A Review of the-State-of-the-Art in Data-driven Approaches for Building Energy Prediction

Ying Sun[a], Fariborz Haghighat[a]*, and Benjamin C. M. Fung[b]

[a] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Canada
[b] School of Information Studies, McGill University, Montreal, Canada

Corresponding author: Fariborz.Haghighat@concordia.ca

Official publication: https://doi.org/10.1016/j.enbuild.2020.110022

Abstract

Building energy prediction plays a vital role in developing a model predictive controller for consumers and optimizing energy distribution plan for utilities. Common approaches for energy prediction include physical models, data-driven models and hybrid models. Among them, data-driven approaches have become a popular topic in recent years due to their ability to discover statistical patterns without expertise knowledge. To acquire the latest research trends, this study first summarizes the limitations of earlier reviews: seldom present comprehensive review for the entire data-driven process for building energy prediction and rarely summarize the input updating strategies when applying the trained data-driven model to multi-step energy prediction. To overcome these gaps, this paper provides a comprehensive review on building energy prediction, covering the entire data-driven process that includes feature engineering, potential data-driven models and expected outputs. The distribution of 105 papers, which focus on building energy prediction by data-driven approaches, are reviewed over data source, feature types, model utilization and prediction outputs. Then, in order to implement the trained data-driven models into multi-step prediction, input updating strategies are reviewed to deal with the time series property of energy related data. Finally, the review concludes with some potential future research directions based on discussion of existing research gaps.

Keywords: Data-driven approach; Machine learning; Energy prediction; Building

1. Introduction

1.1. Motivation for building energy prediction using data-driven approaches

The buildings and buildings construction sectors consumed 36% of the global final energy and nearly 40% of total CO_2 emission in 2018 [1]. Moreover, these percentages are expected to further rise. To lower the environmental and economic burden caused by the increasing building energy demand, improving the energy efficiency of buildings would be an effective solution. The energy saving potential of energy conservation measures could be quantified and compared through building energy prediction [2], [3]. This indicates that building energy prediction could be utilized as a tool for designing and selecting proper energy conservation methods. Besides, multi-step energy prediction could be integrated into a model predictive controller to predefine an optimized HVAC operation schedule in order to achieve peak shifting or energy/cost saving. Furthermore, accurate energy demand forecasting makes it possible for utilities to optimize energy distribution plan and for governments to formulate standards for energy saving.

Common approaches to predict building energy performance mainly include physical models ("white box"), hybrid methods ("grey box") and data-driven approaches ("black box"). Physical models predict the thermal behavior by numerical equations by considering detailed physical properties of building materials and characteristics. Plenty of energy prediction software have been developed and implemented, such as EnergyPlus, TRNSYS, DOE-2, eQUEST, DeST, etc. A detailed review of these physical models is available in [4]–[6]. The main advantage of physical models is the ability to describe heat transfer mechanisms, while their disadvantages include: (1) requirement of expertise; (2) difficulties in making proper assumptions; (3) time-consuming; and (4) inability to adapt to environmental/social-economic vicissitudes. Hybrid methods combine physical models and data-driven approaches to simulate building energy. For instance, Dong et al. [7] integrated data-driven techniques into a physical model to forecast hour and day ahead load for a residential building. Their study shows that the hybrid model improves the prediction accuracy and reduces the computational complexity of traditional physical models. However, hybrid models still face the issues which usually present in physical models, such as improper assumption and requirement of expertise. In contrast, data-driven approaches show the ability to overcome above mentioned limitations of physical models and hybrid methods, due to their ability of discovering statistical patterns from the available dataset instead of on-site physical information. Therefore, recently, data-driven approaches have drawn significant attention in building energy prediction.

1.2. Literature reviews

As the application of data-driven approach on building energy prediction attracts more attention, a variety of review papers have been published in recent years on this topic. To better understand the latest research interest, the content of review papers published after 2013 is summarized in Table 1. This table is organized in the order of general data-driven modeling procedures: (1) feature engineering which is a process of preparing transforming, constructing, and filtering features with the goal of optimizing the performance of a data analysis task. In this part, whether potential feature types for building energy prediction and feature extraction methods are introduced by these review papers is summarized; (2) data-driven algorithms. The presence of reviews about commonly used models, e.g. Linear Regression (LR), AutoRegression-Moving Average (ARMA) and AutoRegression Integrated Moving Average (ARIMA), Regression Tree (RT), Support Vector Machine (SVM), Artificial Neural Network (ANN) and Ensemble models that include boosting and bagging, etc., is listed; and (3) factors considered for expected outputs (i.e. temporal granularity, scale, energy type, building type and validation criteria). Clearly distinguishing these aspects could give an inspiration to feature engineering as well as data-driven model selection. Besides, proper selection and application of validation criteria could ensure the prediction accuracy and generalization of trained models for building energy prediction. Therefore, whether existing review papers summarize these aspects are shown in Table 1.

Note that data collection and data cleaning are generally considered as steps needed prior to research and are rarely reviewed in previous review papers; thus, these aspects are not included in Table 1.

Table 1: Review contents in terms of general data-driven procedures

Reference	Year	Feature		Data-driven algorithms							Factors considered for expected outputs				
		Type	Extraction method	LR	ARMA and ARIMA	RT	SVM	ANN	Ensemble method	Other	Building type	Energy type	Scale	Temporal granularity	Criteria
[6]	2013	×	×	×	×	×	√	√	×	×	×	×	×	×	×
[8]	2014	×	×	×	×	×	√	√	Hybrid	×	×	×	×	×	×
[9]	2016	×	×	√	√	×	√	√	√	Semi-parametric additive models; exponential smoothing models; Fuzzy regression models	×	×	Electric utility	√	×
[10]	2017	√	×	√	√	×	√	√	×	×	×	×	√	√	×
[11]	2017	√	√	√	√	√	√	√	×	×	Commercial	√	×	×	×
[12]	2017	×	×	√	√	×	√	√	Hybrid	×	×	×	×	×	×
[13]	2017	√	×	√	×	×	√	√	√	×	√	√	×	√	×
[14]	2017	×	×	×	√	×	√	√	Hybrid	Fuzzy time series; moving average and exponential smoothing, k-Nearest Neighbor (kNN)	×	×	×	×	×
[15]	2018	√	√	√	√	×	√	√	×	×	√	√	×	√	√
[16]	2018	√	×	√	×	√	√	√	×	Clustering	×	×	×	×	×
[17]	2018	×	×	×	√	×	√	√	×	Clustering	×	×	Urban and rural	×	×
[18]	2019	×	×	×	×	×	×	√	×	×	×	√	×	×	√
[19]	2019	√	×	√	√	√	√	√	√	kNN	×	×	×	×	√
[20]	2019	×	×	√	√	×	√	√	×	Deep learning	×	×	√	√	√

Note:

1. '√' means the literature includes the corresponding contents, while '×' means exclusion.
2. Hybrid model refers to the integration of two data-driven models, instead of the grey box model.

Limitations of existing literature reviews are summarized below:

(1) Comprehensive review for the entire data-driven process in the field of energy prediction was missing. No existing review papers in Table 1 covers all aspects in feature engineering, data-driven algorithms and expected outputs. The missing part mainly include:

   i. Data utilized by the reviewed studies to predict building energy through data-driven algorithms are rarely summarized. However, the number of data points utilized for model training and validation, number of meters and buildings for data collection, as well as accessibility of data would affect the reproducibility and generalization of studied techniques.
   ii. The time series property of energy related data was not highlighted. For instance, Kuster et al. [10] reviewed the number of papers that utilized the time index for different horizon and scale predictions, but they did not consider the situation that utilized historical data as one of inputs.
   iii. Systematic review for feature extraction methods was missing. For instance, Yildiz et al. [11] introduced several feature selection algorithms, but the fundamentals of each method as well as their advantages and disadvantages still need to be further summarized.
   iv. Relatively novel technologies were generally not included. For instance, autoencoders were not reviewed as feature extraction methods, while deep learning and ensemble methods were not well revised when summarizing data-driven algorithms.
   v. Factors (such as temporal granularity, scale, energy type and building type, criteria) reflected from prediction outputs were rarely reviewed at the same time.
   vi. For temporal granularity, the difference between time horizon (the length of time-ahead energy prediction) and time resolution (duration of a time step) was not clearly distinguished. Many existing literature reviews, such as references [15] and [13], focus mainly on time horizon; therefore, the effect of prediction steps (i.e. time horizon divided by time resolution) on prediction accuracy could not be analyzed.

(2) No review summarized the input updating strategies when applying the trained data-driven model in realistic multi-step energy prediction. As time series data, historical energy consumption would affect the predicted future values. The most recent historical energy consumption data lies within the prediction horizon when doing multi-step energy prediction. Thus, the problem of how to deal with the effect of unmeasurable most recent historical data should be solved by proper updating strategies.

1.3. Objectives, contributions and structure of the review

To overcome limitations in existing literature, this paper aims to provide a comprehensive review of building energy prediction data-driven approaches. To be specific, the objectives of this paper include: (1) to give a systematic and comprehensive overview for developing data-driven models to predict building energy consumption; (2) to summarize input updating strategies for applying the developed data-driven model to achieve realistic energy prediction; (3) to highlight future research opportunities in the field of building energy prediction with data-driven approaches.

The main contributions of this paper can be summarized as following: (1) Present a comprehensive review for the entire procedure of data-driven energy prediction approaches. Potential feature types for energy prediction are listed. Commonly-used and novel feature extraction methods and data-driven algorithms are reviewed in terms of principles and their strengths and weakness. Besides, factors reflected from the expected outputs are summarized and clearly distinguished. Then, the distribution of studies from 2015 to 2019 is reviewed for better understanding the recent research interest. (2) Summarize input updating strategies for multi-step energy prediction by the developed data-driven models. These strategies could solve the following problems: i. Whether to consider the effect of historical energy consumption data; ii. How to consider the effect of historical data; iii. How to consider the effect of most recent unmeasurable historical data lied in the prediction horizon.

This paper is organized as follows: Section 2 gives a comprehensive review for the general procedure of developing data-driven energy prediction models, which mainly include feature engineering, data-driven algorithms and factors reflected from expected outputs. Then, Section 3 summarizes four input updating strategies for multi-step building energy prediction. Section 4 presents conclusions and opportunities for future work.

2. General data-driven modeling procedure

Before a data-driven procedure, data is first collected from simulation, measurement/survey, or public database. Then, the data should be thoroughly processed to remove/correct the missing/incorrect data. This process is called data cleaning. Commonly utilized outlier/anomaly detection methods can be found in [21], [22], while approaches to impute/replace missing data were presented [23].

After data collection and data cleaning, features contributing most to prediction results need to be constructed and extracted. Therefore, in Section 2.1, the most commonly used features for energy prediction and feature extraction methods are presented. After data preparation, proper data-driven algorithms should be selected and trained. A summary for data-driven algorithms is shown in Section 2.2. The developed data-driven models could be utilized for building energy prediction after validation. Factors reflected from the expected prediction outputs are introduced in Section 2.3.

2.1. Feature engineering

In this section, potential features types that contribute to building energy consumption are firstly introduced in Section 2.1.1. Then, feature extraction methods which select valuable features or reconstruct feature vectors are summarized in Section 2.1.2.

2.1.1. Feature types

2.1.1.1. Meteorological information

Meteorological information mainly includes ambient dry bulb temperature, wet bulb temperature, dew point temperature, humidity, wind speed, solar radiation, rainfall, air pressure, etc. [13].

Prior to data-driven model construction, the correlation between weather variables and building load (except heating load) has been studied by Cai et al. [24] for three buildings located in Alexandria VA, Shirley NY, and Uxbridge MA, respectively. Among these weather variables, outdoor temperature was found to be positively correlated to building load, while the relation between other variables and building load were insignificant. However, when the ambient temperature was lower than 24.4 °C, it was found to be irrelevant to electricity demand of residential buildings in Italy [25]. This is because the main heating fuel for homes in Italy is natural gas, while electricity is used for cooling systems. Besides ambient temperature, Solar radiation is also commonly utilized in building energy prediction, due to its significant effect on thermal demand and its accessible from weather forecasting [26].

2.1.1.2. Indoor environmental information

Except for weather information, indoor conditions that include set-point temperature of thermostats, indoor temperature, indoor humidity, indoor carbon dioxide concentration, etc. have been identified as a priority for residential cooling and heating load calculation [27]. Note that unlike constant design values of indoor conditions during design stage, these values are dynamic during reality operation. Therefore, to predict building loads precisely, indoor environmental information needs to be considered as a potential feature.

Chammas et al. [28] considered indoor temperature and humidity when predicting energy consumption for a residential house. However, their study did not compare the importance of meteorological information, indoor conditions and time indexes (which will be introduced in Section 2.1.1.4). Ding et al. [29] presented that interior variables would further improve heating load prediction accuracy. However, due to unpredictable interior temperature, the variables could not be utilized for 24 hour ahead heating demand prediction. Wei et al. [30] found that indoor relative humidity, dry-bulb temperature and carbon dioxide concentration are among the top 10 important variables for energy consumption prediction of an office building. These three features were also included for predicting desk fan usage preferences[31]. Furthermore, indoor temperature and humidity have been used as inputs in predicting air conditioning operation[32].

It is interesting to note that studies with considering set-point temperature as inputs for predicting building loads generally aimed at developing demand response control strategy, such as in the research by Behl et al. [33]. Otherwise, studies tend to ignore the effect of set-point temperature in energy prediction accuracy, even though residents in residential buildings have the ability to adjust the set-point temperature to meet their thermal comfort requirement and save energy [34].

2.1.1.3.Occupancy related data

Occupancy related data, such as number of occupants and types of occupant activities, would affect internal gain and then influence the pattern of energy usage [35], [36]. Therefore, it would be a potential feature for building energy prediction.

The principal component analysis of Wei et al. [30] indicated that the number of occupants is even more important than meteorological information for energy prediction in an office building. Wang et al. [37] utilized linear regression to observe the strong linear relation between plug load power and occupant count for working days, and then selected it as one of features for plug load prediction. Sala-Cardoso et al. [38] predicted the activity indicator through a recurrent neural network (RNN) and then integrated it with a power demand prediction model to improve the prediction accuracy of HVAC thermal power demand for a research building.

However, short leave of occupants would not affect the load consumption. Besides, if a public building is controlled without taking into account the occupancy status, its energy consumption might not be strongly related to occupancy patterns [39]. Furthermore, in most cases, the types of occupant activities are not flexible to be collected.

2.1.1.4.Time index

Time index means the stamps series for time, which mainly include time of the day, day of the week, hour type (peak hour or off-peak hour), day type (weekday or weekends), calendar day, etc. The purpose of introducing time index into energy prediction is to indicate the occupancy related effect. For instance, occupants tend to do similar activities at the same time on different days or at the same day on different weeks. Therefore, time index would be a good option when occupancy related data is unavailable.

Fan et al. [40] found that due to the similar energy consumption pattern on the same weekday, 7-days and 14-days ahead peak power demand and energy consumption were the four most important inputs for next-day energy consumption and peak power demand prediction of a commercial building. This indicates that day of the week could be selected as one of the input features able to represent similar energy consumption patterns during the same weekday. Similar justification could be used for selecting time of the day, holiday/workday, peak hour/off-peak hour as inputs.

2.1.1.5.Building characteristic data

Building characteristic features mainly include relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, heat transfer coefficient of building envelopes, absorption coefficient for solar radiation of exterior walls, window-wall ratio, shading coefficient etc. [15].

Once a building is constructed, these data would remain relatively constant. Therefore, it is meaningless to contain this information when using data-driven models to predict dynamic load for a specific building. However, when the study object is multiple buildings or when the objective is using the known load of an existing building to predict the load of a new building, building characteristic features would be beneficial. Seyedzadeh et al. [41] drew feature correlation maps for building characteristic and building heating/cooling loads, and utilized these features as input for data-driven models to predict building loads. Wei et al. [42] predicted annual heating, cooling and electricity intensity for different office buildings based on input factors relevant to building form, e.g. aspect ratio, window-wall ratio, number of floors, orientation and building scale. Talebi et al. [43] utilized thermal mass as one of input features to predict heating demand of a district. Similar studies could also be found in references [44]–[47].

