from the IoT. It standardizes semantic data models and web services with the goal of making it easier to unlock value from the vast quantity of data generated by smart devices that permeate our homes, buildings, factories, and cities. Applications include automation, control, energy, HVAC, lighting, and other environmental systems [192]. Charpenay et al. [193] aligned Haystack tagging ontology with the widespread Semantic Sensor Network upper ontology and designed a configuration environment for BAS based on semantic data to illustrate, so as to discuss the added-values of semantics in automation.
3	Building Topology Ontology (BOT)	In the paper written by Rasmussen et al. [194], the Building Topology Ontology (BOT) was suggested in early 2017 to the W3C community group for Linked Building Data as a simple ontology covering the core concepts of a building. It has since been extended to cover a building site, elements hosted by other elements, zones as a super-class of spaces, stories, buildings and sites, interfaces between adjacent zones/elements, a transitive property to infer implicit relationships between building zone siblings among other refinements. The paper also described in detail the changes and the reasons for implementing them.
4	Smart Appliances REFerence Ontology (SAREF)	In the paper [195], Daniele et al. developed SAREF, the Smart Appliance REFerence ontology, to enable semantic interoperability for smart appliances. They presented SAREF and described their experience in creating this ontology in close interaction with the industry, pointing out lessons learned and identifying topics for follow-up actions.
5	Industry Foundation Classes (IFC) and Building Information Models	IFC is a standardized, digital description of the built asset industry. It is an open, international standard (ISO 16739-1:2018) and promotes vendor-neutral, or agnostic, and usable capabilities across a wide range of hardware devices, software platforms, and interfaces for many different use cases [196]. buildingSMART enables the entire built asset industry to improve the sharing of information throughout the lifecycle of project or asset. By breaking down the silos of information, end users can better collaborate and cooperate regardless of which software application they are using. buildingSMART's technical core is based around Industry Foundation Classes (IFC) which was ISO certified in 2013. IFC covers static building data, e.g., construction and design data, which is also very important in building load prediction for design purposes.

4.3.1.4 Integration of System/Sensor/Data Level:

Plageras et al. [175] surveyed IoT, cloud computing, big data, and sensors technologies with the aim of finding their common operations and combining them. Moreover, regarding the smart city concept, the authors tried to propose new methods to collect and manage sensor data in a smart building, which operates in the IoT environment. Finally, this study states that the proposed solutions for collecting and managing sensors' data in a smart building could lead to an energy-efficient smart building.

4.3.2 Data Pre-Processing

The raw data from hardware often have data quality issues such as: (1) missing values, (2) duplicate data, (3) inconsistent data, (4) noise, and (5) outliers. Data from BAS is not an exception. In applying machine learning, three common data pre-processing steps are formatting, cleaning, and sampling. Formatting converts the raw data into a suitable format to work with. Cleaning is the removal or fixing of missing data. Sampling takes a smaller representative portion of data to aid faster exploration and prototype solution investigation before involving the whole dataset.

Like all other machine learning applications, data for building load prediction need pre-processing from the raw BAS data to ready-to-use data for machine learning. Overall, data pre-processing is a common practice for forecasting [26]. Kuster et al. [26] identified four data pre-processing actions in load prediction: (1) smoothing and filling missing values, (2) measurement of variables dependency and significance, (3) data decomposition and classification and (4) check order of integration and stationarity. It appeared that pre-processing has been done in 66.0% of the papers. Note that the remaining 33% of the papers do not necessarily mean a lack of pre-processing but simply that it has not been mentioned [26]. Table 9 summarizes the data pre-processing research in building load prediction.

Table 9. Summary of data processing techniques in machine-learning-based building load prediction

No	Data Pre-processing Techniques	Example Reference
1	Data cleaning, the previous value for missing value imputation	[74]
2	Data resolution processing	[74]
3	Timestamp formatting	[74]
4	Processing noisy and missing data	[64]
5	Scaling and normalization	[197]
6	Outlier detection of building abnormal energy data	[167, 198]
7	Reducing the impact of measurement noises	[199]

Missing data, or missing values, occur when no data value is stored in observation. Missing data is commonly existing with a significant effect on the data analysis results [200]. Missing data are common in BAS measurement due to network connection, data transmission, sensor fault, and so on. Chujai et al. [74] used the building electricity consumption in the previous time step for missing value imputation with the assumption that the current data will be similar to the previous ones. But generally speaking, there is a lack of discussion on how missing data affects machine learning performance and how to optimally address the missing data issue.

Data resolution processing is another common data pre-processing step because data resolution from BAS (ranges from 1 to 15 minutes) is often higher than that is used for load prediction (typically hourly to monthly). Chujai et al. [74] converted the raw data with one-minute resolution to monthly resolution.

Data normalization of input data is a process of transforming data into the normalized form. In the standard normalization process, each input data point was transformed into interval of 0 and 1 (or -1 and 1). The data of each input can be normalized separately or in groups of input variables. The normalization process may enhance the learning process of network, resulting in better performance of ANN-based forecast models [17]

Outlier detection (or anomaly detection) is the identification of rare items, events, or observations that raise suspicions due to the significant deviation from the majority of the data[201]. In building data, outlier detection is often applied to energy meters and sensor data with potential sensor faults. Xiao and Fan [167] applied association rules to detect operation abnormalities to improve building operational performance. Seem [202] described the application of generalized extreme studentized deviate (GESD) many-outlier procedures for detecting abnormal energy consumption in buildings based on daily readings of energy consumption and peak energy consumption . Li et. al [198] proposed an intelligent data-analysis method for modeling and prediction of daily electricity consumption in buildings. The outlier-detection method of combining GESD and the Q-test were proposed to identify abnormally high or low energy use in a building.

Noises are common in energy meters and sensors of buildings. In terms of reducing impacts of measurement noises, Huang et al. [199] presented a scheme adopting the data fusion technique to improve the quality of building cooling load measurement from BAS. A data fusion algorithm was developed to remove outliers and system errors as well as to reduce the impacts of measurement noises.

4.3.3 Feature Extraction and Feature Selection

Feature is the term for input, variable, and predictor in the language of machine learning. Well-engineered features empower flexibility and robustness (i.e., even if a simple algorithm is used, satisfactory prediction results can still be achieved). The flexibility associated with good features allows molders to use fewer complex models that are faster to run, understand, and maintain. Well-engineered features also ensure the approach's robustness, meaning that even if poorly tuned, the model can still get good results. Modelers do not need to work as hard to choose the right models and the optimal parameters. With good features, the modelers are closer to the truth of the underlying problem with a subsection of data that best characterize that underlying problem.

Feature engineering is the process of using domain knowledge to extract features from raw data via data mining techniques. These features can be used to improve the performance of machine learning algorithms. Feature engineering can be considered as applied machine learning itself [203].

On the algorithm side, increasingly complex predictive modeling algorithms perform feature importance and selection internally during model establishment, such as Multivariate Adaptive Regression Splines (MARS), Random Forest and Gradient Boosted Machines. On the application side, especially the machine learning application in load prediction, feature engineering is also widely studied.

Feature extraction and selection are the major topics in feature engineering. Feature extraction is the automatic or manual construction of new features from raw data; feature selection is the automatic identification and use of features in raw data. Section 4.3.3.1 will discuss feature extraction in building load prediction in detail and Section 4.3.3.2 will discuss feature selection.

4.3.3.1 Feature Extraction

There are two categories of feature extraction in building load prediction. The first one is the manual extraction of features from raw data. Manual extraction is dependent on domain knowledge of building energy consumption. The most common manual extractions are time indicators and time-lag variables.

