Energy & Buildings 326 (2025) 115001

Contents lists available at ScienceDirect

Energy & Buildings

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

Data-driven building load prediction and large language models: Comprehensive overview

Yake Zhang [a], Dijun Wang [b], Guansong Wang [c], Peng Xu [a,*], Yihao Zhu [b]

[a] School of Mechanical Engineering, Tongji University, Shanghai 200092, China
[b] Guangzhou Metro Design and Research Institute CO., LTD., Guangzhou 510030, China
[c] School of Software Engineering, Tongji University, Shanghai 200092, China

ARTICLE INFO

Keywords:
Data-driven approach
Building load prediction
Machine learning
Large language models
Feature engineering
Data preparation
Room-scale load prediction
Retrieval augmented generation

ABSTRACT

Building load forecasting is essential for optimizing the architectural design and managing energy efficiently, enhancing the performance of Heating, Ventilation, and Air Conditioning systems, and enhancing occupant comfort. With advancements in data science and machine learning, the focus on predicting building loads through data analysis has significantly intensified as a research domain. However, previous studies have typically faced challenges such as data scarcity, improper feature extraction methods, and weak model generalization capabilities. To gain a deeper understanding of these issues, a comprehensive review of data processing, feature selection, and model selection methods in previous research is conducted from the perspective of the entire load forecasting process. The aim is to identify the most suitable methods for each step of load forecasting to enhance prediction accuracy. This review surveys the research progress of statistical learning methods, traditional machine learning methods, deep learning methods, and hybrid methods in different application scenarios of building load prediction. Then, it emphasized the critical role of data preprocessing and focused on techniques like data fusion and transfer learning to overcome data shortages and bolster the models' ability to generalize. Moreover, the obtainment of significant features from building characteristics, weather data, and operational statistics to boost prediction accuracy is explored. A notable contribution of this review is the proposed technical framework for EnergyPlus model generation using LLM-based Retrieval Augmented Generation (RAG) technology and room- level load prediction with Spatio-Temporal Graph Neural Networks. This framework utilize architectural design drawings to achieve an "end-to-end" prediction process, aiming to reduce the professional threshold of load prediction and provide technical support for fine-grained regulation of building operation. Exploratory experiment is conducted using a single-zone building model to verify the feasibility of LLM-generated EnergyPlus models, with IDF simulation file generation taking only 196 s. Room-level load forecasting with LLMs remains to be explored further. It is reasonable to believe that the methods proposed in this review hold promise for advancing data-driven building load forecasting technologies.

1. Introduction

The 2022 International Energy Agency report states that energy consumption during the construction and operational stages of buildings represents almost 30 % of worldwide energy consumption, positioning it as a significant source of worldwide carbon emissions [1]. Building load prediction is essential for designing and optimizing building energy systems [2], and it is also fundamental to the optimization models for predicted air conditioning energy consumption, which was first applied in this domain in 2003 [3] and has evolved since then.

Abbreviations: AE, Autoencoders; ANN, Artificial Neural Networks; AR, Autoregressive models; ARIMA, Autoregressive Integrated Moving Average; CNN, Convolutional Neural Networks; DT, Decision Trees; GA, Genetic Algorithm; GAN, Generative Adversarial Networks; HVAC, Heating, Ventilation, and Air Conditioning systems; KNN, K-Nearest Neighbors; KPCA, Kernel Principal Component Analysis; LLM, Large Language Models; LR, linear regression; LSTM, Long Short-Term Memory networks; LTLF, Long-Term Load Forecasting; MA, Moving Average models; MLR, multivariate linear regression; PCA, Principal Component Analysis; RBFN, Radial Basis Function Networks; RBFNN, Radial Basis Function networks; RF, Random Forests; RNN, Recurrent Neural Networks; STGNN, Spatio-Temporal Graph Neural Network; STLF, Short-term Load Forecasting; SVR, Support Vector Regression; XGBoost, Extreme Gradient Boosting.