2.1.1.6.Socio-economic information

Socio-economic information shows the socio-economic situation of the studied area [10]. It mainly includes income, electricity price, GDP, population, etc.

These features are commonly utilized to do long term (e.g. months or years) load prediction for large scale (e.g. district, region or country) [48]. For instance, He et al. [49] found that average electricity price and number of electricity customers/permanent residents could be important in forecasting annual electricity consumption of a city. He et al. [50] identified that historical energy consumption, average annual GDP growth rate and total GDP were the key factors for annual energy consumption prediction of Anhui province, China. However, GDP was revealed by Beyca et al. [51] to be insignificant in natural gas consumption prediction of Istanbul, while price of natural gas and population showed a high correlation to the prediction result.

2.1.1.7.Historical data

Due to the thermal mass of building envelopes, building loads could be affected by historical factors, such as historical values of exogenous features or historical energy consumption. For example, Wang et al. [52] found that the historical heating consumption is the leading factor for heating demand prediction of district heated apartment buildings. Similarly, Ahmad et al. [53] concluded that previous hour's electricity consumption was more important than meteorological information, time index and occupancy related data for 1-hour-ahead HVAC energy consumption prediction of a hotel in Spain. Ding et al. [54] proved that historical cooling capacity data are the most important data for cooling load prediction of an office building. Huang et al. [55] proposed a historical energy comprehensive variable named EVMA to improve the energy demand prediction accuracy for residential buildings based on ensemble methods. Furthermore, He et al. [50] found the historical annual energy consumption of Anhui province in China significantly affected its future annual energy consumption. Due to the ability to increase the prediction accuracy of dynamic loads, the interests in applying historical data as features for data-driven models have been increasing in recent years.

2.1.2. Feature extraction methods

Properly constructed features could reduce the computation time of a data-driven model without sacrificing prediction accuracy [56]. The commonly applied feature extraction methods with the ability to select useful features or reconstruct feature vectors are introduced in the following sections.

2.1.2.1.Variable ranking

The idea of variable ranking is to choose the desired number of features most relevant to the output (i.e. building energy consumption/demand) by a scoring function.

In terms of energy prediction, one popular function for variable ranking is the Pearson correlation coefficient (see Equation 1 [57]) for its quick and easy use. This method determines the strength and direction of the linear relationship between two variables. To calculate the monotonic relationship between two continuous or ordinal variables, Spearman's rank correlation (see Equation 2 [58]) could be utilized. Note that Spearman's rank correlation between two variables equals to the Pearson correlation of rank values of these two variables.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}} \tag{1}$$

where:

r_{xy} is the Pearson correlation coefficient between input feature *x* and target output *y*;

n is sample size;

x_i, y_i are the *i*-th individual sample points;

$\overline{x}$, $\overline{y}$ are the mean value of input feature and target output, respectively.

$$rho_{xy} = \frac{\sum_{i=1}^{n}(x_i'-\overline{x'})(y_i'-\overline{y'})}{\sqrt{\sum_{i=1}^{n}(x_i'-\overline{x'})^2}\sqrt{\sum_{i=1}^{n}(y_i'-\overline{y'})^2}} \tag{2}$$

where:

rho_{xy} is the Spearman's ranking correlation between input feature *x* and target output *y*;

n is sample size;

x_i', y_i' are the ranks of i-th individual sample points;

$\overline{x'}$, $\overline{y'}$ are the mean rank values of input feature and target output, respectively.

One challenge for variable ranking is to determine the desired number of features, which could be considered as a hyperparameter (i.e. a pre-defined parameter which affects the running time of feature engineering process and prediction accuracy of the developed data-driven model [59]). Another drawback of variable ranking is that it could only calculate the relationship between individual variables and output, instead of between subsets of features and output. For instance, Aaron et al. [60] utilized standardized association factors to find out that dry bulb temperature, wet bulb temperature and enthalpy are most relevant to building electricity use. However, they failed to estimate the possible inter-relevance between temperatures and enthalpy. To solve this problem, filter and wrapper methods could be utilized to select the best subset.

2.1.2.2.Filter and Wrapper methods

Both filter and wrapper methods could be utilized for best-subset selection, which means they could consider the interrelationship between features. Among them, filter methods evaluate the importance of individual or subset of features through statistical measures. Filter methods have two different categories: Rank Based (i.e. variable ranking) and Subset Evaluation Based [61]. The filter methods mentioned here refer to the later types, since the former one was described in Section 2.1.2.1. Unlike filter methods, wrapper methods consider all possible subsets of features and measure their performance through supervised learning algorithms.

Filter methods are more efficient than wrapper techniques in terms of computational complexity, while wrapper methods are more stable [61]. Yuan et al. [62] applied partial least squares regression (PLSR) and random forests (RF) to rank the top 10 important input features for predicting weekly coal consumption for space heating. The reason for employing these two filter methods is that they can consider the inter-dependence between input variables. Then, they utilized an SVM based wrapper method to evaluate the proper number of features. The prediction accuracy based on the selected top 6 features met the requirement of ASHRAE Guidelines 14-2014 [63].

2.1.2.3.Embedded method

Unlike the wrapper method, which selects the best subset with the highest prediction performance in a specific learning algorithm, the embedded method integrates feature selection into the learning algorithm. For instance, regularization added to data-driven models could be considered as an embedded method. Jain et al. [64] employed Lasso, a linear regression model which adds an L1 penalty to the squared error loss, to forecast energy consumption of a multi-family residential building. Their results confirmed that in certain cases, Lasso could outperform a Support Vector Regression (SVR) model that did not consider feature selection.

One challenge of the embedded method is that the selected regularization method should adapt the optimization procedure to ensure the existence of optimum solution. Furthermore, this method could not present the importance of features.

2.1.2.4.Principal component analysis (PCA)

The idea of traditional PCA is to project features into a lower-dimensional sub-space with linearly uncorrelated variables [65], while kernel PCA utilizes a kernel function to map nonlinear related original inputs into a new feature space and then perform a linear PCA in this new space [66].

Li et al. [67] compared the building load prediction accuracy between SVR with PCA, SVR with kernel PCA, and SVR without any feature selection techniques. Their results illustrate that SVR with PCA increased the cooling load prediction accuracy compared to the SVR model, while kernel PCA could further improve prediction performance.

Furthermore, Yuldiz et al. [11] showed the way to apply PCA to tackle the multi-collinearity problem in original input variables, and gave a detailed description about how to determine the dimension of reduced feature space. A similar application of PCA has been introduced by Wei et al. [30]. From these studies, one limitation of PCA has been revealed: the dimension of final feature space needs to be manually selected. Besides, when applying kernel, the type of kernel function should be determined.

2.1.2.5.Autoencoder (AE)

AE is a type of unsupervised artificial neural network (ANN) that can learn a compressed nonlinear representation of the input data. As shown in Figure 1, an autoencoder generally consists of two networks: (1) Encoder: maps the original inputs to a compressed low dimension; (2) Decoder: recovers original inputs from the compressed representation.

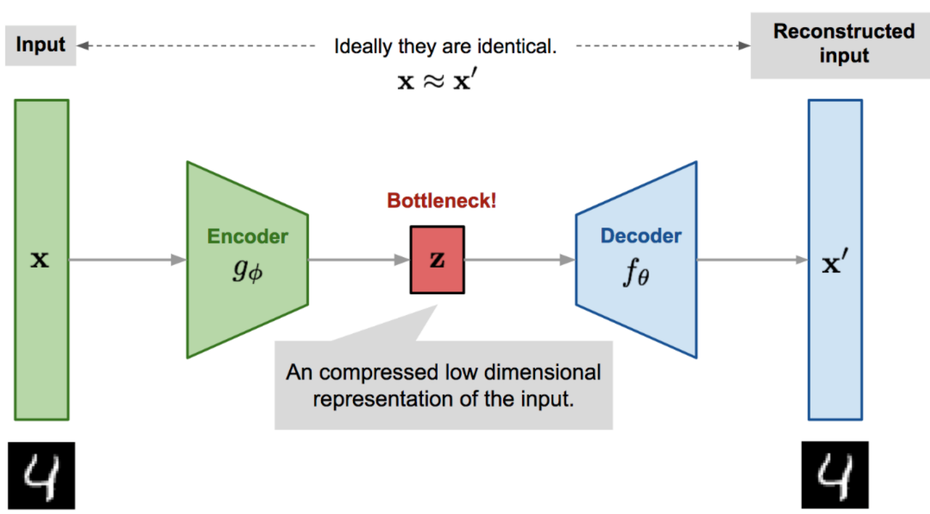


Figure 1 : Illustration of autoencoder model architecture [68]

Fan et al. [69] compared three types of deep learning based feature selection methods (i.e. fully connected AE, convolutional AE and generative adversarial networks) to variable ranking and PCA. The result shows that the deep learning-based feature selection method enhanced the one-step-ahead cooling load prediction performance for an educational building. Furthermore, Mujeeb and Javaid [70] proposed an efficient sparse autoencoder as feature extraction method, and then utilized the compressed feature space as inputs for an non-linear autoregressive network. The proposed method decreased forecasting error of the non-linear autoregressive network for regional load forecasting.

Note that the application of autoencoder for feature extraction in the field of building energy prediction is still uncommon. One reason is that the dimension of original input features is usually small, thus, AE would be computing intensively compared to other feature extraction methods. Following the explosive growth of collected data and implementation of deep learning, the interests in AE would increase.

Strengths and weaknesses of previous introduced feature selection methods are summarized in Table 2.

Table 2 : Strengths and weaknesses of feature selection methods

Type of feature selection	**Strengths**	**Weaknesses**
Variable ranking	1. Fastest and easiest to use 2. Quantitatively calculate the relevance between individual variables and outputs	1. Hard to determine number of desired features 2. Unavailable for considering the effect of inter-relevance between features on the output 3. Could not select the best subset
Filter method	1. Fast and easy to use 2. Subset selection 3. Robust to overfitting	1. Less stable
Wrapper method	1. Subset selection that considers inter-relevant of input features 2. More stable	1. Computational expensiveness 2. High risk of overfitting
Embedded method	1. Easy to use 2. Unnecessary to eliminate features	1. Unable to quantitatively present the importance of features
PCA	1. Relatively easy to use 2. Effective when original feature space dimension is not too large 3. Unnecessary to eliminate features	1. Hard to determine number of desired features 2. For kernel PCA, kernel function needs to be properly selected
AE	1. Learn nonlinear representation of original input 2. More powerful for compressing the dimension of features with lower loss of information	1. Computational expensiveness

2.2. Data-driven algorithms

Data-driven algorithms introduced in this paper are shown in Figure 2. A detailed description is presented in following sections.

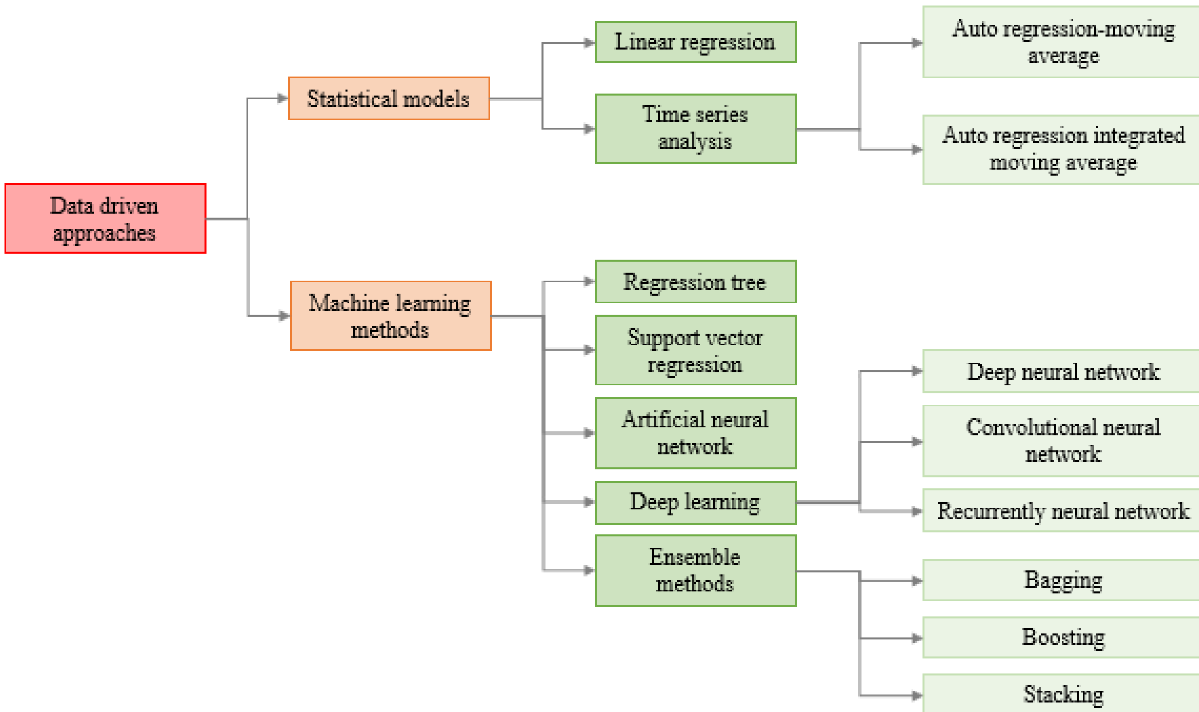


Figure 2 : Data-driven models for building energy prediction

2.2.1. Statistical models

2.2.1.1. Linear regression (LR)

Linear regression is one of the traditional statistical approaches to study the relationship between a dependent variable (i.e. response or output) and one or more independent variables (i.e. predictor or input features). Its general form is shown in Equation 3.

$$\hat{y} = w_0 + wx \tag{3}$$

where,

$\hat{y}$ is the predicted output,

w_0 is the bias term,

w is a weight matrix for features x.

Note that the general form could only discover the linear relationship between features and output. To extend the applicability of linear regression, the input variables could be converted to other forms through different active functions, such as polynomial (Equation 4) or natural logarithm function (Equation 5).

$$\hat{y} = w_0 + wx^m \tag{4}$$

where m means m-th polynomial.

$$\hat{y} = w_0 + wlog(x) \tag{5}$$

The main advantage of linear regression is that it is very easy to use and intuitive to understand. The contribution of individual variables on the prediction result could be directly found from the weight matrix. Besides, extended linear regression could be applied in solving nonlinear problems. However, its limitations should also be noted: (1) General form of linear regression could not consider nonlinear relationships between inputs and outputs; (2) The prediction performance of extended linear regression is highly dependent on the proper selection of active function, which could be a significant challenge; (3) Multicollinearity of input features would hurt the prediction result of linear regression. Therefore, feature extraction methods are recommended to be applied before developing linear regression models.

Applications of linear regression approach into building energy prediction have been sufficiently reviewed in the literature, see Table *1*.

2.2.1.2.Time series analysis

The most commonly used methods for time series analysis are AutoRegressive-Moving Average (ARMA) and AutoRegressive Integrated Moving Average (ARIMA) [10]. ARMA mainly includes two parts: an autoregressive model (AR) with order *p* and a moving average model (MA) with order *q*,

$$\hat{y_t} = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{6}$$

where $\varphi_1, \dots, \varphi_p$ are weights for AR, $\theta_1, \dots, \theta_q$ are weights for MA, ε is white noise, c is a constant.

ARMA could only handle stationary time series. When predicting nonstationary time series, ARIMA would be a better choice since it integrated an initial differencing step to eliminate the non-stationary [71].

ARMA and ARIMA show the ability to consider the effect of historical data, thus, their prediction performance would be acceptable if the output is highly impacted by previous values. However, determining the orders for AR and MA models and the times of initial difference would be a challenge. A detailed summary for applying ARMA and ARIMA models can be found in references listed in Table 1.