The example format of timestamp from BAS is a string like "2017-02-03 12:30:14". The time indicator extracts the timestamp information to day of the week [40], month of the year, weekday/weekend [37], minute of hour [5], time index of 15min (1-96) [64], and so forth. Time indicators are important because time is strongly correlated to occupancy and schedule of building operation. Time-lag variables are also an important example of manual extraction of features. Researchers use energy consumption in the previous time steps[40], peak energy consumption in the previous time steps, weather variables in the previous time steps [37], time indicators in the previous timesteps [64], and so on as model inputs. Time-lag variables are important since it reflects the consideration of thermal inertia of building energy corresponding to indoor and outdoor dynamics. For some heavy-mass buildings, the time lag can be three or more hours, while for light-mass buildings, the time lag can be one or two hours. Other manual feature extraction by domain knowledge include: Climate Z extracted from dry-bulb temperature and solar radiation [204] and virtual cooling extracted from flow rate and temperature difference [199].

The second category of feature extraction in building load prediction is automated extraction of features. The most widely used method is Principle Component Analysis (PCA). The PCA method, is another method to reduce dimension of a dataset, yet is different from the feature selection method. Both methods seek to reduce the number of attributes for machine learning algorithms, but the PCA method achieves this by creating new and transformed combinations of attributes, whereas feature selection methods identify useful attributes from original data without changing them. For knowledge discovery, interpreting the output of algorithms based on PCA often proves to be problematic, as transformed features usually do not carry physical meaning to domain experts [205]. In contrast, the dimensions retained by a feature selection procedure are generally interpretable. It is worth mentioning that PCA can only be used to visualize data or speed up the machine learning algorithm, but it cannot solve overfitting problems. The papers [106, 206-208] applied PCA, and the paper [209] applied Kernel Principal Component Analysis (KPCA) to the building load prediction.

4.3.3.2 Feature Selection

After collecting data from hundreds of sensors in BAS, features or input variables for building energy modeling should be determined. If too many features are used, the complexity of data-driven models will dramatically increase, leading to high computing cost, and low generalization (bias-variance tradeoff). Thus, selection of features has become the focus of many research efforts where original datasets often contain tens or hundreds of variables. For a data-driven model, the feature selection strongly affects the model performance. Hence, the science of feature selection, also known as variable selection, attribute selection or variable subset selection, is essential to effectively identify features for a data-driven building load prediction model. The objective of a feature selection process is three-fold: (1) improving the prediction performance of the model, (2) providing faster and more cost-effective prediction model, and (3) providing a better understanding of the underlying process that generates the data [210].

There are two difficulties when applying feature selection in building load prediction model. First, among hundreds of candidate features from BAS, it is difficult to decide the optimum model complexity, along with number of features, that can fully describe the variation of the output with the least model complexity. Overwhelming or insufficient complexities of model will lead to overfitting and underfitting, respectively, both of which can cause poor model generalization. Second, features collected from BAS vary from building to building. Moreover, the impact of the same feature on building energy varies from building to building. The decision of feature selection requires a good understanding of both building physics and the specific building itself. This process could require much involvement of expert knowledge and customized modeling, which will increase development engineering cost.

Existing studies on building energy modeling often select model features (or input variables) based on underlining physics and domain knowledge, such as heat transfer, thermodynamics, meteorology, building physics, and experiences. Leung et. al roughly divided the inputs or features that affect a building's energy consumption into three groups: external factors, internal factors, and operation of HVAC systems [211]. Karati's book about energy modeling process recommended the following input variables: external climatic data, hour-type, day-type, pretreated air unit operation schedule, and the occupancy space electrical power demand [212]. Li et al. used six input variables for building cooling load prediction: the outdoor dry-bulb temperature and relative humidity at three time steps [213]. Yang et.al summarized typical input variables for on-line building energy prediction including environmental data, time, and operating parameters. It also considered time-lag and tried different combinations of time-lagged temperature measurements [214]. Dong et al. used three weather features including monthly mean outdoor dry-bulb temperature, relative humidity, and global solar radiation as three input features when applying support vector machines to predict building energy consumption in the tropical region [104].

The paper [210] reviewed four types of feature selection techniques applied in building load prediction: the filter, wrapper, hybrid, and embedded method. The filter method, a category of statistical methods, assigns a score to each feature; then ranks the features by the score, and eventually either keeps or removes the feature from the candidate feature set based on the ranking [210]. The wrapper method compares the prediction performance of data-driven models using subsets of candidate features to assess the usefulness of each subset of candidate features [210]. The embedded method is a feature selection method embedded in machine learning algorithms that include feature selection as an intrinsic module such as random forest and multivariate adaptive regression splines (MARS). The hybrid method is the application of at least two of the above techniques as a combined method to select features. Table 10 lists the feature selection techniques from existing building load prediction papers.

Table 11 lists features that are often used or selected for building load prediction. Table 11 categorizes features into time indicators, building operation status, weather variables, economic indicators, time lags, occupancy, and building parameters. The categorization of input data is more detailed than the review paper written by Wang and Srinivasan [23] where there are only three input categories: meteorology, occupancy, and others. According to the papers summarized in the Table 11, if machine learning model is used for energy prediction of multiple building or building design evaluation, building parameters are often used as model inputs.

Table 10. Summary of feature selection techniques in machine-learning-based building load prediction

No.	Feature Selection Type	Feature Selection Algorithm	Example Reference
1	Filter method	Wald's test	[215]
2		Gradient guided feature selection	[216]
3		Correlation coefficients	
4		Permutation importance	[43]
5		Node impurity	[43]
6		p-value	[217]
7		Mutual information	[56]
8		Sliding window EMD (SWEMD)	[27]
9		Minimum redundancy maximum relevance	[56]
10		ANOVA test	[31]
11		Cross-correlation analysis	[33]
12		ℓ2-norm (Euclidean distance)	[156]
13		ReliefF	[218]
14	Wrapper method	Recursive feature elimination	[37]
15		Forward feature selection	[34, 219]
16		Forward genetic search	[47]
17		Stepwise feature selection	[218]
18		Genetic algorithms and information theory	[218]
19		Fuzzy-rough feature selection	[220]
20	Embedded	Boosting tree algorithm	[44]
21		Information gain in decision tree	[36]
22		Random forest algorithm	[32, 43]
23		C5.0 (decision tree)	[64]
24		Fuzzy Inductive Reasoning (FIR)	[221]
25		C4.5 (decision tree)	[36]
26		Linear regression	[216]
27	Hybrid	Pearson correlation coefficient and recursive feature elimination	[222]
28		Minimum-redundancy-maximum-relevance-based Pearson's correlation coefficient (MRMRPC)	[55]
29	Others	Least Absolute Shrinkage and Selection Operator (LASSO)	[40]
30		Chaotic feature selection	[223]

Table 11. Common features for machine-learning-based building load prediction

Category	Features	Example Reference
Time indicator	Day of the week	[40]
	Month of the year	[37]
	Weekday/weekend indicator	[37]
	Minute of hour	[5]
	Time index of 15min (1-96)	[64]
Building operation status	Supply and return chilled water temperature	[5]
	Flow rate of the chilled water	[5]
	Constant air volume systems operation schedule	[43]
	Infiltration rate	[43]
	Heating/cooling temperature setpoint	
	Heating terminals: fan coils (FC)/hot water radiators	
	Boiler energy efficiency	
	Chiller energy efficiency ratio	
	Inlet air temperature	[48]
	Supplied air flow into room	

Category	Features	Example Reference
	Indoor relative humidity	
	Indoor room temperature	
	Water temperature	
Weather variables	Outdoor air dry-bulb temperature	[40]
	Outdoor relative humidity	[38]
	Solar radiation	[38]
	Wind speed	[34]
	Outdoor dew point temperature	[32]
	Precipitation probability	[43]
	Rain indicator	
	Sky condition	
	Direct normal radiation	[31]
	Diffuse horizontal radiation	
Economic indicator	Composite leading index, composite coincident index, composite lagging index, monitoring indicator, average rate of exchange to U.S. dollar, economic growth rate, gross domestic product, gross national product, national income, unemployment rate, consumer price index, oil prices, prices of liquid petroleum gas, and average of the Taiwan stock exchange.	[64]
Time lags	Previous energy consumption	[40]
	Previous peak energy consumption	[37]
	Previous weather variables	[37]
	Previous time indicators	[64]
Occupancy	Number of occupancies	[36]
	Number of rooms booked or number of guests (for hotels)	[32]
	People density	[59]
	Wi-Fi connections	[46]
Building parameters	House type	[36]
	Construction type	
	Floor area	
	Heat loss coefficient	
	Equivalent leakage area	
	Shapes of building	[61]
	Relative compactness	
	Glazing area	
	Roof area	
	Surface area	
	Wall area	
	Orientation	
	Overall height	
	Glazing area distribution	
	Form factor	[29]
	Net area in use	[58]
	Year of construction	
	Area of each floor	[59]
	Floor height	
	Window to wall ratio	
	Number of floors	
	Envelop thickness and thermal conductivity	
	Coefficient of performance (COP)	[63]
	Energy efficiency ratio (EER)	
	Heating emission factors (HEF)	
	Cooling emission factors (CEF)	

4.3.4 Knowledge Gaps in Data-Related Topics

In this section, data-related topics are reviewed, including data generation, data collection, data pre-processing, feature extraction and selection techniques. There are still several key research topics missing in this field.