* Corresponding author.
E-mail address: xupeng@tongji.edu.cn (P. Xu).

https://doi.org/10.1016/j.enbuild.2024.115001
Received 2 August 2024; Received in revised form 2 October 2024; Accepted 3 November 2024
Available online 9 November 2024

Building systems are primarily composed of air conditioning, heating, lighting, equipment, occupants, etc. The complexity of these factors makes building load forecasting a variable and multifaceted issue. Building load prediction models can be developed using physical principles or historical data.

Physical models (white box model) are based on thermodynamic rules and leverage the heat and mass balance equations, analyzing the physical relationship between load influence parameters and loads. By establishing a mathematical representation that links input variables to output variables, these models can forecast load by solving the equations. The quest for accurate building load prediction spurred the development of numerous building energy simulation software based on physical models since the mid-1970s. Examples include well-established tools like BLAST [4], DOE-2 [5], EnergyPlus [6,7], TRNSYS [8], ESP-r [9], DeST [10,11], Dymola [12], and others. The modular, flexibility, and step-by-step calculation capabilities of these simulation tools make them highly versatile for comprehensively analyzing building performance. Owing to their foundation in physical principles, physical models have high interpretability. However, their modeling process is complex, computation-intensive, and demands high levels of expertise from the users.

Data-driven models have gained prominence due to their powerful model expression capabilities and ability to learn complex relationships. Unlike physical models, data-driven models form predictive capabilities by learning and extracting features from large historical multi-dimensional datasets of buildings. Their predictive performance depends on both the quality and the volume of the data used for training. Data-driven models, with their dependence on historical data only, simplicity of operation, and strong nonlinear approximation ability, have significant advantages in modeling complex systems that are difficult to accurately describe by explicit physical processes, which in turn has become a crucial area of research related to building load forecasting.

The precision of building load forecasts is influenced by the model selected, alongside the data's quality and quantity [13]. First, the complexity and time-variability of data in building systems, compounded by environmental uncertainties, make prediction more difficult. Secondly, data acquisition and processing also face technical difficulties, such as how to handle incomplete data and deal with noise [14].

Despite its limitations, data-driven load prediction has received considerable research attention in recent years, with several review studies focusing on existing data-driven works. To better understand the development trends and research hotspots of data-driven building load prediction, Table 1 comprehensively reviews the content of review papers published after 2014 from the perspectives of data preparation, feature engineering, data-driven algorithms, and application scenarios. Ahmad et al. [15] focused on the use of SVM and ANN for electrical load forecasting. Deb et al. [16] conducted a thorough review of time series methods in building load prediction. Runge et al. [13] offered a detailed analysis of ANN and its variants used in estimating building energy consumption. However, most reviews remained focused on the algorithmic aspects. In 2018, Amasyali et al. [17] reviewed building load prediction research from the perspectives of prediction models, prediction scope, data types and dataset sizes, data preprocessing methods, and feature types, suggesting that long-term, residential, and lighting building energy consumption prediction deserves more attention. As data-driven building load prediction research has matured, researchers have gradually recognized the importance of the data and input features required to build models. Since then, review articles have begun to focus not only on the models but also on data preprocessing and feature engineering to model construction. Wei et al. [18] reviewed various types and granularities of data-driven methods, summarized the current development and practical applications of these methods, and suggested that appropriate modifications and incorporation of renewable energy technologies could promote micro-changes in the future energy use of specific buildings. Bourdeaua et al. [19] particularly focused on different machine learning methods, discussing the importance of integrating occupant behavior into data-driven models. Tom M. Mitchell [20] defined machine learning as "a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P." Zhang et al. [21] reviewed the machine learning techniques applied in building load prediction within the context of machine learning architecture. This review included the application of building load prediction models (task T), modeling algorithms aimed at improving machine learning performance and accuracy (performance P), and an examination of data-related aspects such as data collection, data preprocessing, feature extraction, and selection (experience E).