2.2.2. Machine learning methods

2.2.2.1.Regression tree (RT)

RT is a type of decision tree with continuous target variables, see Figure 3. An RT starts with a root node where the input data are split into different internal nodes or leaf nodes. For internal nodes, the inputs are continuously split into subsets, while leaf nodes represent the output of the RT model. This implies that there is a chance that the RT make predictions without involving entire feature space.

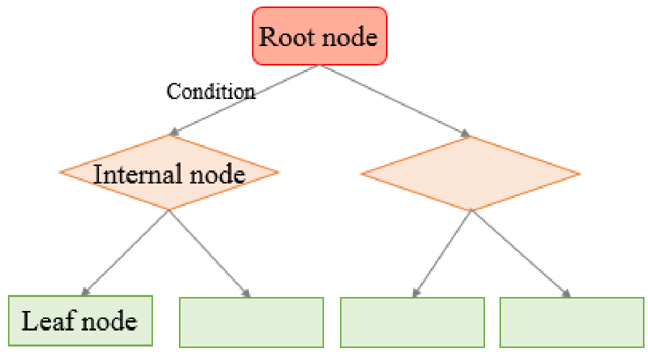


Figure 3 : Schematic of decision tree

The main advantage of RT is easy to understand and interpret due to the fact that it could be displayed graphically [40], [72]. Besides, RT could outperform traditional statistical methods once proper features are selected [73]. The disadvantages of RT are: (1) It could be sensitive to small changes of data; (2) Its structure fails to determine smooth and curvilinear boundaries. Furthermore, to enhance RT prediction performance, groups of RT could be combined as an ensemble model, which would be reviewed in Section 2.2.2.5.

2.2.2.2.Support vector regression (SVR)

SVR is a regression application of SVM, which maximizes the margin between different categories as shown in Figure 4. For SVR, the goal is to find a linear regression function that could predict the result with acceptable deviation from the actual target [74]. For nonlinear regression problems, a kernel function should first be selected to map the original inputs to a high-dimensional feature space, and then apply the SVR. Therefore, one challenge of SVR is the proper selection of kernel function.

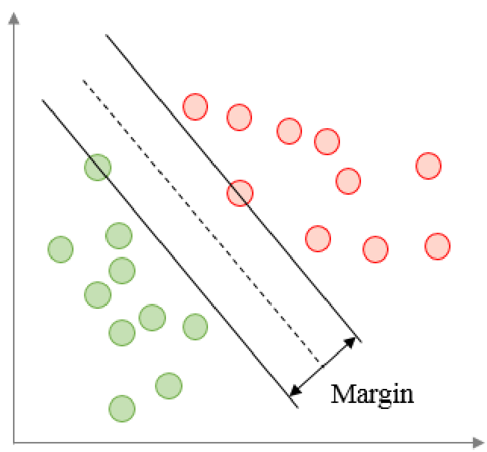


Figure 4 : Schematic of margin between different categories

The advantages of SVR are: (1) It has the ability to solve global minima instead of local minima[75]; (2) Its computational complexity is not determined by the dimensionality of feature space[76]; and (3) Its prediction performance is not sensitive to the noisy data. The application of SVR will not be discussed here since it has been sufficient summarized.

2.2.2.3.Artificial neural network (ANN)

ANN is a machine learning technique inspired by biological neural network [77]. As shown in Figure 5, a typical ANN usually consists of three layers: input layer, hidden layer and output layer. The training goal of an ANN model is to learn the weights and bias (as shown in Equation 7) with proper number of neurons and hidden layers as well as activation functions. Note that although ANN with a single hidden layer can present any Boolean function and ANN with two hidden layers shows the ability to train any function to arbitrary accuracy, the number of hidden layers should be carefully selected to achieve better accuracy with fewer neurons. Furthermore, once the number of hidden layers is increased, the ANN could be considered as deep learning (see Section 2.2.2.4).

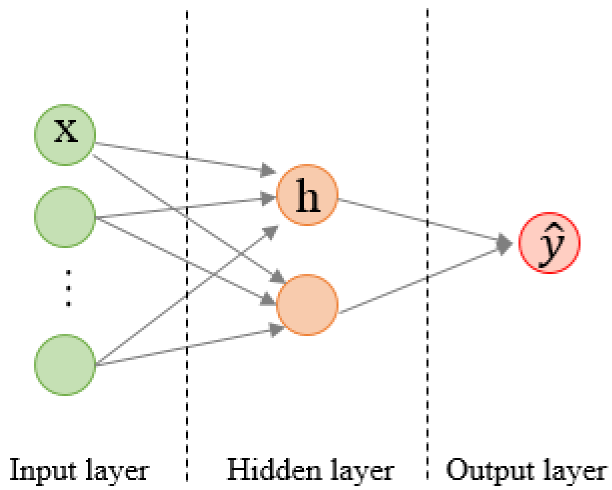


Figure 5 : Schematic of typical ANN

$$\hat{y} = \emptyset(w_{out}h + b_{out}) = \emptyset[w_{out}\sigma(wx + b) + b_{out}] \tag{7}$$

where:

$\emptyset$ is the activation function of output layer;

h is the output of the hidden layer, $h = \sigma(wx + b)$;

σ is the activation function for the hidden layer;

w_{out} and w are the weight matrix;

b_{out} and b are the bias terms.

The advantages and disadvantages of ANN have been described in References [8] and [12]. The main advantage of ANN is the ability to deal with non-linear problems without expertise, while the main disadvantage is the long time required for training models with large number of networks.

2.2.2.4.Deep learning

Deep learning based on ANN includes three categories: deep neural networks (DNN), convolutional neural networks (CNN) and recurrent neural networks (RNN).

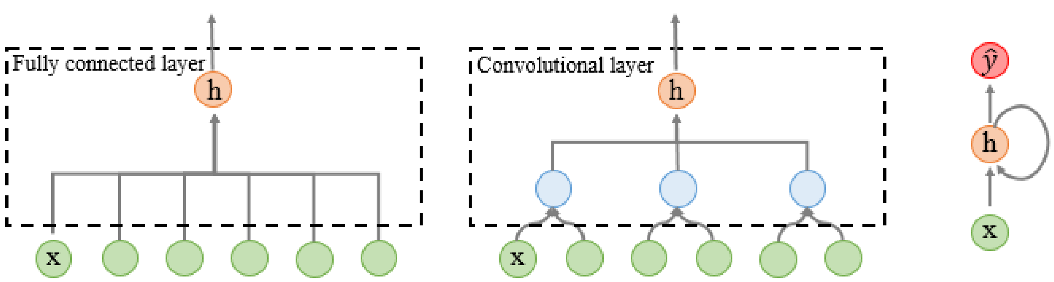
(1) DNN

A DNN is a complex version of ANN containing multiple hidden layers between input and output layers [78]. Typical DNN is a feedforward network without lopping back [79]. Generally, DNN refers to fully connected networks (shown in Figure 6(a)), which means that each neuro in one layer receives information from all neuros from previous layer.

The motivations of utilizing DNN instead of simple ANN have been argued by Good Fellow et al.[80]: (1) DNN requires less neurons than simple ANN in representing complex tasks; (2) In practice, DNN generally presents higher prediction accuracy than ANN. However, implementing DNN models should be done with careful attention to two common issues: overfitting and computing intensive.

(2) CNN

CNN is a special class of DNN, which adopts convolutional layers (shown in Figure 6(b)) to group input unites and apply the same function to gathered groups (i.e. parameter sharing). Compared with general DNN, CNN decreases the risk of overfitting by reducing the connectedness scale and structure complexity. Therefore, CNN could also be treated as a regularized version of typical DNN.



(a) (b) (c)

Figure 6 : Schematic of (a) fully connected layer, and (b) convolutional layer (c) loop in RNN

CNN is well-known in the field of visual imagery analysis, such as image recognition [81], image classification[82], medical image analysis [83] and natural language processing [84]. To implement CNN into load prediction, Sadaei et al. [85] converted hourly load data, hourly temperature data and fuzzified version of load data into multi-channel images, and then fed it to a CNN model. The prediction performance of the developed CNN model was even better than Long Short-Term Memory (LSTM) models, a kind of RNN.

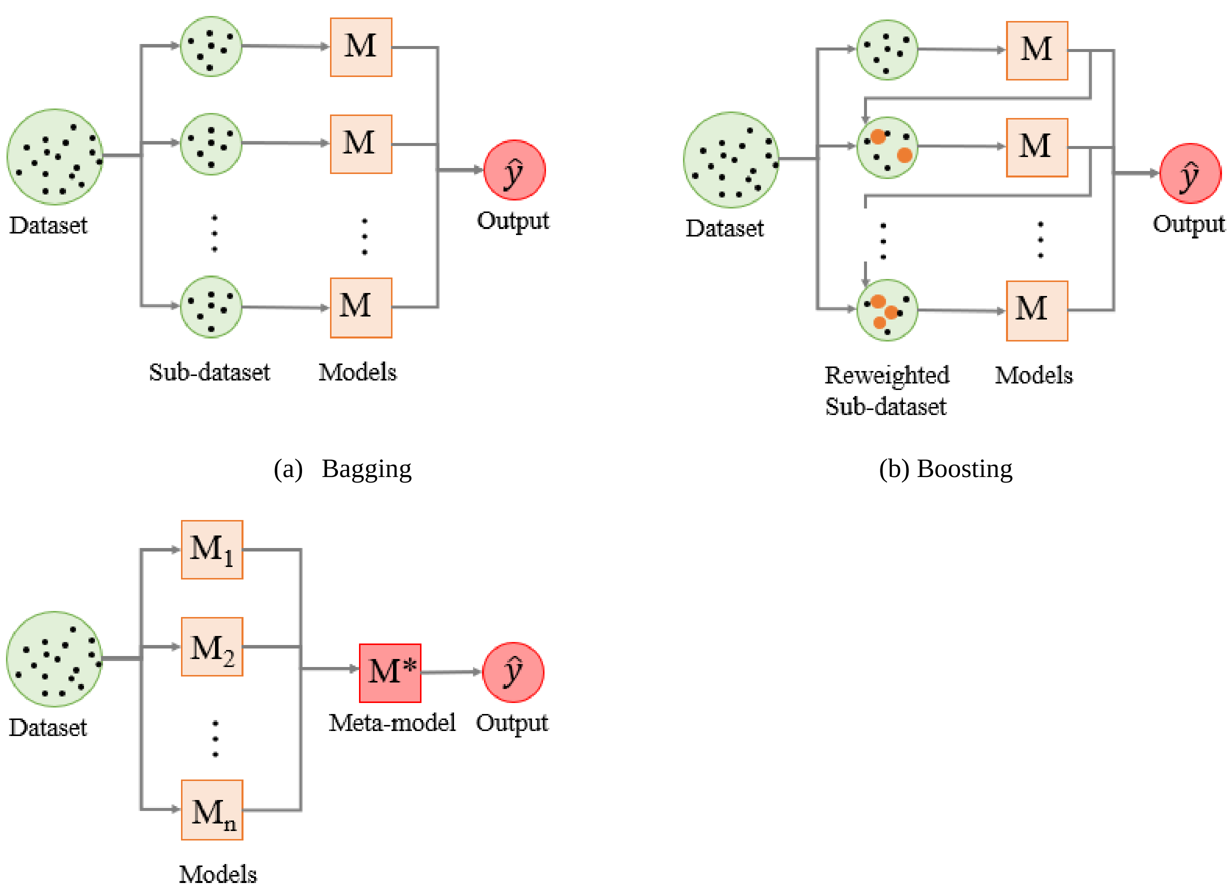
(3) RNN

The distinction between RNNs and other deep learning algorithms is that RNNs involve loops (shown as the cycle in Figure 6(c)) in their structure and makes it possible that information flow in any direction. These cycles introduce time delay in RNN and make RNN more suitable to exhibit temporal dynamic behavior. Therefore, the utilization of RNNs in energy prediction has attracted increasing research interests in recent years.

However, as the weight for the loop is the same for each time step, gradients in the traditional RNN tend to explode or vanish when the loop runs for many times. This problem is called long dependency. To solve this problem, one commonly utilized RNN model, called LSTM, could be applied to remember information for a long period.

2.2.2.5.Ensemble methods

An ensemble method combines the output of multiple learning algorithms in order to enhance the prediction performance of single data-driven models [86]. Commonly used ensemble methods could be classified into three categories: bagging, boosting and stacking models (also called parallel homogeneous, sequential homogeneous and heterogeneous ensemble methods [19]). Schematics of these three types of ensemble methods are shown in Figure 7.



(a) Bagging

(b) Boosting

(c) Stacking

Figure 7 : Schematic of different ensemble methods

(1) Bagging

Bagging, also called bootstrap aggregating, predicts the output by training the same baseline models parallel on different sub-datasets, which are sampled from original input datasets uniformly by replacement. This algorithm tends to decrease the variance when running the trained model on the validation set, due to the independence of each baseline model.

The most commonly utilized bagging method is random forests (RF), for which the baseline models are decision trees. Wang et al. [87] reported that RF is more accurate than RT and SVR in hourly electricity consumption prediction. Furthermore, Johannesen et al. [88] found that RF provides better 30 min-ahead electrical load prediction for urban area compared with kNN and LR. Wang et al. [56] proposed an ensemble bagging tree model to predict hourly educational building electricity demand. Their result shows that the proposed ensemble model is more accurate than RT. However, the larger training time of the bagging tree model than RT would be an issue. Besides, the required additional process for generating sub-dataset and the less interpretable than RT also limits the application of the proposed bagging method.

(2) Boosting

The difference between bagging and boosting is that boosting trains the baseline models incrementally, which means every successive model tries to fix the mistake made by previous models. To achieve this goal, the basic solution is to increase the weight for misclassified data (i.e. orange points in Figure 7(b)). As a result of boosting, the training error would be decreased.

Robinson et al. [89] utilized a gradient boosting regression model to predict annual energy consumption for different types of commercial buildings located in different regions. Their results indicate that the gradient boosting regression model outperforms general linear models (e.g. LR and SVR) and even bagging models with limited number of features. Besides, Walter et al. [90] reported that the gradient boosting decision trees (GBDT) is flexible and accurate for very short term load forecasting for a factory.

Besides comparing the prediction accuracy between different models, interpretability, robustness and efficiency of different models should also be studied. Wang et al. [52] compared these four aspects of five models (i.e. extreme gradient boosting (XGB), GBDT, RF, ANN and SVM) based on a case study of 2-hour ahead heating load prediction for a residential quarter. They concluded that there is no best model when considering all performance. For instance, RF shows the highest accuracy, interpretability and robustness, while XGB presents better efficiency.

(3) Stacking

Unlike bagging and boosting, which utilize the same baseline models, stacking works on an arbitrary set of models. As shown in Figure 7(c), different models are trained on the available input dataset, and then a meta-model is trained based on the outputs of these models to make the final prediction.

Huang et al. [55] combined XGB, extreme learning machine (ELM), LR and SVR as an ensemble learning method, and then utilized it to do a 2-hour ahead heating load prediction for a ground source heat pump that supplies space heating for a residential area. Their result shows the proposed ensemble model is more accurate than XGB, ELM, LR and SVR. Fan et al. [40] developed an ensemble model integrated by eight learning algorithms to enhance the prediction accuracy for next day energy consumption and peak power demand.

2.3. Outputs

In this section, research objects of outputs, i.e. the type of buildings, the types of energy, the scale of buildings, the length and number of steps expected for prediction, and the criteria of evaluation for the accuracy of developed date-driven models, etc. will be presented.

2.3.1. Building Type

When evaluating building loads, the type of buildings (i.e. residential or non-residential) should be distinguished, because the percentage of end use and the influence factors for energy consumption would be different for different types of buildings. For instance, the load consumed by cooking could be a huge contribution for peak load in residential buildings, while official equipment would consume a considerable percentage of commercial building loads. Besides, the set-point temperature for air conditioning is generally controllable for occupants in residential buildings, while it shows a large chance of being constant for nonresidential buildings.