First, in terms of data generation and collection (Section 4.3.1), the sensor-level topic is the least studied sub-topic. Besides the occupancy sensor and energy-environment sensors, there are more advanced techniques and research topics, including virtual sensors, sensor auto-calibration, sensor layout/location, and sensor data analysis/mining, can be further explored; the system-level topics are relatively well studied by covering advanced BAS, IoT, AMI, and sensor network. However, most papers provide conceptual analysis and there lacks engineering implementation of the system-level techniques to provide engineering references on these techniques in real building load forecasting; the data-level topics cover the topics of big data, metadata, and semantic schema. The missing part of this subsection is that there are no study comparing among metadata and semantic schema: a clear guidance of selecting among the existing metadata and semantic schema is missing in the current literature.

In terms of data pre-processing (Section 4.3.2), most topics including (1) missing values, (2) duplicate data, (3) inconsistent data, (4) noise, and (5) outliers are covered. For future development of this subsection, the key knowledge gap is a lack of automated and streamlined data-preprocessing tool that can be embedded to BAS and IoT.

Since machine learning is data hungry and has high requirement for data quality, the techniques to enhance data quality by actively generating extra data with better quality and representative to improve machine learning algorithm are important. There is a great need to develop better methods to effectively collect unbiased building operation data to generate high-fidelity and cost-effective data-driven building energy models. Active learning, similar to optimal experiment design in statistics and excitation in system identification, is essential to effectively generating unbiased and informative training data. However, no studies are found to discuss the topic of active learning in building load prediction.

5. Conclusions and Future Directions

This paper reviewed both review and technical papers on the topic of building load prediction with machine learning techniques. The Sub-keyword Synonym Searching (SSS) methodology (Section 2) was proposed and applied for paper searching in this study: it uses sub-keywords and synonyms to conduct multiple searches to comprehensively and exhaustively capture relevant papers. All papers were reviewed under the topics of applications, machine learning algorithms, and data-related topics.

In Section 3, we reviewed the review papers and concluded that the algorithm development of machine learning is the most attractive topic in building load prediction because most review papers are organized in terms of the machine learning algorithms; the data side of machine learning and supporting techniques, like feature selection, extraction, clustering, and weather forecasting, are also drawing increasing attentions.

In Section 4, we reviewed the technical papers in terms of tasks T (Section 4.1, application), performance measure P (Section 4.2, algorithm), and experience **E** (Section 4.3, data), from the definition of machine learning by Tom M. Mitchell. First in Task **T** or applications of load prediction model, in terms of energy type of load to predict, can be categorized to (1) cooling load/energy prediction, (2) heating load/energy prediction, (3) HVAC load/energy prediction, (4) district heating load/energy prediction, (5) primary energy load/consumption prediction, (6) natural gas consumption prediction, (7) electricity load/consumption prediction, and (8) steam load prediction. In terms of application scenarios of load prediction model, there are three major categories: (1) model predictive control/demand response/control optimization, (2) building parameter design/retrofit, and (3) building energy planning/climate change impact. The forecasting horizon has three major categories: (1) hour(s), (2) day(s), and (3) month(s) or longer. Machine learning techniques are widely applied in various application fields of building load prediction.

In terms of Performance **P**, or the algorithm-side topics (Section 4.2), Table 6 summarizes the 128 different machine learning algorithms that were observed to be applied in building load prediction, covering algorithms of Regression, Support Vector Machine (SVM), Neural Network (NN), Deep Learning, Tree-based Algorithms, Hybrid Algorithms, Autoregressive methods, Extreme Learning, Bayesian Networks, Case-based Reasoning (CBR), Meta Learning, k-Nearest Neighbors (kNN), Gaussian Process and Mixture Models, Ensemble Methods, Fuzzy Timeseries Algorithm. Not only the reviewed algorithms, but also the development and trend of machine learning algorithms in building load prediction are summarized:

(1) Traditional algorithms, such as ANN, SVM, etc., are combined with clustering algorithms, sophisticated parameter optimization, rule-based modeling, feature extraction, and other data-driven algorithms to increase prediction accuracy.

(2) With the development of machine learning and information science, more novel algorithms and modeling approaches are applied for building load prediction. According to authors' literature review, Table 6 shows that many machine learning algorithms have already been applied in building load prediction and analysis.

(3) Improving comprehensibility and interpretability of the modeling process is another important topic in the building load prediction modeling.

(4) Many approaches are brought up to predict dynamics and non-linearity of building energy consumption.

Finally, in Experience **E** or the data side topics (Section 4.3), the paper reviewed this topic ranging from data generation from hardware to final inputs of machine learning models, specifically including (a) data generation and collection, (b) data pre-processing, and (c) feature extraction and selection. The conclusions are summarized:

(1) On the data generation and collection side, there is a trend of moving from traditional BAS system with a local area network, to wireless sensor network (WSN), and finally to IoT. Also, metadata schema and semantic data for buildings are increasing applied.

(2) On the data pre-processing side, different aspects including missing data imputation, data resolution processing, outlier detection, and noises processing are reviewed. A better description of data pre-processing in most of the current technical papers, as well as papers to comprehensively study the topic of data pre-processing, are still lacking.

(3) On the feature engineering side, feature extraction is categorized into manual feature extraction with domain knowledge and automated feature extraction with PCA techniques; in terms of feature selection, Table 11 lists features that are often used for load prediction and Table 10 summarized the feature selection techniques from existing building load prediction papers. There is increasing interest in the feature engineering process in the current studies.

5.1 Primary Limitations and Gaps

Most of reviews and technical papers focus on the machine learning algorithm development. Compared with the algorithm side, the data side is still not well investigated: about 110 papers (in Section 4.2) are focusing on the algorithm side but only around 50 papers (in Section 4.3) are mentioning the data side.

Among the reviewed technical papers, one issue is that the information on reported machine learning model development is often sporadic with important information missing. For example, many papers simply mentioned what data they use for modeling but without giving more details such as data resolution, training/testing data split, etc. Hence, in this paper, a somewhat standardized information table, Table 12, is recommended to be followed for future publication, in order to provide a completed picture on the reported machine learning and modeling techniques, and to make it easier for readers to collect information from the paper.