These reviews are indeed useful, but they mainly scrutinize prediction models from earlier studies, which are often quite basic. Most reviews only briefly mention auxiliary techniques like data and feature engineering in building load prediction, without detailed analysis. Limitations of existing literature reviews are summarized below:

1) The characteristics of the data used for model training and validation will affect the reproducibility and generalization of the studied techniques. Data collection and data cleaning are often considered preliminary steps before conducting research, and they are rarely summarized in previous review papers.
2) There is a lack of systematic review on feature selection methods. Most reviews only introduce a few feature selection algorithms, but the basic principles, respective advantages and disadvantages, and applicable scenarios of each method still need further summarization.
3) Diverse and high-quality test data are the foundation for building high-performance prediction models. However, the test data used in existing studies are often not publicly available, making it difficult to share the insights gained during the load prediction process and to validate the constructed data-driven models. Even though a few reviews have summarized the data characteristics and selection processes used in different studies, these only provide a rough overview, and there is still a lack of discussion on public datasets and a systematic review of feature engineering methods.
4) Relatively novel data-driven models are often not included. Thanks to the rapid development of artificial intelligence, models such as Long Short-Term Memory networks (LSTM), Convolutional Neural Networks (CNN), and Generative Adversarial Networks (GAN) are increasingly being applied to load prediction, but reviews of the latest research are lacking.

With the advancement of artificial intelligence, data-driven technologies have gradually evolved from early statistical-based models to modern variants based on deep learning. Furthermore, recently emerging large language models like BERT [22] and GPT3 [23], which demonstrate strong feature representation learning and few-shot learning capabilities [24], have been used to solve specific domain tasks in multiple fields [24–27], achieving significant results. They are also expected to become an ideal choice for building load prediction problems.

As a relatively traditional application, load forecasting shows significant potential with the integration of LLMs: 1) Key information can be extracted from complex, unstructured data (e.g., maintenance logs, user feedback) to identify equipment anomalies and usage patterns; 2) Different data modalities are efficiently integrated, improving prediction accuracy; 3) Natural language interaction simplifies system operation, enhancing usability; 4) LLMs' context understanding and transfer learning capabilities enhance model generalization using similar buildings' data, enabling load forecasting for buildings lacking historical energy consumption data.

Therefore, this review systematically revisits research on building load prediction using data-driven algorithms. It delves into the application of data-driven methods in different scenarios, the classification of

Table 1
Review contents of data-driven building load prediction.

No.	Author(s)	Ref.	Year	Data Preparation		Feature Engineering		Models	Application	Comments
				source	methods	type	methods			
1	Ahmad et al.	[15]	2014	-	-	-	-	Support Vector Machine(SVM); Artificial Neural Network (ANN); Hybrid	Electrical energy consumption	• Emphasized the effectiveness of ANN and SVM in forecasting building electrical energy consumption and explored hybrid approaches.
2	Raza and Khosravi	[28]	2015	-	-	√	-	ANN; Hybrid ANN	Smart grid and buildings;Short-term electricity load forecasting	• Reviewed the factors influencing the accuracy of ANN-based forecasting models, including model architecture, input combinations, activation functions, and network training algorithms. • Demonstrated the potential of artificial intelligence techniques for short-term electricity load forecasting.
3	Deb et al.	[16]	2017	-	-	-	-	ANN; AutoRegressive Integrated Moving Average model (ARIMA); SVM;Case-Based Reasoning (CBR); Fuzzy time series; Grey prediction model; Moving average and exponential smoothing (MA & ES); K – Nearest Neighbor prediction method (kNN); Hybrid models	Time series forecasting	• Reviewed the applied research on nine major time series forecasting techniques and qualitatively analyzed their advantages and disadvantages.
4	Yildiz et al.	[29]	2017	-	-	√	√	Regression Trees (RT); SVM; ANN	Commercial building	• Reviewed the different applications of ANN, SVM, and especially regression modeling in electrical load forecasting for commercial buildings. • Summarized the most commonly used regression variables and methods to improve model performance and accuracy.
5	Kuster et al.	[30]	2017	-	-	√	-	ANN; SVM; Timeseries analysis; Regression	Electrical load forecasting	• Emphasized the potential correlation between scenario parameters, such as time horizon and/or input variables, and a particular model, aiming to determine which predictive model was best suited to specific situations and variables. • Proposed a classification methodology for predictive model selection.
6	Amasyali and El-Gohary	[17]	2018	-	√	√	-	SVM; ANN; DT; Other statistical algorithms	Building energy consumption	• Reviewed research on data-driven predictive modeling of building energy consumption in terms of the scope of prediction, data attributes, data preprocessing methods, machine learning algorithms, and performance measures used for evaluation. • Identified limitations of the proposed data-driven algorithms, including poor generalization and poor interpretability.
7	Wei et al.	[18]	2018	-	-	√	-	ANN, SVM, statistical regression, DT; Genetic Algorithm(GA); Clustering	Building energy consumption	• Reviewed data-driven methods for predicting and classifying building energy analyses at different prototypes and granularities, and developed a systematic overview of different aspects of each method, such as rationale, current R&D status, latest practical applications, and potential future developments. • Recommended modifying the framework of different data-driven methods according to building characteristics and implementing multi-objective forecasting.