Note that non-residential buildings further include commercial buildings, educational buildings, industrial buildings and hotels.

2.3.2. Energy type

The predicted energy could be separated into electricity, natural gas, fuel oil, and steam in terms of energy source, while it can be divided as air conditioning (space heating and cooling), domestic water heating, plug-load and lighting in terms of end-use [91], [92]. Besides, the forecasted energy could be classified as energy consumption and power demand. Energy consumption is the amount of energy consumed during a period of time, while power demand means how fast energy needs to be supplied. Thus, energy consumption is the integral of power demand over time. For a given time interval, if power demand is constant, its prediction within the given resolution would be consistent with energy consumption forecasting. On the other hand, when power demand fluctuates among the given time interval, energy consumption prediction and power demand forecasting should be distinguished.

The benefits of distinguishing the type of predicted energy are:

(1) Making it possible to quantitatively evaluate the environmental impact (e.g. global warming, ozone layer depletion, human toxicity, and photochemical oxidation, etc.) of building energy use and then give basis to take measures in order to reduce the environmental impact [93].
(2) Providing a foundation in feature construction. For instance, meteorological information might be the critical factor for heating/cooling demand prediction, while occupancy related data and time index could be the most promising features for predicting energy consumed by lighting.
(3) Offering opportunities for more targeted energy/cost saving methods. Through developing energy prediction models for a specific end-use, unique operation/control

plans could be formulated to achieve the minimum energy consumption or cost during a given period.

2.3.3. Scale

Scale can be classified into four classes: sub-building, building, district, region (city), country. Note that a sub-building refers to an individual room or component in a building.

Energy prediction for larger scale (e.g. region and country) should not be considered as a simple aggregation of smaller scale (e.g. sub-building and building), since the effective and available features for different scales would vary [10]. For instance, socio-economic information tends to be collectable and useful for large scale energy prediction, while its effect declines in predicting building/sub-building level energy consumption. Besides, the application of energy prediction models in reality varies for different scales. The model developed for sub-building and building scale could be utilized for demand response control, while large scale energy prediction model is applicable in energy distribution.

2.3.4. Temporal granularity

Two types of temporal granularity need to be determined: horizon and resolution. Horizon means the length of time-ahead load prediction, while resolution means the duration of a time step. When horizon is longer than resolution, the developed model makes an n-step ahead prediction, where n equals to horizon divided by resolution. For instance, if the prediction horizon is 1 hour and the resolution of data points is 15 min, then the model does a 4-step ahead prediction. One thing to note is that under the same resolution, longer horizon prediction takes more risk for higher error. For instance, Ding et al. [94] utilized historical data and meteorological parameters to predict one hour ahead and one day ahead cooling load with 30 min intervals. Their results indicate that a shorter horizon (i.e. 1 hour) prediction presents higher accuracy than a one day ahead prediction.

Time horizon of electrical load prediction models is usually classified as four categories: very short term, short term, medium term and long term [10], [9]. However, the cut-off horizon for these categories varies among different references. Generally, when the time horizon is less than one month, it belongs to short term prediction or even very short term prediction. Very short term and short term predictions help users to implement proper control strategy for load shifting and benefits utilities to design energy distribution plans, while medium and long term prediction could be beneficial for utilities to upgrade their equipment and for governments to formulate standards for energy saving and modify plans for the electricity market.

2.3.5. Criteria

Commonly used validation criteria for evaluating the performance of prediction models include [63], [95]–[98]:

Mean Absolute Error (MAE) = $\frac{1}{n}\sum_{i=1}^{n}|\hat{y_i} - y_i|$ (8)

Mean Absolute Percentage Error (MAPE) (%) = $\frac{1}{n}\sum_{i=1}^{n}|\frac{\hat{y_i}-y_i}{y_i}| * 100$ (9)

Mean Bias Error (MBE) = $\frac{\sum_{i=1}^{n}(\hat{y_i}-y_i)}{n}$ (10)

Normalized MBE (NMBE) (%) = $\frac{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}}{\overline{y}} * 100$ (11)

Mean Squared Error (MSE) = $\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$ (12)

Root Mean Square Error (RMSE) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$ (13)

Coefficient of Variation of the Root Mean Square Error (CV(RMSE)) = $\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\overline{y}}$ (14)

R Square (R^2) = $1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}$ (15)

where $\overline{y}$ is the average value of measured outputs.

These criteria could also be utilized to evaluate the prediction accuracy during model training. Through comparing the prediction accuracy between model training and model validation, overfitting or underfitting could be detected [99]. For instance, if training accuracy is much higher than validation, it might indicate overfitting which means the trained data-driven model fits too closely to the training set with covering the noise and outlier. Besides, if both training and validation accuracy are not acceptable, underfitting occurs to show that the developed data-driven model cannot capture the structure of the studied problem. Both overfitting and underfitting undermine the developed models' generalization, which refers to the ability to predict unseen data.

Here, a short description is given in the following to help the criteria selection.

MAE is the mean value of the sum of absolute errors, while MBE is the average prediction error which could be understood as how far the average predicted values is above or below the average of measured output value. Both MAE and MBE have units that should be taken into consideration when utilizing them to compare the results of different works. Note that the under-predicted outputs would reduce the value of MBE, which means cancellation errors. Therefore, if choosing MBE, other criteria without cancellation errors should be considered.

MAPE is a commonly utilized measure of prediction accuracy because it calculates the mean relative prediction error without units. However, it cannot be utilized when there are zero values in the measured output. By contrast, zero values would not be a big concern when utilizing NMBE, which also shows the advantages of having no units. However, NMBE is limited by cancellation errors.

MSE has the ability to evaluate both variance and bias of the predicted value to the measured output. Note that the unit of MSE would be square of the unit of predicted outputs. To have the same unit as the predicted outputs, RMSE could be utilized. In terms of principle, CV(RMSE) is calculated by dividing RMSE by the mean value of measured outputs; therefore, it evaluates how much the predicted error varies with respect to the mean target value. It is not limited by cancellation errors. Furthermore, NMBE and CV(RMSE) have been recommended as evaluation criteria for building energy prediction models by several standards, such as ASHRAE [63], FEMP [100] and IPMVP [101].

R^2 indicates the goodness of fit. The bigger the value of R^2, the closer its predicted value will be to its target value.

2.4. Discussion

To better understand recent development of energy series prediction by data-driven approaches, papers from 2015 to 2019 are reviewed a total of 105. These literature are found by following steps: (1) Search keywords "data driven; building" or "machine learning; building" from Scopus [102]; (2) Quickly review title of the searched papers, and remain papers which work on energy prediction; (3) Review in depth and keep the relevant 105 papers.

Among the reviewed 105 papers, 17% of these studies are based on public datasets, such as [103]–[110], etc. To the best of the authors knowledge, studies based on the same dataset are lacking. Even if utilize the same dataset, these studies are not compared to each other. Besides, most of existing studies utilize private datasets which are not published due to some reasons, such as privacy and ethics issues. It makes it further difficult for other researchers to reproduce and improve the existing studies. Therefore, as more public data available in the future, quantitatively comparison of new techniques to the existing studies would be effective to improve the usability of data-driven models in building energy prediction. Besides, the number of data points utilized by model training and validation, as well as the number of meters and buildings utilized for data collection are recorded for each reviewed paper. However, these aspects are not analyzed in this paper, because the amount and quality of utilized data are affected by many factors and varies case by case.

Furthermore, the distribution of the reviewed studies among the feature selection, model utilization and prediction objective (output) are summarized in the following subsections.

2.4.1. Study distribution based on features

The utilization of different types of features in studies from year 2015 to 2019 is summarized in Figure 8. Meteorological information, historical data and time index are the top-3 important factors for building energy prediction. Indoor environment information is not commonly used because air-conditioned buildings (especially nonresidential buildings) usually have a nearly constant indoor condition. However, the dynamic change of indoor conditions should be considered for energy prediction to achieve peak shifting by controlling the indoor environment within an acceptable range. The relatively lower utilization of occupancy related data is caused by the complexity of data collection and its replacement by time index data. Building characteristic data is usually ignored because it is generally kept as constant through the life cycle of a specific building. One possible reason for less utilization of socio-economic information among recent studies is that it is only useful for large scale prediction. This conclusion can be deduced from Figure 9. Another interesting observation from Figure 9 is that indoor environment information, occupancy related data and building characteristic are just utilized by relatively small scale (i.e. sub-building, building and district level).

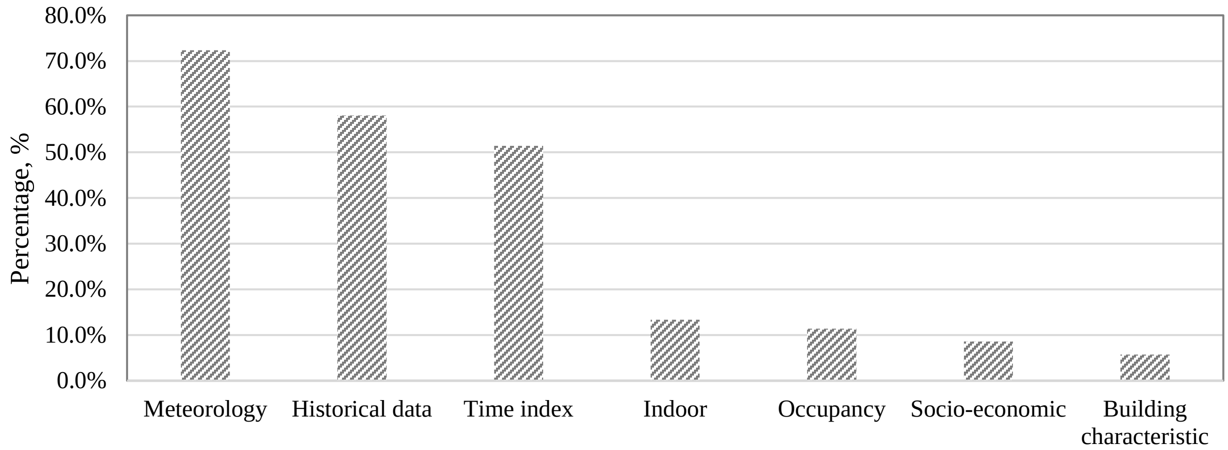


Figure 8 : Feature utilization in recent studies

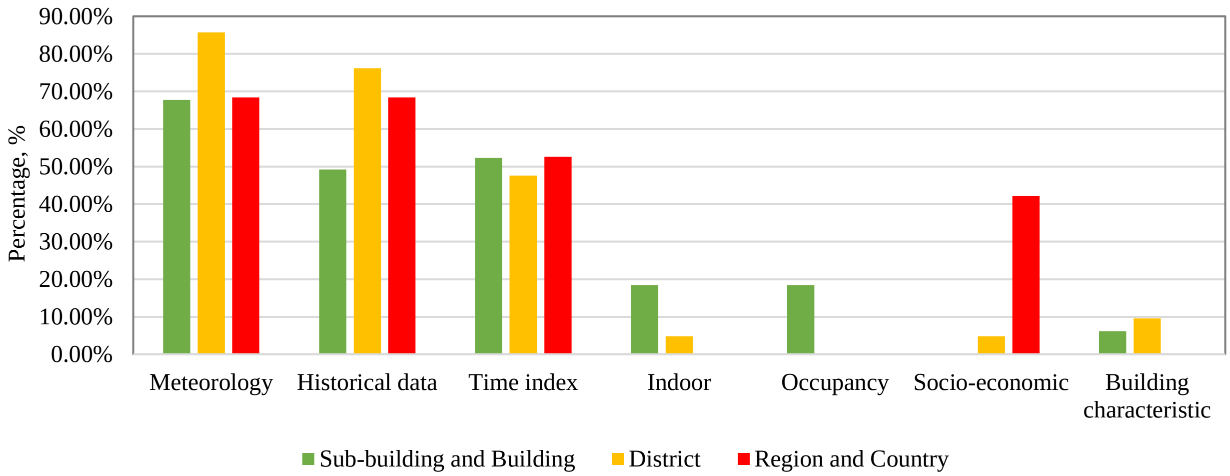


Figure 9 : Feature utilization among different scales in recent studies

2.4.2. Study distribution based on models

The percentage of studies utilizing different kinds of data-driven models are summarized in Figure 10. ANN, SVR and LR seem to be the popular models, while the concentration on time series analysis and RT is less. The application of RT is less due to its unacceptable prediction accuracy when applied to validation dataset or test dataset. However, RT is a common base model in ensemble methods, which have attracted considerable attention in recent years. Besides this, as Figure 10 shows, deep learning has started to draw interest in recent years. Moreover, around 80% of studies implement more than one model to the collected data, because the generalization of data-driven models varies among different factors, e.g. size and structure of dataset.

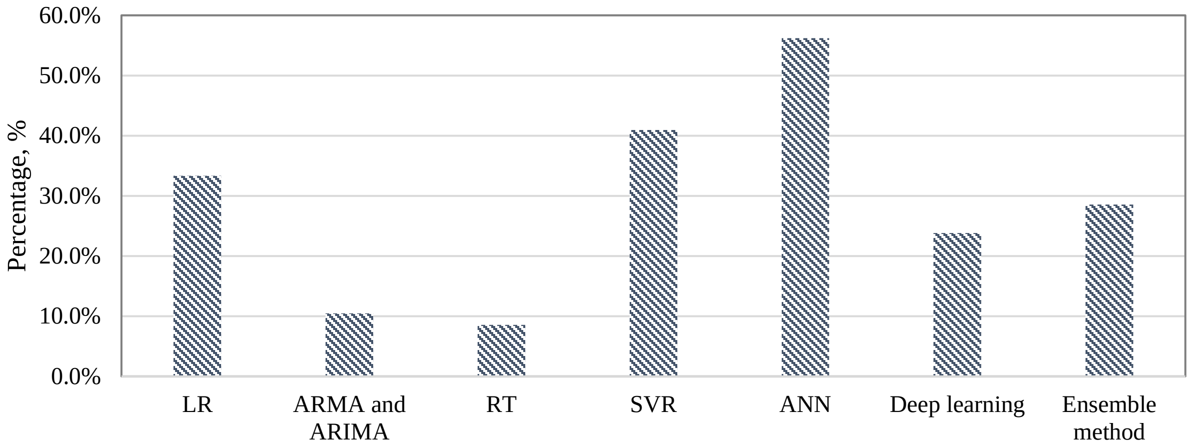


Figure 10 : Model utilization among studies

2.4.3. Study distribution based on outputs

Distribution of research among building types is shown in Figure 11. Earlier studies have mainly concentrated on residential, commercial and educational buildings, while studies based on industrial buildings and hotels are lacking. The reason for few studies focusing on industrial buildings is that the production from different factories varies a lot and thus the influencing factors cannot be easily realized and selected. Besides, for hotels, occupant numbers and occupancy activity are unstable, thus, energy prediction for hotels based on data-driven approaches could be challenging.