Table 12. Recommended summary table for future paper writing of machine-learning-based building load prediction studies

Category	**Key Information**	**Description**
General information	Building type	Commercial, residential, commercial/residential
	Building numbers	Single building, multiple buildings
	Energy type to predict	Cooling load, heating load, natural gas load, electricity load/consumption, steam load, and so on
	Application scenario	Model predictive control/demand response/control optimization, building parameter design/retrofit, building energy planning/climate change impact
	Forecasting horizon	Short/medium/long, minutes/hours/days/weeks/months/years
Data description	Source of data	Simulation, real building BAS, AMI data
	Sampling interval of data	Seconds, minutes, hours
	Data cleaning method	Missing data imputation, etc.
	Training/validation/testing data description	e.g. training data from July 1st to August 31st, 60% data for training, 40% for validation and testing, etc.
Feature engineering	Feature extraction	Features extracted from raw sensor/measurements
	Feature selection	Feature selection techniques to select from measurements and extracted features
	Dimension reduction	e.g. PCA
	Final features used for algorithm	A list of feature names after feature extraction and feature selection
Algorithm	Main algorithm structure	e.g. random forest, ANN, SVM, etc.
	Tuning method	e.g. GA, SOM, etc.
	Other support techniques	Clustering techniques, etc.
Performance evaluation	Error metrics	e.g. mean absolute percentage error (MAPE), root mean squared error (RMSE)

One of the key limitations on the application of machine learning algorithm side is a lack of high-quality and real-world-collected datasets and testbeds to evaluate the performance of algorithms. Although there are some datasets such as "ASHRAE - Great Energy Predictor III, the Kaggle competition" [171] which provides the dataset for training and testing, diversified and high-quality testbeds are still lacking to evaluate algorithm performance. Besides, tested under standardized experimental environment (for example, Flexible Research Platform at Oak Ridge National Lab), the conclusion of the algorithm performance will be more convincing.

5.2 Future Trends and Challenges

We believe that the trend of future studies on applying machine learning techniques for building load prediction is to advance from a simple application of certain specific machine learning algorithms to the development of an integrated process and/or framework including data engineering, feature engineering, and machine learning algorithm training and application. More specifically, the following trends are predicted:

- The development emphasis will be transferred from the algorithm side to the data side.
- Novel methods to integrate domain knowledge in a machine learning process will be increasingly explored. The machine learning development is not only the development of algorithms itself but also the exploration of algorithm design for different application scenarios.
- More advanced machine learning supporting techniques such as clustering techniques, data augmentation, active learning techniques, and so forth will be applied in the model development process.

- The future machine learning algorithms and models will be developed in plug-n-play format without much manual tuning.

From our point of view, the major challenges of the continuous study and development of machine learning-based load prediction are the real building applications of load prediction models, the realization of automation, and reduction of engineering cost. To be more specific:

- The realization of automation and the reduction of the engineering cost are the key advantages of machine learning based load prediction, but it is very challenging to embed the load prediction algorithms and models into the existing BAS and IoT systems in the industry.

- To realize the automation, improving the extendibility of algorithms to be applicable to different building types, energy system types, and weather is the starting point. An automated framework with automated tuning based on specific building and data is challenging.

- It is also challenging to apply varieties of real building testbeds to test the extendibility of algorithms because real building testbeds with high data quality are not common.

5.3 Future Research Directions

Although there are a great number of studies and wide research topics covered in machine learning-based load prediction as reviewed in this paper, many other important topics are not or hardly covered: there are many research opportunities for future studies. Similar to the previous structure, the future opportunities and research directions are grouped by applications (task T), algorithms (performance P), and data (experience E) and are listed as follows:

- Applications (task T)
    - With the rapid development of smart city and connected communities, an increasing demand for the load prediction for multiple buildings or building clusters is observed. Moreover, the building load prediction under demand response and building-grid interaction are becoming more complicated and challenging. The new research topics include the interdependency of energy load of building clusters, features used for building clusters load prediction, and load prediction framework under building-grid-integration.
    - Equipment faults and control errors are pervasive in today's commercial buildings [224] and building faults waste a substantial amount of energy each year—0.3 to 1.9 exajoule (EJ, or 10^{18} J) of primary energy in the United States [225]. Load prediction that automatically take the impact of building faults on building energy into consideration is very challenging. The uncertainty quantification of faults on building energy can be a potential research opportunity.

- Algorithms (performance P)
    - The fast-growing algorithm development in the field of machine learning and artificial intelligence can continue to provide new algorithms to be used for building load prediction.
    - With the development of increasingly sophisticated algorithms, there will be a bottleneck of model performance when algorithms have higher complexity than data or the algorithm is unnecessarily complicated for the existing data. To solve this problem, algorithms that can automatically generate new data and include the extra data into the training process is the future opportunities in research.
    - Robust autotuning algorithms that is plug-n-play to different buildings and data are essential to the wide industrial applications of building load prediction.

- The integration of machine learning techniques and other techniques that have been developed for dynamic system modeling, such as system identification, should be better explored.

- Data (experience E)
  - For better development of algorithm side in this field, testing data with high quality are in great need to provide a relatively standard criteria to evaluate algorithm performance.
  - The integration of smart sensors, especially occupancy sensors, with machine learning algorithm has great opportunities to provide extra information and features to building load prediction.
  - There is a lack of engineering implementation of the system-level sensor techniques (Section 4.3.1) to provide engineering references on these techniques in real building load prediction.
  - These are few studies comparing among metadata and semantic schema: a clear guidance of selecting among the existing metadata and semantic schema is missing in the current literature, which can be a good opportunity in the future research.
  - There lacks an automated and streamlined data-preprocessing tool that can be embedded to BAS and IoT.
  - Since machine learning is data hungry and has high requirement for data quality, the techniques to enhance data quality by actively generating extra data with better quality and representative to improve machine learning algorithm are important. There is a great need to develop better methods to effectively collect unbiased building operation data to generate high-fidelity and cost-effective data-driven building energy models. Active learning in machine learning is essential to effectively generating unbiased and informative training data. However, no studies are found to discuss the topic of active learning in building load prediction, making it a good research direction in the future.

Reference

1. EIA, Monthly energy review – June 2020. US Energy Information Administration, 2020.
2. Li, X. and J. Wen, Review of building energy modeling for control and operation. Renewable and Sustainable Energy Reviews, 2014. 37: p. 517-537.
3. Jordan, M.I. and T.M. Mitchell, Machine learning: Trends, perspectives, and prospects. Science, 2015. 349(6245): p. 255-260.
4. Mitchell, T.M., Machine learning. 1997. Burr Ridge, IL: McGraw Hill, 1997. 45(37): p. 870-877.
5. Fan, C., F. Xiao, and Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms. Applied energy, 2017. 195: p. 222-233.
6. Peng, Y., et al., Using machine learning techniques for occupancy-prediction-based cooling control in office buildings. Applied energy, 2018. 211: p. 1343-1358.
7. Kim, W. and S. Katipamula, A review of fault detection and diagnostics methods for building systems. Science and Technology for the Built Environment, 2018. 24(1): p. 3-21.
8. Nosratabadi, S., et al. State of the art survey of deep learning and machine learning models for smart cities and urban sustainability. in International Conference on Global Research and Education. 2019. Springer.
9. Potter, C.W., A. Archambault, and K. Westrick. Building a smarter smart grid through better renewable energy information. in 2009 IEEE/PES Power Systems Conference and Exposition. 2009. IEEE.
10. Li, X. and J. Wen, Building energy consumption on-line forecasting using physics based system identification. Energy and Buildings, 2014. 82: p. 1-12.
11. Mocanu, E., et al., Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks, 2016. 6: p. 91-99.