(continued on next page)

Table 1 (*continued*)

No.	Author(s)	Ref.	Year	Data Preparation		Feature Engineering		Models	Application	Comments
				source	methods	type	methods			
8	Ahmad et al.	[31]	2018	-	-	-	-	Clustering, Statistical regression, ANN, SVM	Building energy consumption	• Focused on data-driven methods for predicting building energy demand and energy forecasting methods for large-scale buildings. • Conducted a comprehensive study of building energy analysis for different building types, such as industrial, commercial, and residential, in urban and rural environments, exploring the need for analyzing the energy performance of buildings at both urban and rural levels.
9	Runge and Zmeureanu	[13]	2019	-	-	-	-	ANN	Building energy consumption	• Reviewed research on artificial neural networks for predicting energy use and demand in buildings since 2000. • Provided a focused review of applications, predictive models, and performance metrics used in model evaluation.
10	Bourdeau et al.	[19]	2019	√	√	√	-	Autoregressive models; Statistical regressions; Classification-based methods; KNN; DT; SVM; ANN; Ensemble models	Building energy consumption	• Reviewed six single techniques and two combined approaches in building energy modeling and prediction techniques. • Classified input feature selection methods into supervised and unsupervised learning. • Focused on the impact of input data characteristics (i.e., source, input type, time series time step, data volume, and training-validation-testing ratios) and preprocessing methods on prediction accuracy. • Suggested that occupant behavior led to significant modeling uncertainty and should not be reduced to theoretical occupancy scenarios.
11	Sun et al.	[32]	2020	-	-	√	√	Linear regression (LR); Time series analysis; RT; SVM; ANN; DNN; RNN; CNN	Building energy consumption	• Proposed an updated multi-step building energy forecasting strategy. • Reviewed feature engineering, data-driven modeling, and prediction outputs in building energy prediction within the data-driven process.
12	Zhang et al.	[21]	2021	-	√	√	√	LR; SVM; NN; Deep Learning; Tree-based Algorithms; Hybrid; Auto-regressive methods; fuzzy timeseries model	Building energy consumption	• Reviewed the application of machine learning techniques in building load prediction within the context of a logical architecture for machine learning, focusing on using a performance metric P and performing a task T based on empirical learning (E). • Suggested that, compared to the algorithmic side, the data aspect was still not well studied and showed weak model generalization, highlighting the need for high-quality test data to provide a relatively standard criterion for assessing algorithm performance. • Noted the increasing need for load forecasting for multiple buildings or building clusters, emphasizing the interdependence of energy loads in building clusters and the unique challenges in the load forecasting framework under building grid integration.