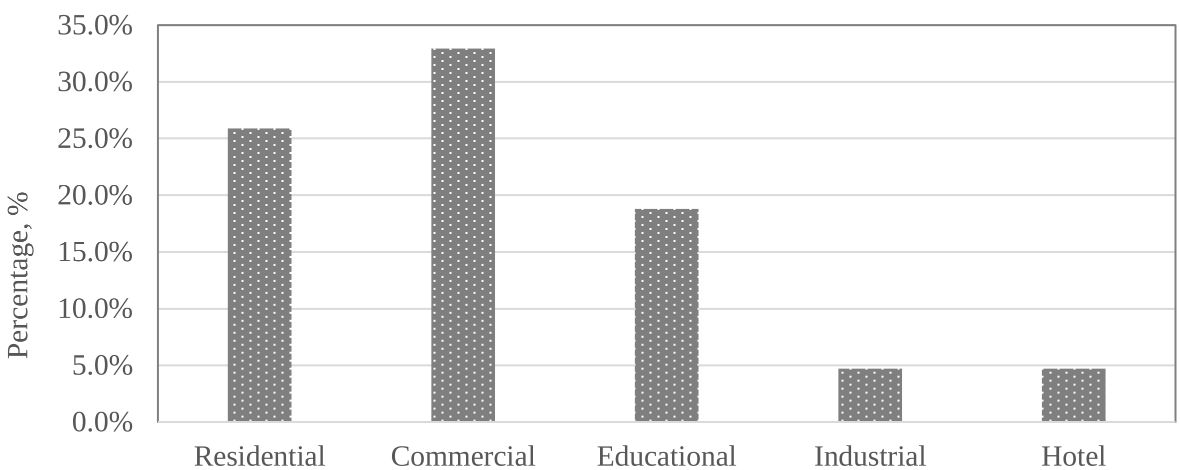


Figure 11: Study distribution by building type

Research distribution by energy type is not summarized here because most studies predicted aggregated end-use energy consumption instead of individual end-use and the effect of primary energy type on model development process is negligible.

Figure 12 presents the distribution of studies for different scales. More than 60% of studies predict the energy for an entire building, followed by around 20% of studies for district level. The high number of studies on these two scales is caused by more collectable data and applicability of the developed model to realistic demand response control and grid distribution. However, prediction for sub-building level is lacking due to the limitation of data collection. For instance, Geyer et al. [111] utilized simulation data instead of real measured data to predict heat flux through individual components, such as walls, windows, and roofs. The research focusing on sub-building level energy prediction would increase as the demand for individual room control and for energy saving potential analysis of different envelopes goes higher.

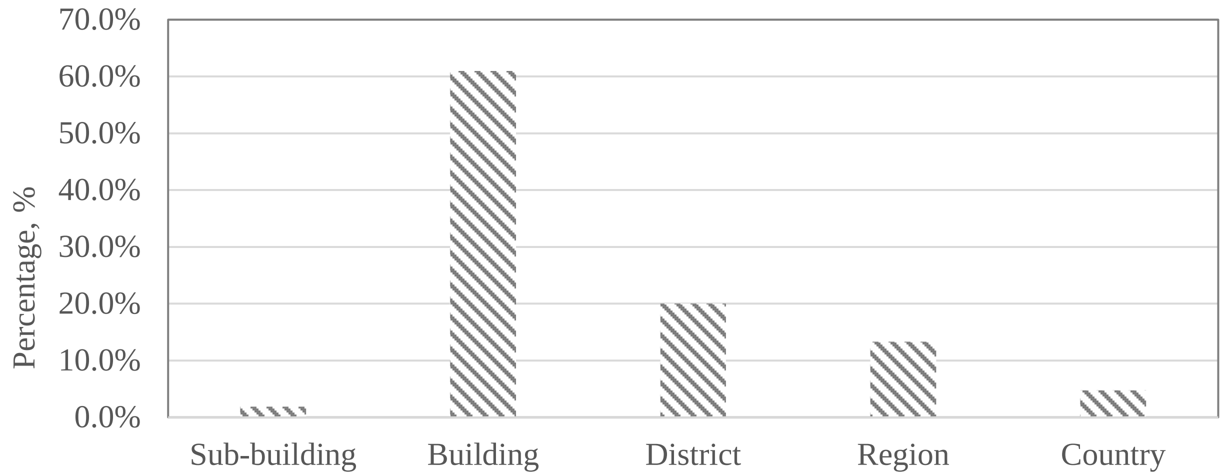


Figure 12 : Study distribution by scale

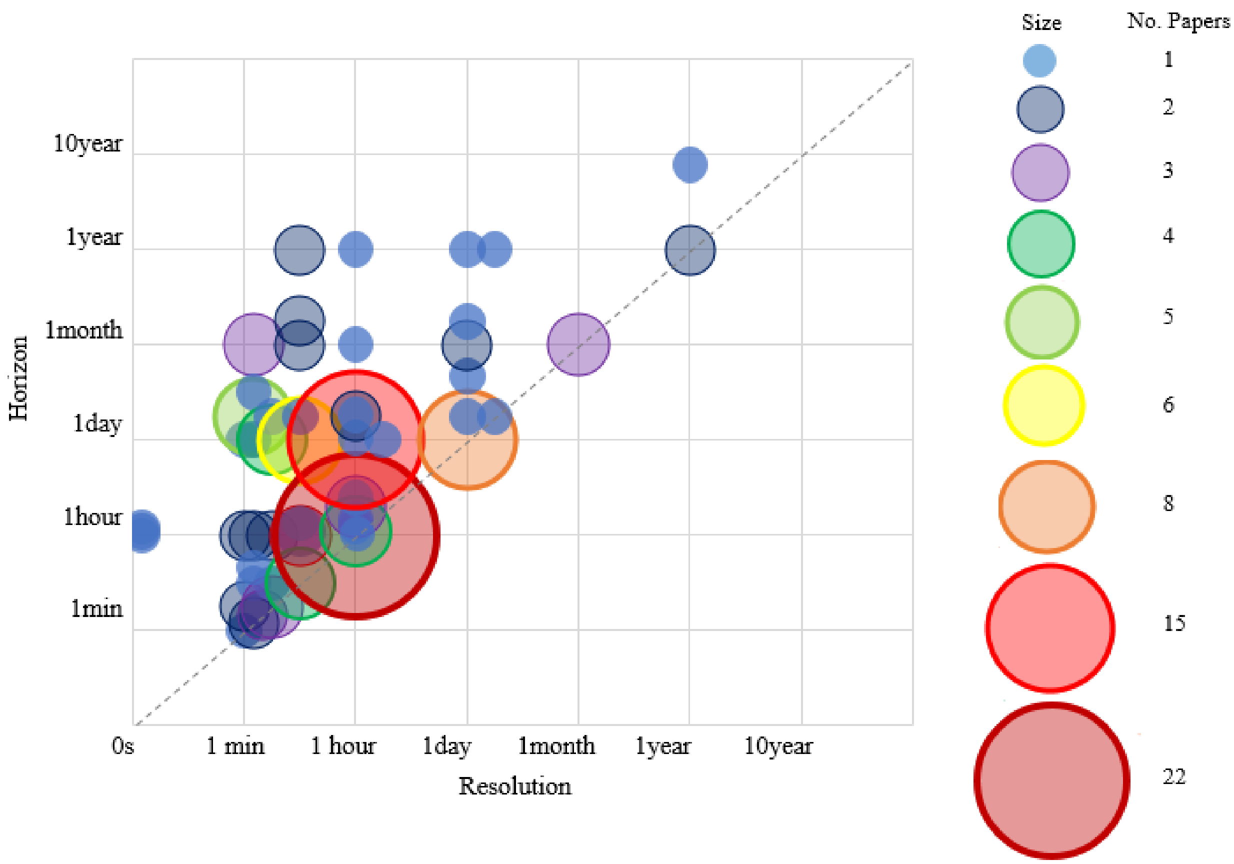


Figure 13 : Study distribution by temporal granularity

Horizon and resolution of the 105 studies are shown in Figure 13. The centroid of each circle means the resolution and horizon of studies; the bigger the circle, the more studies are located in that temporal granularity. Circles lies on the dash line are single step predictions, meaning horizon is equal to resolution. All circles above the dash line are multi-step predictions. Note that one paper could present results for several temporal granularities. Most (64.75%) studies present multi-step prediction, which is useful for continuous control and monitoring. Besides, the resolution and horizon for most studies are higher than 1 min.

3. Updating strategies for multi-step building energy prediction

After training and validating a data-driven model, the developed model could be utilized for energy prediction in real life. Prior to implementation, the process of updating prediction inputs should be resolved. Therefore, this section summarizes several updating strategies for multi-step building energy prediction.

3.1. Strategy 1: Updating inputs by real values without historical data

In this strategy, energy consumption during a specific period or power demand at a specific time is predicted only based on exogenous inputs (e.g. forecasted meteorological information, time index, etc.) during that specific time period or time point. Studies that did multi-step building energy prediction based on Strategy 1 are summarized in Table 3.

Table 3 : Studies for multistep energy prediction based on Strategy 1

Reference	No. training points	No. validation points	No. meters	No. buildings	Features	Models	Building type	Energy type	Scale	Resolution	Horizon	No. steps	Criteria
[112]	3562	1526	-	1	Meteorological information, time index	LR, SVR, RF	Educational	Electricity energy consumption	Building	-	-	1-3	MAE, RMSE
[113]	-	-	-	-	Meteorological information, time index, socio-economic information, operation characteristic of HVAC system	DNN	Commercial, industrial parks	Heating and cooling energy consumption	District	1hour	1hour-3hour	1-3	MSE
[60]	2784	96	6	6	Meteorological information, time index	ANN, SVR, LR, Gaussian process	Commercial, Hotel	Electricity energy consumption	Building	15min	24hour	96	RMSE, NMBE, CV(RMSE), R^2
[33]	12240	2880	-	8	Meteorological information, indoor environmental information, time index, operation characteristic of HVAC system	RT, RF	Educational	Electricity power demand	Building	5min	1hour	12	-
[114]	1440; 4368; 17520	480; 1488; 5760	-	-	Meteorological information	ANN, LR, ensemble method (AdaBoost)	ALL	Energy consumption	District	30min	1month, 1season, 1year	1440, 4320, 17520	MAE, MAPE, CV(RMSE)
[115]	21924	3132	-	1	Meteorological information, time index	ANN	Educational	Electricity power demand	Building	1hour	24hour, 9days	24, 216	MAE, MAPE, RMSE

Since no historical data would be utilized as input for prediction, the multi-step ahead prediction performance is mainly limited by the forecasting uncertainty of exogenous inputs. Another drawback of this strategy is its inability to treat the expected output as time series data. Therefore, it is usually utilized in three cases: (1) Doing single-step prediction [56], [116]; (2) Doing multi-step prediction with longer horizon (usually higher than 1 day) and larger scale (generally greater than district). For instance, Salcedo-Sanz et al. [117] and Sánchez-Oro et al. [118] presented acceptable one-year ahead energy demand prediction for Spain based on socio-economic factors for the corresponding year; (3) Furthermore, for predicting energy for appliances whose operation schedule is not affected by previous conditions, such as lighting. For instance, Amasyali and El-Gohary [119] predicted daily lighting energy consumption using SVM based on daily average sky cover and day type.

3.2. Strategy 2: Updating inputs by real values with historical data

In this strategy, historical data is one of the input factors for energy prediction. However, the historical data is not treated as continuous information, which means energy consumption in the next step might be affected by data in discontinuous previous steps. Therefore, ground truth of historical data is utilized as input during multi-step energy prediction.

Guo et al. [120], [121] observed the thermal response time of the building as 40 min, and thus utilized current heating load, meteorological information, indoor temperature, and time index to predict the heating load after 40 min by SVM, MLR and ANN. Note that their study only focused on single-step prediction. Dedinec et al. [122] employed the average load of the previous day, the load for the same hour of the previous day and the load for the same hour-day combination of previous week as historical data to forecast day ahead hourly electricity load for buildings in Macedonia. These discontinuous historical data ensure the availability of ground truth data. Similarly, Idowu et al. [123] utilized data at time t-k and t-$2k$ to predict heating load at t+k (where t means current time and k is the forecast horizon). The drawback of this study is the inconsistent lag time for different time steps in a specific horizon, which means a larger lag period for a further future point. More studies doing multi-step energy prediction based on Strategy 2 can be found in Table 4.

The main limitation of Strategy 2 is that the intermittent historical data could weaken the energy prediction accuracy, because recent historical data has more effect on future energy consumption than before.

Table 4 : Studies for multistep energy prediction based on Strategy 2

Reference	**No. training points**	**No. validation points**	**No. meters**	**No. buildings**	**Features**	**Models**	**Building type**	**Energy type**	**Scale**	**Resolu tion**	**Horizon**	**No. steps**	**Criteria**
[94]	768	768	4	1	Meteorological information, historical data (previous hour's cooling load/one day prior cooling load)	SVR	Commercial	Cooling load	Building	30min	1hour, 24hour	2, 48	MAE, RMSE, R^2
[29]	144	24	-	1	Historical Meteorological data	MLP, SVR	Commercial	Heating load	Building	1hour	24hour	24	MRE, CV(RMSE), R^2
[124], [125]	2400; 696	48; 24	2	2	Meteorological information, historical data	SVR	Commercial, Hotel	Electricity power demand intensity	building	1hour	24hour	24	MAE, MAPE, RMSE
[126]	35064	17544	-	-	Meteorological information, time index, historical data (previous 24 hr average load, 24-hr lagged load, and 168-hr lagged load)	ANN-IEAMCG M-R	ALL	Electricity energy consumption	Region	30min, 1hour	24hour	48, 24	MAE, MAPE
[127]	14400	5040	-	10	Meteorological information, time index, historical data (thermal load and control signal with a lag of 1 day and 1 week)	LR, ANN, SVR, RT	Commercial	Heating load	District	1hour	24hour	24	MAPE
[123]	936	168	10	10	Meteorological information, historical data (Meteorological, thermal load and operational variables)	MLR, ANN, SVR, RT	Commercial, residential	Heating load	Building	1hour	1hour-48hour	1-48	RMSE
[122]	43800	17520	-	-	Meteorological information, time index, historical data (average load of the previous day, load for the same hour of the previous day, and load for the same hour-day combination of previous week)	Deep belief network	ALL	Electricity power demand	Country	1hour	24hour	24	MAPE
[128]	-	-	-	-	Meteorological information, time index, historical data (previous day hourly load, previous week hourly load),	Ensemble methods	-	Electric power demand	-	1hour	24hour	24	-

Reference	No. training points	No. validation points	No. meters	No. buildings	Features	Models	Building type	Energy type	Scale	Resolu tion	Horizon	No. steps	Criteria
					1st and 2nd derivatives of previous load and current temperature								
[129]	17520	8760	-	-	Time index, historical data (previous two days' consumption, previous day exterior temperature)	ANN	ALL	Energy consumption	Region	1hour	24hour	24	MAPE
[130]	14892	2628	-	5	Meteorological information, time index, historical data (energy consumption at the same timestep the previous day, and energy consumption at the same timestep the previous week)	ANN	ALL	Heating energy consumption	Building	1hour	24hour	24	R^2

3.3. Strategy 3: Updating inputs by predicted data

In this strategy, multistep energy prediction with historical data as one of input features is achieved by updating predicted value as historical data for upcoming steps, as shown in Figure 14. Note that the moving box in Figure 14 could also be called as sliding window method with length n.