12. Li, X., et al., Commercial building cooling energy forecasting using proactive system identification: A whole building experiment study. Science and Technology for the Built Environment, 2016. 22(6): p. 674-691.
13. Cugliari, J., Y. Goude, and J.-M. Poggi. Disaggregated electricity forecasting using wavelet-based clustering of individual consumers. in 2016 IEEE International Energy Conference (ENERGYCON). 2016. IEEE.
14. Zhao, H.-x. and F. Magoulès, A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 2012. 16(6): p. 3586-3592.
15. Ahmad, A., et al., A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews, 2014. 33: p. 102-109.
16. Fumo, N., A review on the basics of building energy estimation. Renewable and Sustainable Energy Reviews, 2014. 31: p. 53-60.
17. Raza, M.Q. and A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews, 2015. 50: p. 1352-1372.
18. Yildiz, B., J.I. Bilbao, and A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable and Sustainable Energy Reviews, 2017. 73: p. 1104-1122.
19. Daut, M.A.M., et al., Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renewable and Sustainable Energy Reviews, 2017. 70: p. 1108-1118.
20. Deb, C., et al., A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews, 2017. 74: p. 902-924.
21. Seyedzadeh, S., et al., Machine learning for estimation of building energy consumption and performance: a review. Visualization in Engineering, 2018. 6(1): p. 5.
22. Lazos, D., A.B. Sproul, and M. Kay, Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. Renewable and Sustainable Energy Reviews, 2014. 39: p. 587-603.
23. Wang, Z. and R.S. Srinivasan. A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. in 2015 Winter Simulation Conference (WSC). 2015. IEEE.
24. Molina-Solana, M., et al., Data science for building energy management: A review. Renewable and Sustainable Energy Reviews, 2017. 70: p. 598-609.
25. Amasyali, K. and N.M. El-Gohary, A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 2018. 81: p. 1192-1205.
26. Kuster, C., Y. Rezgui, and M. Mourshed, Electrical load forecasting models: A critical systematic review. Sustainable Cities and Society, 2017. 35: p. 257-270.
27. Liu, Y., W. Wang, and N. Ghadimi, Electricity load forecasting by an improved forecast engine for building level consumers. Energy, 2017. 139: p. 18-30.
28. Gerwig, C. Short term load forecasting for residential buildings—An extensive literature review. in International Conference on Intelligent Decision Technologies. 2017. Springer.
29. Ekici, B.B. and U.T. Aksoy, Prediction of building energy needs in early stage of design by using ANFIS. Expert Systems with Applications, 2011. 38(5): p. 5352-5358.
30. Kwok, S.S. and E.W. Lee, A study of the importance of occupancy to building cooling load in prediction by intelligent approach. Energy Conversion and Management, 2011. 52(7): p. 2555-2564.
31. Sholahudin, S. and H. Han, Simplified dynamic neural network model to predict heating load of a building using Taguchi method. Energy, 2016. 115: p. 1672-1678.
32. Ahmad, M.W., M. Mourshed, and Y. Rezgui, Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 2017. 147: p. 77-89.
33. Sala-Cardoso, E., et al., Activity-aware HVAC power demand forecasting. Energy and Buildings, 2018. 170: p. 15-24.
34. Jovanović, R.Ž., A.A. Sretenović, and B.D. Živković, Ensemble of various neural networks for prediction of heating energy consumption. Energy and Buildings, 2015. 94: p. 189-199.
35. Powell, K.M., et al., Heating, cooling, and electrical load forecasting for a large-scale district energy system. Energy, 2014. 74: p. 877-885.
36. Yu, Z., et al., A decision tree method for building energy demand modeling. Energy and Buildings, 2010. 42(10): p. 1637-1646.

37. Fan, C., F. Xiao, and S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 2014. 127: p. 1-10.
38. Fumo, N. and M.R. Biswas, Regression analysis for prediction of residential energy consumption. Renewable and Sustainable Energy Reviews, 2015. 47: p. 332-343.
39. Braun, M., H. Altan, and S. Beck, Using regression analysis to predict the future energy consumption of a supermarket in the UK. Applied Energy, 2014. 130: p. 305-313.
40. Jain, R.K., et al., 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, 2014. 123: p. 168-178.
41. Edwards, R.E., J. New, and L.E. Parker, Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, 2012. 49: p. 591-603.
42. Marino, D.L., K. Amarasinghe, and M. Manic. Building energy load forecasting using deep neural networks. in IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. 2016. IEEE.
43. Chae, Y.T., et al., Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings, 2016. 111: p. 184-194.
44. Kusiak, A., M. Li, and Z. Zhang, A data-driven approach for steam load prediction in buildings. Applied Energy, 2010. 87(3): p. 925-933.
45. Ahmad, T. and H. Chen, Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. Energy and Buildings, 2018. 166: p. 460-476.
46. Pombeiro, H., et al., Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. Energy and Buildings, 2017. 146: p. 141-151.
47. Massana, J., et al., Short-term load forecasting in a non-residential building contrasting models and attributes. Energy and Buildings, 2015. 92: p. 322-330.
48. Yuce, B., et al., Utilizing artificial neural network to predict energy consumption and thermal comfort level: An indoor swimming pool case study. Energy and Buildings, 2014. 80: p. 45-56.
49. Zhang, F., et al., Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. Energy and Buildings, 2016. 126: p. 94-103.
50. Ruiz, L.G.B., et al., An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. Energies, 2016. 9(9): p. 684.
51. Jetcheva, J.G., M. Majidpour, and W.-P. Chen, Neural network model ensembles for building-level electricity load forecasts. Energy and Buildings, 2014. 84: p. 214-223.
52. Fernández, I., C.E. Borges, and Y.K. Penya. Efficient building load forecasting. in ETFA2011. 2011. IEEE.
53. Fu, Y., et al., Using support vector machine to predict next day electricity load of public buildings with sub-metering devices. Procedia Engineering, 2015. 121: p. 1016-1022.
54. Li, K. and H. Su, Forecasting building energy consumption with hybrid genetic algorithm-hierarchical adaptive network-based fuzzy inference system. Energy and Buildings, 2010. 42(11): p. 2070-2076.
55. Mohammadi, M., et al., Small-scale building load forecast based on hybrid forecast engine. Neural Processing Letters, 2018. 48(1): p. 329-351.
56. Chitsaz, H., et al., Short-term electricity load forecasting of buildings in microgrids. Energy and Buildings, 2015. 99: p. 50-60.
57. Escrivá-Escrivá, G., et al., New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. Energy and Buildings, 2011. 43(11): p. 3112-3119.
58. Aman, S., Y. Simmhan, and V.K. Prasanna. Improving energy use forecast for campus micro-grids using indirect indicators. in 2011 IEEE 11th International Conference on Data Mining Workshops. 2011. IEEE.
59. Ascione, F., et al., Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. Energy, 2017. 118: p. 999-1017.
60. Mottahedi, M., et al., Multi-linear regression models to predict the annual energy consumption of an office building with different shapes. Procedia engineering, 2015. 118: p. 622-629.
61. Kumar, S., S.K. Pal, and R.P. Singh, A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. Energy and Buildings, 2018. 176: p. 275-286.
62. Rahman, A., V. Srikumar, and A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy, 2018. 212: p. 372-385.
63. Pino-Mejías, R., et al., Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions. Energy, 2017. 118: p. 24-36.