(continued on next page)

Table 1 (*continued*)

No.	Author(s)	Ref.	Year	Data Preparation		Feature Engineering		Models	Application	Comments
				source	methods	type	methods			
13	Zhu et al.	[33]	2022	√	√	√	√	Multiple regression, ARMA, DT, FNN, RNN	Integrated energy systems	• Focused the research objective on integrated energy systems (IES) and revealed the uniqueness of the IES load forecasting problem compared to conventional power system problems. • Reviewed the fundamental issues of data-driven techniques in current research, including variable decision-making, data preparation, feature engineering, and model identification. • Proposed automated machine learning as an effective way to address load forecasting problems in complex industrial systems.

Note: '√' means the literature includes the corresponding contents, while '-' means exclusion.

methods, and the forefront issues in building load prediction. It also focuses on reviewing existing research on data processing methods and feature engineering. The discussion concludes with an analysis of the review and potential avenues for future research. Within this framework, the review will concentrate on the following essential inquiries:

1) How can data-driven methods more accurately predict building loads? How are data-driven building load prediction methods applied in different scenarios? What benefits and drawbacks do these methodologies offer in practical implementations?
2) What strategies can be employed in data processing and feature selection to boost the dependability of forecasts for building loads?
3) Given the swift advancements in large language models, how to leverage their powerful feature representation learning capability, few-shot learning capability, and generalization capability to capture more hidden information from multi-dimensional energy usage data, thereby improving the precision and dependability of forecasts for building loads?

By thoroughly investigating these issues, this review seeks to offer those studying building load prediction a detailed and thorough insight, aiming to direct subsequent research and practical applications. The remainder of this review is as follows. Section 2 outlines the application scenarios and general forecasting process for data-driven building load prediction. Section 3 reviews past research from various aspects including building types, time scales, prediction objectives, load prediction models, and feature types. Section 4 reviews previous studies concerning the present situation of data on building energy usage, data preprocessing methods, data fusion methods, and transfer learning. Section 5 focuses on feature engineering within historical research on building load prediction, exploring feature types, feature construction, feature selection, feature extraction, and sensitivity analysis. Finally, Section 6 discusses the proposed LLM-based framework for building load prediction, and Section 7 summarizes the findings.

2. Applications and process of data-driven load forecasting

Essential for building design and operations, building load forecasting underpins efficient energy management and occupant comfort. During the design phase, designers and engineers need accurate load information for rational system capacity planning, equipment selection, and energy distribution. Similarly, in the operational phase, building systems rely on dynamic adjustments based on real-time load conditions to ensure efficient use of energy and meet user demands. In large commercial complexes, accurate load forecasting can support energy demand response strategies. By adjusting equipment operation, buildings can adapt to fluctuations in the energy market, achieving cost minimization and sustainable operation of the system.

2.1. Importance of building load prediction in building energy systems

Building energy systems refer to the integrated technologies within a building that manage its energy consumption. These systems typically include HVAC, lighting, and renewable energy sources like solar panels. The goal is to optimize energy use and enhance operational efficiency. In the context of building lifecycle management, load forecasting plays a crucial role in supporting building energy system management and achieving energy savings. By accurately predicting energy demand, it enables better system design, dynamic operational adjustments, and participation in demand response programs, thus optimizing energy use throughout the building's lifecycle.

System planning and energy efficiency in the design phase: Load forecasting aids engineers in understanding anticipated energy consumption patterns, guiding the selection and sizing of equipment across the full load spectrum. Oversized HVAC systems cause unnecessary energy consumption, while undersized systems fail to ensure occupant comfort. Load forecasting enables the correct sizing of equipment, reducing energy waste and ensuring operational efficiency. Designers typically rely on historical data and technical specifications for long-term load forecasting, ensuring the building's overall energy planning aligns with efficiency and comfort goals.