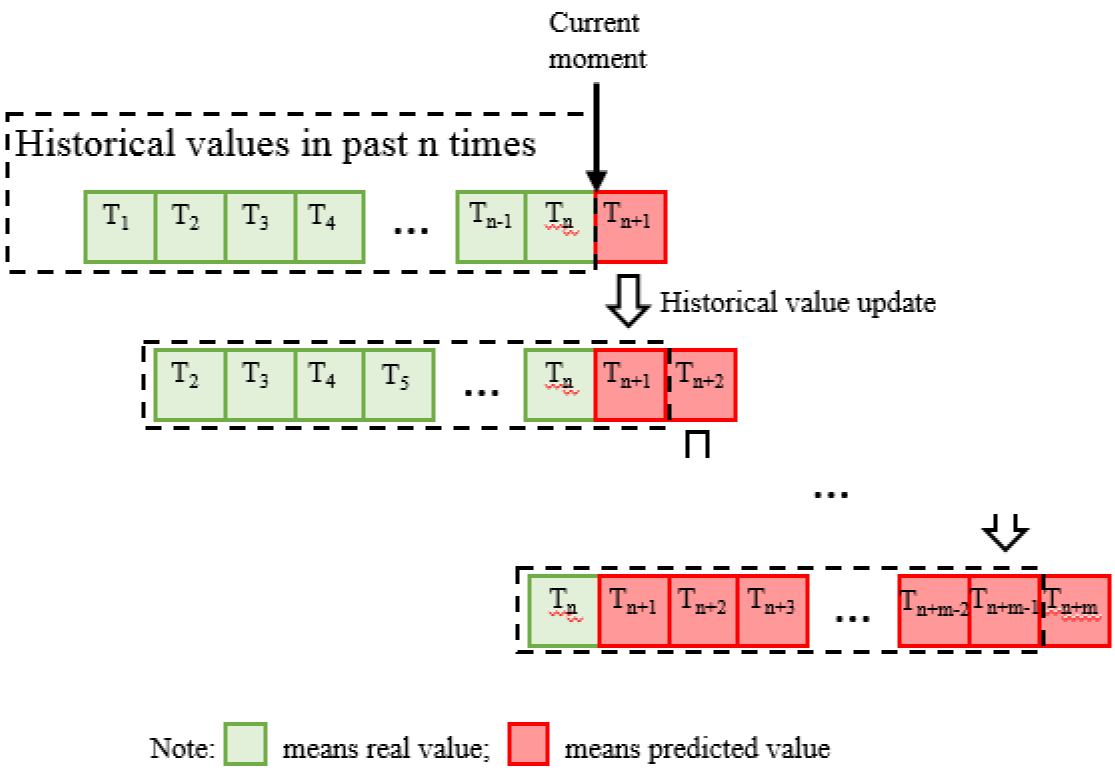


Figure 14 : Updating process of Strategy 3 for m-step ahead energy prediction with n steps of historical data

Deb et al. [131] employed Strategy 3 to predict next 20 days' daily energy consumption using the previous 5 days' cooling load. Their result presents high accuracy (R^2 more than 0.94). Note that the inputs and outputs in their study are all summarized classes instead of continuous values. Fan et al. [132] applied the updating process shown in Figure 14 for 24 hour ahead cooling load prediction based on data with 30 min intervals. Their study shows that involving past 24 hours' data as input features could significantly increase the prediction accuracy compared with only considering time index and meteorological information at the predicted time. Furthermore, the study of Mocanu et al. [133] illustrates that under Strategy 3, one-day ahead energy prediction is worse than one hour ahead prediction with 1-minute resolution data. Summary for more studies with multi-step energy prediction by Strategy 3 are shown in Table 5

Table 5 : Studies for multistep energy prediction based on Strategy 3

Reference	No. training points	No. validation points	No. meters	No. buildings	Features	Models	Building type	Energy type	Scale	Resolution	Horizon	No. steps	Criteria
[24]	3942	216	-	3	Meteorological information, historical data	ARIMA, RNN	Commercial	Electricity energy consumption	Building	1hour	24hour	24	MAPE, CV(RMSE)
[131]	237	158	-	3	Historical data (previous five days' energy consumption)	ANN	Educational	Energy consumption for cooling	Building	1day	20day	20	R^2
[134]	-	-	-	-	Meteorological information, historical data	LR, Gaussian process, ANN	Commercial	Cooling load of water source heat pump	Building	5min	7days, 1month	2016, 8640	MAE, CV(RMSE), MAPE
[132]	11054	2369	-	1	Meteorological information, time index, historical data (previous 24hours' cooling load, exterior temperature and humidity)	MLR, RF, GBM, SVR, XGB, DNN, elastic net	Educational	Cooling load	Building	30min	24hour	48	MAE, RMSE, CV(RMSE)
[135]	10526	1858	-	-	Historical data	ANN, SVR, ensemble method	Commercial	Energy consumption of heat pump system	Building	5min	30min	6	RMSE, MAE, R^2
[136]	2688	96	12	1	Meteorological information, time index, historical data	ARIMA, SVR, MetaFA-LSSVR, SARIMA-MetaFA-LSSVR	Residential	Electricity energy consumption	Building	15min	24hour	96	MAE, MAPE, RMSE
[137]	2688	96	12	1	Meteorological information, time index, historical data	SARIMA-MetaFA-LSSVR	Residential	Electricity energy consumption	Building	15min	24hour	96	Error rate
[138]	40320	2976	19	1 factory with four sections	Meteorological information, time index, historical data (lag of 1,2,3,24,168)	MLR, SVR	Industrial	Reactive power demand	Building	1hour	1hour - 24hours	1-48	Normalized MAE, Normalized RMSE, MASE
[139]	-	-	-	30	Meteorological information, time index; historical data	ANN	Educational	Thermal load	District	15min	24hour	96	NMBE, CV(RMSE)

One drawback of converting predicted value as inputs for multi-step energy prediction is the accumulated prediction error [140], [141]. To avoid this issue, two general solutions could be utilized. The first solution is Strategy 2, which has a longer lagging time than prediction horizon. The second solution is to aggregate training data and reduce the resolution. For example, Gu et al. [142] calculated the daily averaged value from 10 min interval data for daily heating load prediction, while utilizing hourly collected data for hourly prediction. Grolinger et al. [143] found that energy consumption prediction for event venues with 1 day interval data is more accurate than with 15 min or 1 hour interval data. They also presented that the processing time for models with large intervals is much faster than for those with smaller ones. One thing to note when applying this aggregating approach is that the reduced resolution would reflect the flexibility of control strategies developed based on energy prediction model.

Furthermore, to ease the influence of continuous historical data, Chen et al.[124], [125] proposed a clustering based training method for machine learning models when doing 1 day ahead hourly energy consumption prediction. In their study, training data was firstly separated into different groups based on the similarity of electric demand intensity at previous hours and then models for corresponding groups were trained based on meteorological data. Zhang et al. [144] proposed an error correction approach to improve the cooling load prediction accuracy. The error correction terms are calculated by mean difference between real cooling load and predicted cooling load. The proposed approach could significantly increase the 1-hour-ahead prediction accuracy of MLR. However, this approach did not consider accumulated error, thus its effect on multistep ahead prediction is still uncertain.

3.4. Strategy 4: Exploiting time series nature of energy data directly through models

In this strategy, the time series nature of energy related data is considered directly through models, which means the developed models show the ability to predict a sequence of output just based on ground truth data. Schematic for this type of model is shown in Figure 15. Note that models shown in this strategy could also be utilized as a solution for the accumulated error caused by Strategy 3.

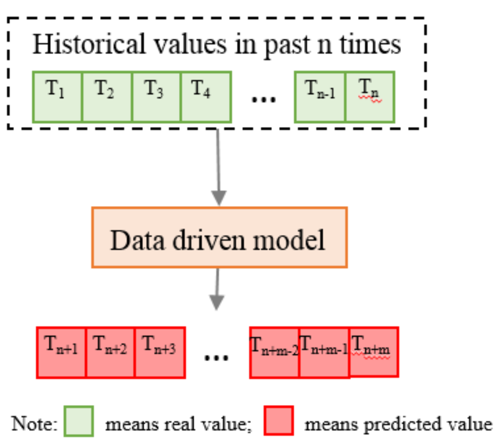


Figure 15 : Schematic for models with a sequence of output

Table 6 : Studies for multistep energy prediction based on Strategy 4

Reference	**No. training points**	**No. validation points**	**No. meters**	**No. buildings**	**Features**	**Models**	**Building type**	**Energy type**	**Scale**	**Resolution**	**Horizon**	**No. steps**	**Criteria**
[24]	3942	219	-	3	Meteorological information, historical data	CNN	Commercial	Electricity energy consumption	Building	1hour	24hour	24	MAPE, CV(RMSE)
[145]	504; 840	24	-	2	Historical data	LSTM	Commercial	Miscellaneous electric loads, lighting, heat gain	Building	1hour	8hour, 24hour	8, 24	Relative RMSE
[146]	8760	1820; 8760	-	31	Meteorological information, time index, historical data	RNN	Commercial, residential	Electricity energy consumption	Building, district	1hour	24hour	24	RMSE
[147]	1844352	204928	3	1	Time index, historical data, sub-metering data	CNN-LSTM	Residential	Electricity energy consumption	Building	1min	1hour	60	MAE, MAPE, MSE, RMSE
[148]	75000	2538	-	100 users	Meteorological information, historical data (previous 4 weeks of a load profile sequence)	LSTM	-	Electricity energy consumption	Building	1day	4weeks	28	Accuracy
[149]	-	-	-	-	Historical data (previous 4 hours' load profile sequence)	LSTM	ALL	Electricity energy consumption. power demand	Region	15min	30min	2	MAPE, RMSE
[150]	-	-	63 .	1	Historical data (previous 240 hours' load or previous 70 days' load)	LSTM	Residential	Electric power demand	Building	1hour 1day	24hour 7day	24 7	MRE, MAE, RMSE, SMAPE
[151]	22000	2000	-	200	Time index, historical data	LSTM	Residential	Electric power demand	Building	30min	1hour 2hour 4hour	2 4 8	Pinball loss
[152]	19104; 18096; 18624	6336; 6048; 6192	-	3	Meteorological information, time index, historical data (previous 3days' load)	Recurrent inception convolution neural network (RICNN)	ALL	Electric energy consumption	Region	30min	24hour	48	RMSE, MAPE
[153]	3592	634	-	1	Meteorological information, historical data	LSTM	Residential	Heating demand	Building	1hour	12hour 24hour	12 24	-

Studies doing multi-step energy prediction based on Strategy 4 is summarized in Table 6. It shows that the most commonly used model for Strategy 4 is RNN. For instance, Rahman et al.[146] developed two RNN models that could predict 24-h sequence of electric load with 1 hour intervals. The proposed RNN models provided more accurate electricity prediction than a three-layer DNN, while DNN is more reliable for a year ahead hourly aggregated energy prediction for a group of residential buildings. The poor performance of RNN in predicting aggregated energy prediction is caused by the fewer long-term dependencies in the aggregated profile. Moreover, the training time of these RNN models limited their application.

4. Conclusions and opportunities for future works

In this paper, studies working on building energy prediction with data-driven approaches are reviewed. To clarify shortages of existing review papers, a revision is done for previous literature reviews focusing on similar subject as this paper. Incomplete work of previous review papers is present in all steps of developing a data-driven model. Besides, no review summarizes how to implement the developed data-driven models in multi-step energy prediction.

In order to fill research gaps in previous review studies, this paper first gives a comprehensive review for developing data-driven models in terms of general procedures, which contains feature engineering, data-driven algorithms and factors reflected from outputs. Prior to this, data collection and data cleaning are briefly introduced. For data collection, the number of data points collected for model training and model validation, the number of collection devices (e.g. meters or buildings), data sources, as well as the type of collected features are the aspects to be considered. Thus, for feature engineering, potential feature types (i.e. meteorological information, indoor environmental information, occupancy related data, time index, building characteristic data, socio-economic information, historical data) are firstly summarized to give an inspiration to data collection. Then, feature extraction methods, such as variable ranking, filter and wrapper methods, embedded method, PCA, and AE, are reviewed. To select proper feature extraction methods, their strengths and weaknesses are presented. As for data-driven algorithms, a variety of models are introduced in terms of principles as well as advantages and disadvantages. Factors reflected from the expected outputs (i.e. building type, energy type, scale, temporal granularity, and criteria) are also reviewed. A discussion is given to analyze the distribution of studies from 2015 to 2019.

Then, four input updating strategies are summarized to give a guide for utilizing the developed models in realistic multi-step energy prediction. The applications of these updating strategies are reviewed. Limitations of them are introduced: Strategy 1 is limited by the disability of considering time series property; Strategy 2 is weakened by the neglect of most recent historical data; Strategy 3 shows a disadvantage of accumulated error; Strategy 4 requires a huge amount of data in model training. Possible solutions for these drawbacks are also explained. However, novel updating strategies with high speed and accuracy are still required.

According to the distribution of reviewed studies and drawbacks shows in existing updating strategies, future works could focus on:

(1) Implementing novel data-driven techniques to real-life cases.

(2) Proposing guidelines for feature engineering and data-driven model selection under different cases.
(3) Fixing problems occurred in existing updating strategies, e.g. accumulated error.
(4) Integrating data-driven energy prediction models into other applications e.g. model predictive control strategies.

Abbreviations

AE	Autoencoder
ANN	Artificial Neural Network
AR	AutoRegressive Model
ARIMA	AutoRegression Integrated Moving Average
ARMA	AutoRegression-Moving Average
ASHRAE	The American Society of Heating, Refrigerating and Air-Conditioning Engineers
CNN	Convolutional Neural Networks
CV(RMSE)	Coefficient of Variation of the Root Mean Square Error
DNN	Deep Neural Networks
ELM	Extreme Learning Machine
FEMP	Federal Energy Management Program
GBDT	Gradient Boosting Decision Trees
IPMVP	International Performance Measurement and Verification Protocol
kNN	k-Nearest Neighbor
LR	Linear Regression
LSTM	Long Short-Term Memory
MA	Moving Average Model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
MSE	Mean Squared Error
NMBE	Normalized Mean Bias Error
PCA	Principal Component Analysis
PLSR	Partial Least Squares Regression
R^2	R Square
RF	Random Forests
RICNN	Recurrent Inception Convolution Neural Network
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks

RT	Regression Tree
SVM	Support Vector Machine
SVR	Support Vector Regression
XGB	Extreme Gradient Boosting

Nomenclature

c	Constant
h	Output of the hidden layer in ANN
p	Order of AR
q	Order of MA
r_{xy}	Pearson correlation coefficient between input feature x and target output y
rho_{xy}	Spearman's ranking correlation between input feature *x* and target output *y*
x_i	Input of i-th sample point
$\overline{x}$	Mean value of input feature
x_i'	Input rank of *i*-th individual sample points
$\overline{x'}$	Mean rank values of input feature
y_i	Target output of i-th sample point
$\overline{y}$	Mean value of target (measured) output
y_i'	Target output rank of *i*-th sample point
$\overline{y'}$	Mean rank values of target output
$\hat{y}$	Predicted output
w_0	Bias term
w	Weight matrix
$\varphi_1, \dots, \varphi_p$	Weights for AR
$\theta_1, \dots, \theta_q$	Weights for MA
ε	White noise
$\emptyset$	Activation function of output layer in ANN
σ	Activation function for the hidden layer in ANN

References

[1] IEA, "Energy Efficiency: Buildings." [Online]. Available: https://www.iea.org/topics/energyefficiency/buildings/. [Accessed: 11-Sep-2019].

[2] C. V. Gallagher, K. Leahy, P. O'Donovan, K. Bruton, and D. T. J. O'Sullivan, "Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0," *Energy Build.*, vol. 167, no. Journal Article, pp. 8–22, 2018, doi: 10.1016/j.enbuild.2018.02.023.

[3] C. V. Gallagher, K. Bruton, K. Leahy, and D. T. J. O'Sullivan, "The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings," *Energy Build.*, vol. 158, no. Journal Article, pp. 647–655, 2018, doi: 10.1016/j.enbuild.2017.10.041.

[4] M. S. Al-Homoud, "Computer-aided building energy analysis techniques," *Build. Environ.*, vol. 36, no. 4, pp. 421–433, May 2001, doi: 10.1016/S0360-1323(00)00026-3.

[5] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Build. Environ.*, vol. 43, no. 4, pp. 661–673, Apr. 2008, doi: 10.1016/j.buildenv.2006.10.027.

[6] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renew. Sustain. Energy Rev.*, vol. 23, no. Journal Article, pp. 272–288, 2013, doi: 10.1016/j.rser.2013.03.004.

[7] B. Dong, Z. Li, S. M. M. Rahman, and R. Vega, "A hybrid model approach for forecasting future residential electricity consumption," *Energy Build.*, vol. 117, no. Journal Article, pp. 341–351, 2016, doi: 10.1016/j.enbuild.2015.09.033.

[8] A. S. Ahmad *et al.*, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, May 2014, doi: 10.1016/j.rser.2014.01.069.

[9] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016, doi: 10.1016/j.ijforecast.2015.11.011.