64. Hsiao, Y.-H., Household electricity demand forecast based on context information and user daily schedule analysis from meter data. IEEE Transactions on Industrial Informatics, 2014. 11(1): p. 33-43.
65. Lusis, P., et al., Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Applied Energy, 2017. 205: p. 654-669.
66. Tahmassebi, A. and A.H. Gandomi, Building energy consumption forecast using multi-objective genetic programming. Measurement, 2018. 118: p. 164-171.
67. Chou, J.-S. and D.-K. Bui, Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy and Buildings, 2014. 82: p. 437-446.
68. Deng, L. and D. Yu, Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 2014. 7(3–4): p. 197-387.
69. Mocanu, E., et al., Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning. Energy and Buildings, 2016. 116: p. 646-655.
70. Kazienko, P., E. Lughofer, and B. Trawiński, Hybrid and ensemble methods in machine learning J. UCS special issue. Journal of Universal Computer Science, 2013. 19(4): p. 457-461.
71. Opitz, D. and R. Maclin, Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 1999. 11: p. 169-198.
72. Penya, Y.K., C.E. Borges, and I. Fernández. Short-term load forecasting in non-residential buildings. in IEEE Africon'11. 2011. IEEE.
73. Yun, K., et al., Building hourly thermal load prediction using an indexed ARX model. Energy and Buildings, 2012. 54: p. 225-233.
74. Chujai, P., N. Kerdprasop, and K. Kerdprasop. Time series analysis of household electric consumption with ARIMA and ARMA models. in Proceedings of the International MultiConference of Engineers and Computer Scientists. 2013.
75. Newsham, G.R. and B.J. Birt. Building-level occupancy data to improve ARIMA-based electricity use forecasts. in Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building. 2010.
76. Mohamed, N., M.H. Ahmad, and Z. Ismail. Short term load forecasting using double seasonal ARIMA model. in Proceedings of the regional conference on statistical sciences. 2010.
77. Yao, Y., et al., Hourly cooling load prediction by a combined forecasting model based on analytic hierarchy process. International journal of thermal sciences, 2004. 43(11): p. 1107-1118.
78. Amjady, N., Short-term hourly load forecasting using time-series modeling with peak load estimation capability. IEEE Transactions on power systems, 2001. 16(3): p. 498-505.
79. Espinoza, M., et al., Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Transactions on Power Systems, 2005. 20(3): p. 1622-1630.
80. Pereira, C.M., N.N. de Almeida, and M.L. Velloso, Fuzzy modeling to forecast an electric load time series. Procedia Computer Science, 2015. 55: p. 395-404.
81. Huang, C.-M., C.-J. Huang, and M.-L. Wang, A particle swarm optimization to identifying the ARMAX model for short-term load forecasting. IEEE Transactions on Power Systems, 2005. 20(2): p. 1126-1133.
82. Yang, H.-T. and C.-M. Huang, A new short-term load forecasting approach using self-organizing fuzzy ARMAX models. IEEE Transactions on Power Systems, 1998. 13(1): p. 217-225.
83. Ju, F.-Y. and W.-C. Hong, Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting. Applied Mathematical Modelling, 2013. 37(23): p. 9643-9651.
84. Chou, J.-S. and N.-T. Ngo, Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. Applied Energy, 2016. 177: p. 751-770.
85. O'Neill, Z. and C. O'Neill, Development of a probabilistic graphical model for predicting building energy performance. Applied energy, 2016. 164: p. 650-658.
86. Monfet, D., et al., Development of an energy prediction tool for commercial buildings using case-based reasoning. Energy and Buildings, 2014. 81: p. 152-160.
87. Cui, C., et al., Short-term building energy model recommendation system: A meta-learning approach. Applied energy, 2016. 172: p. 251-263.
88. Al-Qahtani, F.H. and S.F. Crone. Multivariate k-nearest neighbour regression for time series data—A novel algorithm for forecasting UK electricity demand. in The 2013 international joint conference on neural networks (IJCNN). 2013. IEEE.

89. Colombo, T., I. Koprinska, and M. Panella. Maximum Length Weighted Nearest Neighbor approach for electricity load forecasting. in 2015 International Joint Conference on Neural Networks (IJCNN). 2015. IEEE.
90. Sudheer, G. and A. Suseelatha, Short term load forecasting using wavelet transform combined with Holt–Winters and weighted nearest neighbor models. International Journal of Electrical Power and Energy Systems, 2015. 64: p. 340-346.
91. Zhang, Y., et al. An inverse model with uncertainty quantification to estimate the energy performance of an office building. in Proceedings of 13th international building performance simulation association conference. 2013.
92. Srivastav, A., A. Tewari, and B. Dong, Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. Energy and Buildings, 2013. 65: p. 438-447.
93. Boltürk, E., B. Öztayşi, and I. Sarı. Electricity consumption forecasting using fuzzy time series. in 2012 IEEE 13th International Symposium on Computational Intelligence and Informatics (CINTI). 2012. IEEE.
94. Efendi, R., Z. Ismail, and M.M. Deris, A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand. Applied Soft Computing, 2015. 28: p. 422-430.
95. Liu, X., E. Bai, and L. Luo. Time-variant slide fuzzy time-series method for short-term load forecasting. in 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems. 2010. IEEE.
96. Wi, Y.-M., S.-K. Joo, and K.-B. Song, Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment. IEEE Transactions on Power Systems, 2011. 27(2): p. 596-603.
97. Song, K.-B., et al., Short-term load forecasting for the holidays using fuzzy linear regression method. IEEE Transactions on Power Systems, 2005. 20(1): p. 96-101.
98. Enayatifar, R., et al., Imperialist competitive algorithm combined with refined high-order weighted fuzzy time series (RHWFTS–ICA) for short term load forecasting. Energy Conversion and Management, 2013. 76: p. 1104-1116.
99. Ma, Y., et al. Study on power energy consumption model for large-scale public building. in 2010 2nd International Workshop on Intelligent Systems and Applications. 2010. IEEE.
100. Aydinalp-Koksal, M. and V.I. Ugursal, Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. Applied Energy, 2008. 85(4): p. 271-296.
101. Hygh, J.S., et al., Multivariate regression as an energy assessment tool in early building design. Building and Environment, 2012. 57: p. 165-175.
102. Azadeh, A., M. Saberi, and O. Seraj, An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran. Energy, 2010. 35(6): p. 2351-2366.
103. Cheng, M.-Y. and M.-T. Cao, Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. Applied Soft Computing, 2014. 22: p. 178-188.
104. Dong, B., C. Cao, and S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings, 2005. 37(5): p. 545-553.
105. Zhao, H.X. and F. Magoulès, Parallel support vector machines applied to the prediction of multiple buildings energy consumption. Journal of Algorithms and Computational Technology, 2010. 4(2): p. 231-249.
106. Jinhu, L., et al. Applying principal component analysis and weighted support vector machine in building cooling load forecasting. in 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering. 2010. IEEE.
107. Xuemei, L., et al. Building cooling load forecasting using fuzzy support vector machine and fuzzy c-mean clustering. in 2010 international conference on computer and communication technologies in agriculture engineering. 2010. IEEE.
108. Wang, J., et al., An annual load forecasting model based on support vector regression with differential evolution algorithm. Applied Energy, 2012. 94: p. 65-70.
109. Fan, S. and L. Chen, Short-term load forecasting based on an adaptive hybrid method. IEEE Transactions on Power Systems, 2006. 21(1): p. 392-401.
110. Sun, C. and D. Gong. Support vector machines with PSO algorithm for short-term load forecasting. in 2006 IEEE International Conference on Networking, Sensing and Control. 2006. IEEE.
111. Wang, J., Y. Zhou, and X. Chen. Electricity load forecasting based on support vector machines and simulated annealing particle swarm optimization algorithm. in 2007 IEEE International Conference on Automation and Logistics. 2007. IEEE.

112. Niu, D., Y. Wang, and D.D. Wu, Power load forecasting using support vector machine and ant colony optimization. Expert Systems with Applications, 2010. 37(3): p. 2531-2539.
113. Che, J., A novel hybrid model for bi-objective short-term electric load forecasting. International Journal of Electrical Power and Energy Systems, 2014. 61: p. 259-266.
114. Chen, Y., et al., A hybrid application algorithm based on the support vector machine and artificial intelligence: An example of electric load forecasting. Applied Mathematical Modelling, 2015. 39(9): p. 2617-2632.
115. Hong, W.-C., Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. Energy, 2011. 36(9): p. 5568-5578.
116. Wang, Y., D. Niu, and X. Ma, Optimizing of SVM with hybrid PSO and genetic algorithm in power load forecasting. Journal of Networks, 2010. 5(10): p. 1192.
117. Pai, P.-F. and W.-C. Hong, Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. Electric Power Systems Research, 2005. 74(3): p. 417-425.
118. Hong, W.-C., Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. International Journal of Electrical Power and Energy Systems, 2009. 31(7-8): p. 409-417.
119. Kavousi-Fard, A., H. Samet, and F. Marzbani, A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. Expert Systems with Applications, 2014. 41(13): p. 6047-6056.
120. Baziar, A. and A. Kavousi-Fard, Short term load forecasting using a hybrid model based on support vector regression. Int. J. Sci. Technol. Res, 2015. 4(05): p. 189-195.
121. Hong, W.-C., Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model. Energy Conversion and Management, 2009. 50(1): p. 105-117.
122. Ruhaidah, S., S. Puteh, and S. Ani, Hybridizing GMDH and least squares SVM support vector machine for forecasting tourism demand. International Journal of Research and Reviews in Applied Sciences, 2010. 3(3): p. 274-279.
123. Xuemei, L., D. Lixing, and L. Lanlan. A novel building cooling load prediction based on SVR and SAPSO. in 2010 International Symposium on Computer, Communication, Control and Automation (3CA). 2010. IEEE.
124. Neto, A.H. and F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy and Buildings, 2008. 40(12): p. 2169-2176.
125. Li, Q., et al., Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. Energy Conversion and Management, 2009. 50(1): p. 90-96.
126. Liu, F., R.D. Findlay, and Q. Song. A neural network based short term electric load forecasting in Ontario Canada. in 2006 International Conference on Computational Inteligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA'06). 2006. IEEE.
127. Che, J., J. Wang, and G. Wang, An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. Energy, 2012. 37(1): p. 657-664.
128. Chang, P.-C., C.-Y. Fan, and J.-J. Lin, Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach. International Journal of Electrical Power and Energy Systems, 2011. 33(1): p. 17-27.
129. Li, Q., P. Ren, and Q. Meng. Prediction model of annual energy consumption of residential buildings. in 2010 international conference on advances in energy engineering. 2010. IEEE.
130. Hou, Z., et al., Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. Applied energy, 2006. 83(9): p. 1033-1046.
131. Azadeh, A., et al., Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Applied Mathematics and Computation, 2007. 186(2): p. 1731-1741.
132. Li, K., H. Su, and J. Chu, Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study. Energy and Buildings, 2011. 43(10): p. 2893-2899.
133. Ghanbari, A., S. Abbasian-Naghneh, and E. Hadavandi. An intelligent load forecasting expert system by integration of ant colony optimization, genetic algorithms and fuzzy logic. in 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM). 2011. IEEE.
134. Mishra, S. and S.K. Patra. Short term load forecasting using neural network trained with genetic algorithm & particle swarm optimization. in 2008 First International Conference on Emerging Trends in Engineering and Technology. 2008. IEEE.