Real-time system control and optimization in the operational phase: In the operational phase, load forecasting is essential for dynamic control and system optimization. Building energy management systems use real-time load forecasts, considering factors like occupant density, outdoor weather, and indoor thermal comfort needs, to adjust system operations dynamically. Accurate load predictions help optimize equipment control strategies, improving energy efficiency. Additionally, combining forecast data with real-time operational data allows for proactive maintenance scheduling and fault diagnosis, ensuring reliable and energy-efficient system performance.

Demand response: Load forecasting can not only optimize the operation of HVAC systems but also participate in grid demand response programs by predicting short-term load variations. By flexibly adjusting their energy usage behavior during energy price fluctuations, they can minimize operating costs and achieve sustainable system operations. With the development of renewable energy, the power sources on the supply side of the grid have become increasingly complex and diverse. In addition to traditional coal-fired power, renewable energy sources such as solar and wind power are greatly affected by time and weather, resulting in less stable power generation. The integration of renewable energy into the grid will pose new challenges for balancing supply and demand. This requires more accurate and detailed load forecasting on an hourly or daily basis on the demand side, rather than merely forecasting the total energy consumption on a monthly or yearly basis.

The challenges of load forecasting for building energy systems are as follows

1) **Building data quality and availability:** The data generated by building energy systems may be incomplete or noisy, increasing the difficulty of prediction. Ensuring access to high-quality, real-time data is essential to reduce prediction errors and uncertainties.
2) **Complexity and nonlinearity of prediction scenarios:** Energy consumption patterns are influenced by multiple factors, including weather, occupant behavior, and building characteristics. The interaction among these factors adds complexity to the model, requiring advanced algorithms (e.g., deep learning) to effectively capture these nonlinear relationships.
3) **Model scalability:** A significant challenge is how to effectively use historical data from similar buildings when specific historical data for a given building is lacking. Enhancing the generalization ability of models to extract useful information from similar buildings and make appropriate adjustments is necessary to improve prediction accuracy.
4) **Real-time system integration:** Efficient integration of real-time data streams is crucial, but technical challenges may arise in inputting sensor data and weather forecasts into load prediction models promptly. Furthermore, ensuring the system dynamically adjusts HVAC operation strategies based on load prediction results to enhance energy efficiency is a complex issue.

Although load prediction models play a critical role in building energy system management, data gaps, inaccurate sensors, or changing building usage patterns can affect model performance. Understanding the characteristics of different application scenarios for building load prediction and exploring prediction methods used in various data types, prediction time scales, prediction goals, building scales, and building types is crucial.

2.2. Different scenarios for load forecasting

Building load forecasting can be applied at multiple levels and scenarios, depending on the type of data required, the time scale of the forecast, the goal of the forecast, the size of the building, and the different needs of the building type.

A key challenge in building load forecasting is data availability and quality, which can vary significantly depending on the specific application scenario. In the design phase [34], load forecasting typically relies on historical data from similar buildings, technical specifications, and building simulation models to predict future energy demands. In contrast, during the operational phase, real-time and high-frequency data inputs are necessary [35], such as:

- **Weather conditions:** Real-time outdoor temperature, humidity, and solar radiation.
- **Building occupancy:** Sensors tracking the number of occupants and their movement within the building.
- **Equipment operational status:** Data from HVAC systems, lighting, and other electrical systems.
- **Indoor environmental parameters:** Temperature, humidity, and CO_2 levels within the building.

The complexity of managing these diverse data sources makes data preprocessing and fusion crucial steps in building an accurate and reliable prediction model.