[10] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustain. Cities Soc.*, vol. 35, no. Journal Article, pp. 257–270, 2017, doi: 10.1016/j.scs.2017.08.009.

[11] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renew. Sustain. Energy Rev.*, vol. 73, no. Journal Article, pp. 1104–1122, 2017, doi: 10.1016/j.rser.2017.02.023.

[12] M. A. Mat Daut, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, and F. Hussin, "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 1108–1118, Apr. 2017, doi: 10.1016/j.rser.2016.12.015.

[13] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," *Renew. Sustain. Energy Rev.*, vol. 75, no. Journal Article, pp. 796–808, 2017, doi: 10.1016/j.rser.2016.10.079.

[14] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 74, no. Journal Article, pp. 902–924, 2017, doi: 10.1016/j.rser.2017.02.085.

[15] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, no. Journal Article, pp. 1192–1205, 2018, doi: 10.1016/j.rser.2017.04.095.

[16] Y. Wei *et al.*, “A review of data-driven approaches for prediction and classification of building energy consumption,” *Renew. Sustain. Energy Rev.*, vol. 82, no. Journal Article, pp. 1027–1047, 2018, doi: 10.1016/j.rser.2017.09.108.

[17] T. Ahmad, H. Chen, Y. Guo, and J. Wang, “A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review,” *Energy Build.*, vol. 165, no. Journal Article, pp. 301–320, 2018, doi: 10.1016/j.enbuild.2018.01.017.

[18] S. R. Mohandes, X. Zhang, and A. Mahdiyar, “A comprehensive review on the application of artificial neural networks in building energy analysis,” *Neurocomputing*, vol. 340, pp. 55–75, May 2019, doi: 10.1016/j.neucom.2019.02.040.

[19] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, “Modeling and forecasting building energy consumption: A review of data-driven techniques,” *Sustain. Cities Soc.*, vol. 48, p. 101533, Jul. 2019, doi: 10.1016/j.scs.2019.101533.

[20] N. Wei, C. Li, X. Peng, F. Zeng, and X. Lu, “Conventional models and artificial intelligence-based models for energy consumption forecasting: A review,” *J. Pet. Sci. Eng.*, vol. 181, p. 106187, Oct. 2019, doi: 10.1016/j.petrol.2019.106187.

[21] I. Ben-Gal, “Outlier detection,” in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 131–146.

[22] V. Hodge and J. Austin, “A survey of outlier detection methodologies,” *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

[23] W. Badr, “6 Different Ways to Compensate for Missing Values In a Dataset (Data Imputation with examples),” *Medium*, 12-Jan-2019. [Online]. Available: https://towardsdatascience.com/6-different-ways-to-compensate-for-missing-values-data-imputation-with-examples-6022d9ca0779. [Accessed: 17-Oct-2019].

[24] M. Cai, M. Pipattanasomporn, and S. Rahman, “Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques,” *Appl. Energy*, vol. 236, pp. 1078–1088, Feb. 2019, doi: 10.1016/j.apenergy.2018.12.042.

[25] A. Alberini, G. Prettico, C. Shen, and J. Torriti, “Hot weather and residential hourly electricity demand in Italy,” *Energy*, vol. 177, pp. 44–56, Jun. 2019, doi: 10.1016/j.energy.2019.04.051.

[26] H. Thieblemont, F. Haghighat, A. Moreau, and G. Lacroix, “Control of electrically heated floor for building load management: A simplified self-learning predictive control approach,” *Energy Build.*, vol. 172, pp. 442–458, Aug. 2018, doi: 10.1016/j.enbuild.2018.04.042.

[27] “2013 ASHRAE Handbook - Fundamentals (I-P Edition) - 17. Residential Cooling and Heating Load Calculations - Knovel.” [Online]. Available: https://app.knovel.com/web/view/pdf/show.v/rcid:kpASHRAEB2/cid:kt00TYG4J5/viewerType:pdf//root_slug:17-residential-cooling-and-heating-load-calculations/url_slug:residential-cooling-heating?cid=kt00TYG4J5&b-toc-cid=kpASHRAEB2&b-toc-root-slug=&b-toc-url-slug=residential-cooling-heating&b-toc-title=2013%20ASHRAE%20Handbook%20-%20Fundamentals%20(I-P%20Edition). [Accessed: 22-Jul-2019].

[28] M. Chammas, A. Makhoul, and J. Demerjian, “An efficient data model for energy prediction using wireless sensors,” *Comput. Electr. Eng.*, vol. 76, pp. 249–257, Jun. 2019, doi: 10.1016/j.compeleceng.2019.04.002.

[29] Y. Ding, Q. Zhang, T. Yuan, and K. Yang, “Model input selection for building heating load prediction: A case study for an office building in Tianjin,” *Energy Build.*, vol. 159, no. Journal Article, pp. 254–270, 2018, doi: 10.1016/j.enbuild.2017.11.002.

[30] Y. Wei *et al.*, “Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks,” *Appl. Energy*, vol. 240, pp. 276–294, Apr. 2019, doi: 10.1016/j.apenergy.2019.02.056.

[31] S. S. Shetty, D. C. Hoang, M. Gupta, and S. K. Panda, "Learning desk fan usage preferences for personalised thermal comfort in shared offices using tree-based methods," *Build. Environ.*, vol. 149, pp. 546–560, Feb. 2019, doi: 10.1016/j.buildenv.2018.12.040.
[32] S.-H. Mun, Y. Kwak, and J.-H. Huh, "A case-centered behavior analysis and operation prediction of AC use in residential buildings," *Energy Build.*, vol. 188–189, pp. 137–148, Apr. 2019, doi: 10.1016/j.enbuild.2019.02.012.
[33] M. Behl, F. Smarra, and R. Mangharam, "DR-Advisor: A data-driven demand response recommender system," *Appl. Energy*, vol. 170, no. Journal Article, pp. 30–46, 2016, doi: 10.1016/j.apenergy.2016.02.090.
[34] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy Build.*, vol. 49, no. Generic, pp. 591–603, 2012, doi: 10.1016/j.enbuild.2012.03.010.
[35] Z. Yu, B. C. Fung, F. Haghighat, H. Yoshino, and E. Morofsky, "A systematic procedure to study the influence of occupant behavior on building energy consumption," *Energy Build.*, vol. 43, no. 6, pp. 1409–1417, 2011.
[36] M. Ashouri, F. Haghighat, B. C. M. Fung, and H. Yoshino, "Development of a ranking procedure for energy performance evaluation of buildings based on occupant behavior," *Energy Build.*, vol. 183, pp. 659–671, Jan. 2019, doi: 10.1016/j.enbuild.2018.11.050.
[37] Z. Wang, T. Hong, and M. A. Piette, "Predicting plug loads with occupant count data through a deep learning approach," *Energy*, vol. 181, pp. 29–42, Aug. 2019, doi: 10.1016/j.energy.2019.05.138.
[38] E. Sala-Cardoso, M. Delgado-Prieto, K. Kampouropoulos, and L. Romeral, "Activity-aware HVAC power demand forecasting," *Energy Build.*, vol. 170, no. Journal Article, pp. 15–24, 2018, doi: 10.1016/j.enbuild.2018.03.087.
[39] M. S. Gul and S. Patidar, "Understanding the energy consumption and occupancy of a multi-purpose academic building," *Energy Build.*, vol. 87, pp. 155–165, Jan. 2015, doi: 10.1016/j.enbuild.2014.11.027.
[40] C. Fan, F. Xiao, and S. Wang, "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques," *Appl. Energy*, vol. 127, pp. 1–10, Aug. 2014, doi: 10.1016/j.apenergy.2014.04.016.
[41] S. Seyedzadeh, F. Pour Rahimian, P. Rastogi, and I. Glesk, "Tuning machine learning models for prediction of building energy loads," *Sustain. Cities Soc.*, vol. 47, p. 101484, May 2019, doi: 10.1016/j.scs.2019.101484.
[42] L. Wei, W. Tian, J. Zuo, Z.-Y. Yang, Y. Liu, and S. Yang, "Effects of Building Form on Energy Use for Buildings in Cold Climate Regions," *8th Int. Cold Clim. HVAC Conf.*, vol. 146, no. Journal Article, pp. 182–189, 2016, doi: 10.1016/j.proeng.2016.06.370.
[43] B. Talebi, F. Haghighat, and P. A. Mirzaei, "Simplified model to predict the thermal demand profile of districts," *Energy Build.*, vol. 145, pp. 213–225, Jun. 2017, doi: 10.1016/j.enbuild.2017.03.062.
[44] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools," *Energy Build.*, vol. 49, no. Generic, pp. 560–567, 2012, doi: 10.1016/j.enbuild.2012.03.003.
[45] S. Sekhar Roy, R. Roy, and V. E. Balas, "Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM," *Renew. Sustain. Energy Rev.*, vol. 82, no. Journal Article, pp. 4256–4268, 2018, doi: 10.1016/j.rser.2017.05.249.

[46] S. Kumar, 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 Build.*, vol. 176, no. Journal Article, pp. 275–286, 2018, doi: 10.1016/j.enbuild.2018.06.056.

[47] S. Kumar, S. K. Pal, and R. P. Singh, “Intra ELM variants ensemble based model to predict energy performance in residential buildings,” *Sustain. Energy Grids Netw.*, vol. 16, no. Journal Article, pp. 177–187, 2018, doi: 10.1016/j.segan.2018.07.001.

[48] K. B. Lindberg, P. Seljom, H. Madsen, D. Fischer, and M. Korpås, “Long-term electricity load forecasting: Current and future trends,” *Util. Policy*, vol. 58, pp. 102–119, Jun. 2019, doi: 10.1016/j.jup.2019.04.001.

[49] Y. He, Y. Qin, S. Wang, X. Wang, and C. Wang, “Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network,” *Appl. Energy*, vol. 233–234, pp. 565–575, Jan. 2019, doi: 10.1016/j.apenergy.2018.10.061.

[50] Y. He, Y. Zheng, and Q. Xu, “Forecasting energy consumption in Anhui province of China through two Box-Cox transformation quantile regression probability density methods,” *Measurement*, vol. 136, pp. 579–593, Mar. 2019, doi: 10.1016/j.measurement.2019.01.008.

[51] O. F. Beyca, B. C. Ervural, E. Tatoglu, P. G. Ozuyar, and S. Zaim, “Using machine learning tools for forecasting natural gas consumption in the province of Istanbul,” *Energy Econ.*, vol. 80, pp. 937–949, May 2019, doi: 10.1016/j.eneco.2019.03.006.

[52] R. Wang, S. Lu, and Q. Li, “Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings,” *Sustain. Cities Soc.*, vol. 49, p. 101623, Aug. 2019, doi: 10.1016/j.scs.2019.101623.

[53] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption,” *Energy Build.*, vol. 147, no. Journal Article, pp. 77–89, 2017, doi: 10.1016/j.enbuild.2017.04.038.

[54] Y. Ding, Q. Zhang, T. Yuan, and F. Yang, “Effect of input variables on cooling load prediction accuracy of an office building,” *Appl. Therm. Eng.*, vol. 128, no. Journal Article, pp. 225–234, 2018, doi: 10.1016/j.applthermaleng.2017.09.007.

[55] Y. Huang, Y. Yuan, H. Chen, J. Wang, Y. Guo, and T. Ahmad, “A novel energy demand prediction strategy for residential buildings based on ensemble learning,” *Energy Procedia*, vol. 158, pp. 3411–3416, Feb. 2019, doi: 10.1016/j.egypro.2019.01.935.

[56] Z. Wang, Y. Wang, and R. S. Srinivasan, “A novel ensemble learning approach to support building energy use prediction,” *Energy Build.*, vol. 159, pp. 109–122, Jan. 2018, doi: 10.1016/j.enbuild.2017.10.085.

[57] K. Pearson, “VII. Note on regression and inheritance in the case of two parents,” *Proc. R. Soc. Lond.*, vol. 58, no. 347–352, pp. 240–242, 1895.

[58] C. Spearman, “The proof and measurement of association between two things.,” 1961.

[59] M. Claesen and B. De Moor, “Hyperparameter search in machine learning,” *ArXiv Prepr. ArXiv150202127*, 2015.

[60] A. Zeng, S. Liu, and Y. Yu, “Comparative study of data driven methods in building electricity use prediction,” *Energy Build.*, vol. 194, pp. 289–300, Jul. 2019, doi: 10.1016/j.enbuild.2019.04.029.

[61] U. M. Khaire and R. Dhanalakshmi, “Stability of feature selection algorithm: A review,” *J. King Saud Univ. - Comput. Inf. Sci.*, Jun. 2019, doi: 10.1016/j.jksuci.2019.06.012.

[62] P. Yuan, L. Duanmu, and Z. Wang, “Coal consumption prediction model of space heating with feature selection for rural residences in severe cold area in China,” *Sustain. Cities Soc.*, vol. 50, p. 101643, Oct. 2019, doi: 10.1016/j.scs.2019.101643.

[63] ASHRAE, *ASHRAE Guideline 14–2014, Measurement of Energy, Demand, and Water Savings*. ASHRAE Atlanta, 2014.

[64] Jain R. K., Damoulas T., and Kontokosta C. E., “Towards Data-Driven Energy Consumption Forecasting of Multi-Family Residential Buildings: Feature Selection via The Lasso,” *Comput. Civ. Build. Eng. 2014*, pp. 1675–1682, doi: 10.1061/9780784413616.208.
[65] K. Pearson, “LIII. On lines and planes of closest fit to systems of points in space,” *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.
[66] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.
[67] L. Xuemei, D. Lixing, L. Jinhu, X. Gang, and L. Jibin, “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, pp. 522–526, doi: 10.1109/WKDD.2010.137.
[68] “From Autoencoder to Beta-VAE,” *Lil’Log*, 12-Aug-2018. [Online]. Available: https://lilianweng.github.io/2018/08/12/from-autoencoder-to-beta-vae.html. [Accessed: 24-Jul-2019].
[69] C. Fan, Y. Sun, Y. Zhao, M. Song, and J. Wang, “Deep learning-based feature engineering methods for improved building energy prediction,” *Appl. Energy*, vol. 240, pp. 35–45, Apr. 2019, doi: 10.1016/j.apenergy.2019.02.052.
[70] S. Mujeeb and N. Javaid, “ESAENARX and DE-RELM: Novel Schemes for Big Data Predictive Analytics of Electricity Load and Price,” *Sustain. Cities Soc.*, Jul. 2019, doi: 10.1016/j.scs.2019.101642.
[71] *8.1 Stationarity and differencing | Forecasting: Principles and Practice*. .
[72] G. James and SpringerLink (Online service), Eds., *An introduction to statistical learning: with applications in R*. New York, NY: Springer, 2013.
[73] G. K. F. Tso and K. K. W. Yau, “Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks,” *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep. 2007, doi: 10.1016/j.energy.2006.11.010.
[74] D. Basak, “Support Vector Regression,” *Int. J. Neural Inf. Process. – Lett. Rev.*, vol. 11, pp. 203–224, 2007.
[75] S. M. Clarke, J. H. Griebsch, and T. W. Simpson, “Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses,” *J. Mech. Des.*, vol. 127, no. 6, pp. 1077–1087, Aug. 2004, doi: 10.1115/1.1897403.
[76] M. Awad and R. Khanna, “Support Vector Regression,” in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, M. Awad and R. Khanna, Eds. Berkeley, CA: Apress, 2015, pp. 67–80.
[77] S. C. Kleene, “Representation of events in nerve nets and finite automata,” RAND PROJECT AIR FORCE SANTA MONICA CA, 1951.
[78] Y. Bengio, “Learning Deep Architectures for AI,” *Found Trends Mach Learn*, vol. 2, no. 1, pp. 1–127, Jan. 2009, doi: 10.1561/2200000006.
[79] A. Zell *et al.*, “SNNS (stuttgart neural network simulator),” in *Neural network simulation environments*, Springer, 1994, pp. 165–186.
[80] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016.
[81] “[1409.1556] Very Deep Convolutional Networks for Large-Scale Image Recognition.” [Online]. Available: https://arxiv.org/abs/1409.1556. [Accessed: 06-Nov-2019].
[82] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[83] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network,” in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, 2014, pp. 844–848.