135. Xu, H., et al., Application of GMDH to short-term load forecasting, in Advances in Intelligent Systems. 2012, Springer. p. 27-32.
136. Sun, W., Y.-X. Zhang, and F.-T. Li. The neural network model based on pso for short-term load forecasting. in 2006 International Conference on Machine Learning and Cybernetics. 2006. IEEE.
137. ShangDong, Y. and L. Xiang. A new ANN optimized by improved PSO algorithm combined with chaos and its application in short-term load forecasting. in 2006 International Conference on Computational Intelligence and Security. 2006. IEEE.
138. Lu, N. and J. Zhou. Particle swarm optimization-based RBF neural network load forecasting model. in 2009 Asia-Pacific Power and Energy Engineering Conference. 2009. IEEE.
139. Niu, D., H. Wang, and C. Cai. Application of Neural Network-based Combining Forecasting Model Optimized by Ant Colony In Power Load Forecasting. in 2010 Asia-Pacific Power and Energy Engineering Conference. 2010. IEEE.
140. Wai, R.-J., Y.-C. Huang, and Y.-C. Chen. Intelligent daily load forecasting with fuzzy neural network and particle swarm optimization. in 2012 IEEE International Conference on Fuzzy Systems. 2012. IEEE.
141. He, Y. and Q. Xu. Short-term power load forecasting based on self-adapting PSO-BP neural network model. in 2012 Fourth International Conference on Computational and Information Sciences. 2012. IEEE.
142. Jiang, X., et al., Forecasting electrical energy consumption of equipment maintenance using neural network and particle swarm optimization. Mathematical Problems in Engineering, 2013. 2013.
143. Lee, C.-M. and C.-N. Ko, Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. Neurocomputing, 2009. 73(1-3): p. 449-460.
144. Mao, H., et al., Short-term and midterm load forecasting using a bilevel optimization model. IEEE Transactions on Power Systems, 2009. 24(2): p. 1080-1090.
145. Donate, J.P., et al., Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm. Neural Computing and Applications, 2013. 22(1): p. 11-20.
146. Han, M., et al., A novel reinforcement learning method for improving occupant comfort via window opening and closing. Sustainable Cities and Society, 2020: p. 102247.
147. Xue, J., Z. Xu, and J. Watada, Building an integrated hybrid model for short-term and mid-term load forecasting with genetic optimization. International Journal of Innovative Computing, Information and Control, 2012. 8(10): p. 7381-7391.
148. Nie, H., et al., Hybrid of ARIMA and SVMs for short-term load forecasting. Energy Procedia, 2012. 16: p. 1455-1460.
149. Wu, J., et al., Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model. Energy Conversion and Management, 2013. 70: p. 1-9.
150. Hernández, L., et al., A multi-agent system architecture for smart grid management and forecasting of energy demand in virtual power plants. IEEE Communications Magazine, 2013. 51(1): p. 106-113.
151. Brown, M., C. Barrington-Leigh, and Z. Brown, Kernel regression for real-time building energy analysis. Journal of Building Performance Simulation, 2012. 5(4): p. 263-276.
152. Xuemei, L., et al. A novel air-conditioning load prediction based on ARIMA and BPNN model. in 2009 Asia-Pacific Conference on Information Processing. 2009. IEEE.
153. Fan, G.-F., et al., Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. Neurocomputing, 2016. 173: p. 958-970.
154. Li, P., et al., Application of a hybrid quantized Elman neural network in short-term load forecasting. International Journal of Electrical Power and Energy Systems, 2014. 55: p. 749-759.
155. Chaturvedi, D., A. Sinha, and O. Malik, Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network. International Journal of Electrical Power and Energy Systems, 2015. 67: p. 230-237.
156. Ghofrani, M., et al., A hybrid short-term load forecasting with a new input selection framework. Energy, 2015. 81: p. 777-786.
157. Moazzami, M., A. Khodabakhshian, and R. Hooshmand, A new hybrid day-ahead peak load forecasting method for Iran's National Grid. Applied Energy, 2013. 101: p. 489-501.
158. Mena, R., et al., A prediction model based on neural networks for the energy consumption of a bioclimatic building. Energy and Buildings, 2014. 82: p. 142-155.
159. Wang, X. and M. Meng, A Hybrid Neural Network and ARIMA Model for Energy Consumption Forcasting. JCP, 2012. 7(5): p. 1184-1190.