[84] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *ArXiv14085882 Cs*, Sep. 2014.

[85] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365–377, May 2019, doi: 10.1016/j.energy.2019.03.081.

[86] D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999, doi: 10.1613/jair.614.

[87] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, "Random Forest based hourly building energy prediction," *Energy Build.*, vol. 171, no. Journal Article, pp. 11–25, 2018, doi: 10.1016/j.enbuild.2018.04.008.

[88] N. J. Johannesen, M. Kolhe, and M. Goodwin, "Relative evaluation of regression tools for urban area electrical energy demand forecasting," *J. Clean. Prod.*, vol. 218, pp. 555–564, May 2019, doi: 10.1016/j.jclepro.2019.01.108.

[89] C. Robinson *et al.*, "Machine learning approaches for estimating commercial building energy consumption," *Appl. Energy*, vol. 208, pp. 889–904, Dec. 2017, doi: 10.1016/j.apenergy.2017.09.060.

[90] J. Walther, D. Spanier, N. Panten, and E. Abele, "Very short-term load forecasting on factory level – A machine learning approach," *Procedia CIRP*, vol. 80, pp. 705–710, Jan. 2019, doi: 10.1016/j.procir.2019.01.060.

[91] N. R. C. Government of Canada, "Residential Sector Canada Table 2: Secondary Energy Use and GHG Emissions by End-Use," 30-Nov-2018. [Online]. Available: http://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/showTable.cfm?type=CP&sector=res&juris=ca&rn=2&page=0. [Accessed: 28-Jun-2019].

[92] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," *Energy Build.*, vol. 128, pp. 198–213, Sep. 2016, doi: 10.1016/j.enbuild.2016.06.089.

[93] I. Blom, L. Itard, and A. Meijer, "Environmental impact of building-related and user-related energy consumption in dwellings," *Build. Environ.*, vol. 46, no. 8, pp. 1657–1669, Aug. 2011, doi: 10.1016/j.buildenv.2011.02.002.

[94] Y. Ding, Q. Zhang, and T. Yuan, "Research on short-term and ultra-short-term cooling load prediction models for office buildings," *Energy Build.*, vol. 154, pp. 254–267, Nov. 2017, doi: 10.1016/j.enbuild.2017.08.077.

[95] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005.

[96] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.

[97] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 2006.

[98] A. C. Cameron and F. A. Windmeijer, "An R-squared measure of goodness of fit for some common nonlinear regression models," *J. Econom.*, vol. 77, no. 2, pp. 329–342, 1997.

[99] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural network studies. 1. Comparison of overfitting and overtraining," *J. Chem. Inf. Comput. Sci.*, vol. 35, no. 5, pp. 826–833, 1995.

[100] L. Webster *et al.*, "M&V Guidelines: Measurement and Verification for Performance-Based Contracts," Version 4.0, Technical Report, 2015.

[101] I. Committee, “International Performance Measurement and Verification Protocol: Concepts and options for determining energy and water savings, Volume I,” National Renewable Energy Lab., Golden, CO (US), 2001.
[102] “Scopus - Document search.” [Online]. Available: https://www.scopus.com/search/form.uri?display=basic. [Accessed: 21-Feb-2020].
[103] “ISO New England - Energy, Load, and Demand Reports.” [Online]. Available: https://www.iso-ne.com/isoexpress/web/reports/load-and-demand. [Accessed: 21-Feb-2020].
[104] “Electricity Data - U.S. Energy Information Administration (EIA).” [Online]. Available: https://www.eia.gov/electricity/data.php. [Accessed: 21-Feb-2020].
[105] “China YearBooks | China Statistical Yearbooks Database.” [Online]. Available: https://www.chinayearbooks.com/. [Accessed: 21-Feb-2020].
[106] “UCI Machine Learning Repository.” [Online]. Available: http://archive.ics.uci.edu/ml/index.php. [Accessed: 21-Feb-2020].
[107] I. S. S. D. Archive, “Home, Irish Social Science Data Archive,” *Irish Social Science Data Archive*. [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/. [Accessed: 21-Feb-2020].
[108] “olypen - Trac.” [Online]. Available: https://svn.pnl.gov/olypen. [Accessed: 21-Feb-2020].
[109] “» Natural Gas Demand NaturalGas.org.” [Online]. Available: http://naturalgas.org/business/demand/. [Accessed: 21-Feb-2020].
[110] “EnerNOC Open :: Data.” [Online]. Available: https://open-enernoc-data.s3.amazonaws.com/anon/index.html. [Accessed: 24-Feb-2020].
[111] P. Geyer and S. Singaravel, “Component-based machine learning for performance prediction in building design,” *Appl. Energy*, vol. 228, no. Journal Article, pp. 1439–1453, 2018, doi: 10.1016/j.apenergy.2018.07.011.
[112] A. González-Vidal, F. Jiménez, and A. F. Gómez-Skarmeta, “A methodology for energy multivariate time series forecasting in smart buildings based on feature selection,” *Energy Build.*, vol. 196, pp. 71–82, Aug. 2019, doi: 10.1016/j.enbuild.2019.05.021.
[113] H. Luo, H. Cai, H. Yu, Y. Sun, Z. Bi, and L. Jiang, “A short-term energy prediction system based on edge computing for smart city,” *Future Gener. Comput. Syst.*, vol. 101, pp. 444–457, Dec. 2019, doi: 10.1016/j.future.2019.06.030.
[114] T. Ahmad and H. Chen, “Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment,” *Energy*, vol. 160, pp. 1008–1020, Oct. 2018, doi: 10.1016/j.energy.2018.07.084.
[115] L. Xu, S. Wang, and R. Tang, “Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load,” *Appl. Energy*, vol. 237, pp. 180–195, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.022.
[116] L. M. Candanedo, V. Feldheim, and D. Deramaix, “Data driven prediction models of energy use of appliances in a low-energy house,” *Energy Build.*, vol. 140, no. Journal Article, pp. 81–97, 2017, doi: 10.1016/j.enbuild.2017.01.083.
[117] S. Salcedo-Sanz, J. Muñoz-Bulnes, J. A. Portilla-Figueras, and J. Del Ser, “One-year-ahead energy demand estimation from macroeconomic variables using computational intelligence algorithms,” *Energy Convers. Manag.*, vol. 99, pp. 62–71, Jul. 2015, doi: 10.1016/j.enconman.2015.03.109.
[118] J. Sánchez-Oro, A. Duarte, and S. Salcedo-Sanz, “Robust total energy demand estimation with a hybrid Variable Neighborhood Search – Extreme Learning Machine algorithm,” *Energy Convers. Manag.*, vol. 123, pp. 445–452, Sep. 2016, doi: 10.1016/j.enconman.2016.06.050.
[119] K. Amasyali and N. El-Gohary, “Building Lighting Energy Consumption Prediction for Supporting Energy Data Analytics,” *ICSDEC 2016 – Integrating Data Sci. Constr. Sustain.*, vol. 145, no. Journal Article, pp. 511–517, 2016, doi: 10.1016/j.proeng.2016.04.036.

[120] Y. Guo *et al.*, “Machine learning-based thermal response time ahead energy demand prediction for building heating systems,” *Appl. Energy*, vol. 221, no. Journal Article, pp. 16–27, 2018, doi: 10.1016/j.apenergy.2018.03.125.
[121] Y. Guo, G. Li, H. Chen, J. Wang, and Y. Huang, “A thermal response time ahead energy demand prediction strategy for building heating system using machine learning methods,” *Proc. 9th Int. Conf. Appl. Energy*, vol. 142, no. Journal Article, pp. 1003–1008, 2017, doi: 10.1016/j.egypro.2017.12.346.
[122] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, “Deep belief network based electricity load forecasting: An analysis of Macedonian case,” *Energy*, vol. 115, pp. 1688–1700, Nov. 2016, doi: 10.1016/j.energy.2016.07.090.
[123] S. Idowu, S. Saguna, C. Åhlund, and O. Schelén, “Applied machine learning: Forecasting heat load in district heating system,” *Energy Build.*, vol. 133, no. Journal Article, pp. 478–488, 2016, doi: 10.1016/j.enbuild.2016.09.068.
[124] Y. Chen, H. Tan, and U. Berardi, “Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: A clustering-based hybrid approach,” *Energy Build.*, vol. 148, no. Journal Article, pp. 228–237, 2017, doi: 10.1016/j.enbuild.2017.05.003.
[125] Y. Chen and H. Tan, “Short-term prediction of electric demand in building sector via hybrid support vector regression,” *Appl. Energy*, vol. 204, no. Journal Article, pp. 1363–1374, 2017, doi: 10.1016/j.apenergy.2017.03.070.
[126] P. Singh, P. Dwivedi, and V. Kant, “A hybrid method based on neural network and improved environmental adaptation method using Controlled Gaussian Mutation with real parameter for short-term load forecasting,” *Energy*, vol. 174, pp. 460–477, May 2019, doi: 10.1016/j.energy.2019.02.141.
[127] D. Geysen, O. De Somer, C. Johansson, J. Brage, and D. Vanhoudt, “Operational thermal load forecasting in district heating networks using machine learning and expert advice,” *Energy Build.*, vol. 162, no. Journal Article, pp. 144–153, 2018, doi: 10.1016/j.enbuild.2017.12.042.
[128] A. Khoshrou and E. J. Pauwels, “Short-term scenario-based probabilistic load forecasting: A data-driven approach,” *Appl. Energy*, vol. 238, pp. 1258–1268, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.155.
[129] S. Chapaloglou *et al.*, “Smart energy management algorithm for load smoothing and peak shaving based on load forecasting of an island’s power system,” *Appl. Energy*, vol. 238, pp. 627–642, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.102.
[130] J. Reynolds, M. W. Ahmad, Y. Rezgui, and J.-L. Hippolyte, “Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm,” *Appl. Energy*, vol. 235, pp. 699–713, Feb. 2019, doi: 10.1016/j.apenergy.2018.11.001.
[131] C. Deb, L. S. Eang, J. Yang, and M. Santamouris, “Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks,” *Energy Build.*, vol. 121, no. Journal Article, pp. 284–297, 2016, doi: 10.1016/j.enbuild.2015.12.050.
[132] C. Fan, F. Xiao, and Y. Zhao, “A short-term building cooling load prediction method using deep learning algorithms,” *Appl. Energy*, vol. 195, pp. 222–233, Jun. 2017, doi: 10.1016/j.apenergy.2017.03.064.
[133] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, “Deep learning for estimating building energy consumption,” *Sustain. Energy Grids Netw.*, vol. 6, pp. 91–99, Jun. 2016, doi: 10.1016/j.segan.2016.02.005.
[134] T. Ahmad, H. Chen, J. Shair, and C. Xu, “Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management,” *Int. J. Refrig.*, vol. 98, pp. 399–409, Feb. 2019, doi: 10.1016/j.ijrefrig.2018.10.017.

[135] C. Xu *et al.*, “Modal decomposition based ensemble learning for ground source heat pump systems load forecasting,” *Energy Build.*, vol. 194, pp. 62–74, Jul. 2019, doi: 10.1016/j.enbuild.2019.04.018.

[136] J.-S. Chou and N.-T. Ngo, “Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns,” *Appl. Energy*, vol. 177, no. Journal Article, pp. 751–770, 2016, doi: 10.1016/j.apenergy.2016.05.074.

[137] J.-S. Chou and N.-S. Truong, “Cloud forecasting system for monitoring and alerting of energy use by home appliances,” *Appl. Energy*, vol. 249, pp. 166–177, Sep. 2019, doi: 10.1016/j.apenergy.2019.04.063.

[138] A. Bracale, G. Carpinelli, P. De Falco, and T. Hong, “Short-term industrial reactive power forecasting,” *Int. J. Electr. Power Energy Syst.*, vol. 107, pp. 177–185, May 2019, doi: 10.1016/j.ijepes.2018.11.022.

[139] S. J. Cox, D. Kim, H. Cho, and P. Mago, “Real time optimal control of district cooling system with thermal energy storage using neural networks,” *Appl. Energy*, vol. 238, pp. 466–480, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.093.

[140] E. T. Al-Shammari *et al.*, “Prediction of heat load in district heating systems by Support Vector Machine with Firefly searching algorithm,” *Energy*, vol. 95, pp. 266–273, Jan. 2016, doi: 10.1016/j.energy.2015.11.079.

[141] J. Zhao, W. Wang, and C. Sheng, *Data-Driven Prediction for Industrial Processes and Their Applications*. Springer, 2018.

[142] J. Gu, J. Wang, C. Qi, C. Min, and B. Sundén, “Medium-term heat load prediction for an existing residential building based on a wireless on-off control system,” *Energy*, vol. 152, no. Journal Article, pp. 709–718, 2018, doi: 10.1016/j.energy.2018.03.179.

[143] K. Grolinger, A. L’Heureux, M. A. M. Capretz, and L. Seewald, “Energy Forecasting for Event Venues: Big Data and Prediction Accuracy,” *Energy Build.*, vol. 112, no. Journal Article, pp. 222–233, 2016, doi: 10.1016/j.enbuild.2015.12.010.

[144] C. Zhang, Y. Zhao, X. Zhang, C. Fan, and T. Li, “An Improved Cooling Load Prediction Method for Buildings with the Estimation of Prediction Intervals,” *10th Int. Symp. Heat. Vent. Air Cond. ISHVAC2017 19-22 Oct. 2017 Jinan China*, vol. 205, no. Journal Article, pp. 2422–2428, 2017, doi: 10.1016/j.proeng.2017.09.967.

[145] Z. Wang, T. Hong, and M. A. Piette, “Data fusion in predicting internal heat gains for office buildings through a deep learning approach,” *Appl. Energy*, vol. 240, pp. 386–398, Apr. 2019, doi: 10.1016/j.apenergy.2019.02.066.

[146] A. Rahman, V. Srikumar, and A. D. Smith, “Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,” *Appl. Energy*, vol. 212, no. Journal Article, pp. 372–385, 2018, doi: 10.1016/j.apenergy.2017.12.051.

[147] T.-Y. Kim and S.-B. Cho, “Predicting residential energy consumption using CNN-LSTM neural networks,” *Energy*, vol. 182, pp. 72–81, Sep. 2019, doi: 10.1016/j.energy.2019.05.230.

[148] N. Kim, M. Kim, and J. K. Choi, “LSTM Based Short-term Electricity Consumption Forecast with Daily Load Profile Sequences,” in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, 2018, pp. 136–137, doi: 10.1109/GCCE.2018.8574484.

[149] J. Bedi and D. Toshniwal, “Deep learning framework to forecast electricity demand,” *Appl. Energy*, vol. 238, pp. 1312–1326, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.113.

[150] Z. Zheng, H. Chen, and X. Luo, “Spatial granularity analysis on electricity consumption prediction using LSTM recurrent neural network,” *Energy Procedia*, vol. 158, pp. 2713–2718, Feb. 2019, doi: 10.1016/j.egypro.2019.02.027.

[151] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, “Probabilistic individual load forecasting using pinball loss guided LSTM,” *Appl. Energy*, vol. 235, pp. 10–20, Feb. 2019, doi: 10.1016/j.apenergy.2018.10.078.

[152] J. Kim, J. Moon, E. Hwang, and P. Kang, “Recurrent inception convolution neural network for multi short-term load forecasting,” *Energy Build.*, vol. 194, pp. 328–341, Jul. 2019, doi: 10.1016/j.enbuild.2019.04.034.

[153] Z. Lee, K. Gupta, K. J. Kircher, and K. M. Zhang, “Mixed-integer model predictive control of variable-speed heat pumps,” *Energy Build.*, vol. 198, pp. 75–83, Sep. 2019, doi: 10.1016/j.enbuild.2019.05.060.