160. Tang, F., A. Kusiak, and X. Wei, Modeling and short-term prediction of HVAC system with a clustering algorithm. Energy and Buildings, 2014. 82: p. 310-321.
161. Yokoyama, R., T. Wakui, and R. Satake, Prediction of energy demands using neural network with model identification by global optimization. Energy Conversion and Management, 2009. 50(2): p. 319-327.
162. Xuemei, L., et al. Hybrid genetic algorithm and support vector regression in cooling load prediction. in 2010 Third International Conference on Knowledge Discovery and Data Mining. 2010. IEEE.
163. Aqlan, F., et al. Integrating artificial neural networks and cluster analysis to assess energy efficiency of buildings. in IIE Annual Conference. Proceedings. 2014. Institute of Industrial and Systems Engineers (IISE).
164. Prahastono, I., D. King, and C.S. Ozveren. A review of electricity load profile classification methods. in 2007 42nd international universities power engineering conference. 2007. IEEE.
165. Yang, J., et al., k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. Energy and Buildings, 2017. 146: p. 27-37.
166. Santamouris, M., et al., Using intelligent clustering techniques to classify the energy performance of school buildings. Energy and Buildings, 2007. 39(1): p. 45-51.
167. Xiao, F. and C. Fan, Data mining in building automation system for improving building operational performance. Energy and Buildings, 2014. 75: p. 109-118.
168. Jalori, S. and T.A. Reddy, A New Clustering Method to Identify Outliers and Diurnal Schedules from Building Energy Interval Data. ASHRAE Transactions, 2015. 121(2).
169. Yu, Z., et al., A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and Buildings, 2011. 43(6): p. 1409-1417.
170. Petcharat, S., S. Chungpaibulpatana, and P. Rakkwamsuk, Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings. Energy and Buildings, 2012. 52: p. 145-152.
171. Miller, C., et al., The ASHRAE Great Energy Predictor III competition: Overview and results. Science and Technology for the Built Environment, 2020: p. 1-21.
172. Bogatu, A.T., COST–EFFECTIVE DATA WRANGLING IN DATA LAKES. 2020, The University of Manchester.
173. Levermore, G.J., Building energy management systems: applications to low-energy HVAC and natural ventilation control. 2000: Taylor & Francis.
174. Ahmad, M.W., et al., Building energy metering and environmental monitoring–A state-of-the-art review and directions for future research. Energy and Buildings, 2016. 120: p. 85-102.
175. Plageras, A.P., et al., Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings. Future Generation Computer Systems, 2018. 82: p. 349-357.
176. Yang, J., M. Santamouris, and S.E. Lee, Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. Energy and Buildings, 2016. 121: p. 344-349.
177. Zhao, Y., et al., Virtual occupancy sensors for real-time occupancy information in buildings. Building and Environment, 2015. 93: p. 9-20.
178. Zou, H., et al., Non-intrusive occupancy sensing in commercial buildings. Energy and Buildings, 2017. 154: p. 633-643.
179. Zhou, K., C. Fu, and S. Yang, Big data driven smart energy management: From big data to big insights. Renewable and Sustainable Energy Reviews, 2016. 56: p. 215-225.
180. Osterlind, F., et al. Integrating building automation systems and wireless sensor networks. in 2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007). 2007. IEEE.
181. Mainetti, L., L. Patrono, and A. Vilei. Evolution of wireless sensor networks towards the internet of things: A survey. in SoftCOM 2011, 19th international conference on software, telecommunications and computer networks. 2011. IEEE.
182. Buyya, R., J. Broberg, and A.M. Goscinski, Cloud computing: Principles and paradigms. Vol. 87. 2010: John Wiley & Sons.
183. Domingues, P., et al., Building automation systems: Concepts and technology review. Computer Standards and Interfaces, 2016. 45: p. 1-12.
184. Mantha, B.R., C.C. Menassa, and V.R. Kamat, A taxonomy of data types and data collection methods for building energy monitoring and performance simulation. Advances in Building Energy Research, 2016. 10(2): p. 263-293.
185. Yang, H., et al., Advanced wireless sensor networks for sustainable buildings using building ducts. Sustainability, 2018. 10(8): p. 2628.

186. van de Kaa, G., H.J. de Vries, and J. Rezaei, Platform selection for complex systems: Building automation systems. Journal of Systems Science and Systems Engineering, 2014. 23(4): p. 415-438.
187. Kazmi, A.H., et al., A review of wireless-sensor-network-enabled building energy management systems. ACM Transactions on Sensor Networks, 2014. 10(4): p. 1-43.
188. Bashir, M.R. and A.Q. Gill. Towards an IoT big data analytics framework: smart buildings systems. in 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). 2016. IEEE.
189. Koseleva, N. and G. Ropaite, Big data in building energy efficiency: understanding of big data and main challenges. Procedia Engineering, 2017. 172(2017): p. 544-549.
190. Mathew, P.A., et al., Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. Applied Energy, 2015. 140: p. 85-93.
191. Balaji, B., et al. Brick: Towards a unified metadata schema for buildings. in Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments. 2016.
192. Haystack, P., Project haystack. 2018.
193. Charpenay, V., et al. An ontology design pattern for iot device tagging systems. in 2015 5th International Conference on the Internet of Things (IOT). 2015. IEEE.
194. Rasmussen, M., et al. Recent changes in the building topology ontology. 2017.
195. Daniele, L., F. den Hartog, and J. Roes. Created in close interaction with the industry: the smart appliances reference (SAREF) ontology. in International Workshop Formal Ontologies Meet Industries. 2015. Springer.
196. Zichi, A. and M. Torri, Green BIM and CIM: Sustainable Planning Using Building Information Modelling, in Green Planning for Cities and Communities. 2020, Springer. p. 383-409.
197. Crone, S.F., J. Guajardo, and R. Weber. The impact of preprocessing on support vector regression and neural networks in time series prediction. in DMIN. 2006.
198. Li, X., C.P. Bowers, and T. Schnier, Classification of energy consumption in buildings with outlier detection. IEEE Transactions on Industrial Electronics, 2009. 57(11): p. 3639-3644.
199. Huang, G., et al., A data fusion scheme for building automation systems of building central chilling plants. Automation in Construction, 2009. 18(3): p. 302-309.
200. Allison, P.D., Missing data. Vol. 136. 2001: Sage publications.
201. Liu, L. and M.T. Özsu, Encyclopedia of database systems. Vol. 6. 2009: Springer New York, NY, USA:.
202. Seem, J.E., Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy and Buildings, 2007. 39(1): p. 52-58.
203. Ng, A., Machine Learning and AI via Brain simulations. 2013.
204. Lam, J.C., et al., Principal component analysis and long-term building energy simulation correlation. Energy Conversion and Management, 2010. 51(1): p. 135-139.
205. Cunningham, P., Dimension Reduction, Technical report UCD-CSI-2007-7. University College Dublin, 2007.
206. Li, K., et al., Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energy and Buildings, 2015. 108: p. 106-113.
207. Reddy, T.A. and D. Claridge, Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption. Energy and Buildings, 1994. 21(1): p. 35-44.
208. Chen, T. and Y.-C. Wang, Long-term load forecasting by a collaborative fuzzy-neural approach. International Journal of Electrical Power and Energy Systems, 2012. 43(1): p. 454-464.
209. Xuemei, L., et al. A novel hybrid approach of KPCA and SVM for building cooling load prediction. in 2010 Third International Conference on Knowledge Discovery and Data Mining. 2010. IEEE.
210. Guyon, I. and A. Elisseeff, An introduction to variable and feature selection. Journal of machine learning research, 2003. 3(Mar): p. 1157-1182.
211. Leung, M., et al., The use of occupancy space electrical power demand in building cooling load prediction. Energy and Buildings, 2012. 55: p. 151-163.
212. Krarti, M., Energy audit of building systems: an engineering approach. 2016: CRC press.
213. Li, Q., et al., Applying support vector machine to predict hourly cooling load in the building. Applied Energy, 2009. 86(10): p. 2249-2256.
214. Yang, J., H. Rivard, and R. Zmeureanu, On-line building energy prediction using adaptive artificial neural networks. Energy and Buildings, 2005. 37(12): p. 1250-1259.

215. Dodier, R.H. and G.P. Henze, Statistical analysis of neural networks as applied to building energy prediction. Journal of solar energy engineering, 2004. 126(1): p. 592-600.
216. Zhao, H.-x. and F. Magoulès, Feature selection for predicting building energy consumption based on statistical learning method. Journal of Algorithms and Computational Technology, 2012. 6(1): p. 59-77.
217. Tsanas, A. and A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 2012. 49: p. 560-567.
218. Ceperic, E., V. Ceperic, and A. Baric, A strategy for short-term load forecasting by support vector regression machines. IEEE Transactions on Power Systems, 2013. 28(4): p. 4356-4364.
219. Kolter, J.Z. and J. Ferreira Jr, A large-scale study on predicting and contextualizing building energy usage. 2011.
220. Son, H. and C. Kim, Forecasting short-term electricity demand in residential sector based on support vector regression and fuzzy-rough feature selection with particle swarm optimization. Procedia engineering, 2015. 118: p. 1162-1168.
221. Jurado, S., et al., Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy, 2015. 86: p. 276-291.
222. Zhang, L. and J. Wen, A systematic feature selection procedure for short-term data-driven building energy forecasting model development. Energy and Buildings, 2019. 183: p. 428-442.
223. Kouhi, S., F. Keynia, and S.N. Ravadanegh, A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection. International Journal of Electrical Power and Energy Systems, 2014. 62: p. 862-867.
224. Zhang, L., et al., A systematic feature extraction and selection framework for data-driven whole-building automated fault detection and diagnostics in commercial buildings. Building and Environment, 2020. 186: p. 107338.
225. Roth, K.W., et al., Energy impact of commercial building controls and performance diagnostics: market characterization, energy impact of building faults and energy savings potential. Prepared by TAIX LLC for the US Department of Energy. November. 412pp (Table 2–1), 2005.