The objects of load prediction usually include office buildings [36], residential buildings [37], commercial buildings [38], educational buildings [39], industrial buildings [40], etc. Depending on the prediction target, load prediction are typically divided into cooling [34] /heating [41] load, HVAC electricity load [42], and building electricity load [43]. Cooling/heating load refers to the energy required to maintain a comfortable indoor temperature. This is typically calculated based on the building's heat transfer, internal heat sources (such as occupants and equipment), and external environmental factors (such as solar radiation and outdoor temperature). Accurate cooling/heating loads prediction is essential for energy-efficient design, device selection, and intelligent operation and maintenance, enabling the optimal configuration of heating and cooling devices and improving efficiency. HVAC electricity load describes the energy used by an HVAC system during operation, including the electricity required for cooling/heating loads, as well as for ventilation, air filtration, and humidity control. Accurate HVAC electricity consumption prediction helps optimize system efficiency, develop operation strategies, and predict potential equipment failures, enhancing system reliability. Building electricity load encompasses the total electricity usage of a building, covering not only the electricity consumption of the HVAC system, but also other power-using equipment, such as lighting, equipment and household appliances. It is primarily used for energy procurement planning and grid load management. Accurate prediction of building electricity consumption allows for better balancing of energy supply and demand, reducing peak load pressure, and supporting overall energy management strategies.

The time scale is a pivotal factor in building load forecasting, influencing the construction of prediction models. Building load prediction can be classified as short-term or long-term depending on the timescale. Short-term load forecasting (STLF) [44–47], generally ranging from hours to a week, relies on real-time, such as weather changes and current load patterns, crucial for energy system optimization and demand response control objectives. In contrast, long-term load forecasting (LTLF) [38,48–50], which is usually based on monthly and annual forecasts, focuses more on seasonal trends and historical data analysis, impacting long-term energy management.

Building load prediction can be addressed at various levels of complexity, ranging from individual buildings [38] to regions [46] and even countries [51]. As the scale increases, capturing the wider influence of factors becomes more challenging. Predictions for individual buildings focus on load changes within a single building, requiring an accurate understanding and modeling of the characteristics of each building, such as design, usage patterns, and equipment efficiency. Predictions for building groups [50] study the overall load of a group of adjacent or interconnected buildings, considering interactions between buildings and the integration of various energy systems. Regional predictions involve a broader range of buildings and facilities within a larger area, requiring handling complex datasets and considering the diversity of different building types and uses. These apply to urban planning, regional energy distribution, and demand response strategies.

Fig. 1 shows the data-driven building load prediction framework, which is a multi-step systematic approach.

1) **Data collection:** Acquiring data from building management systems, HVAC sensors, weather stations, and occupancy sensors to form a comprehensive dataset.
2) **Data preparation:** Ensuring data quality by handling missing values, noise reduction, and outlier elimination, which are critical for developing a robust predictive model for building energy systems.
3) **Feature engineering:** Identifying and selecting key variables that significantly affect building energy consumption, such as occupancy, weather, and building system operational statuses, and transforming these into input variables for the prediction model.
4) **Algorithm selection and model training:** Choosing the most appropriate algorithm based on the building's energy system

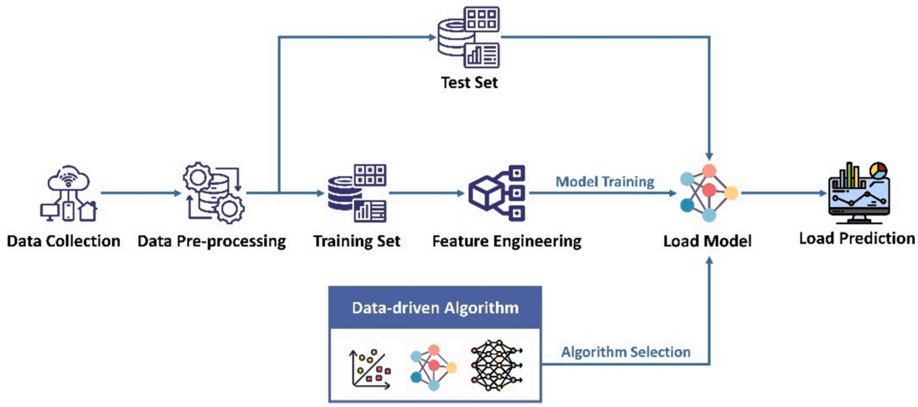


Fig. 1. Data-driven framework for building load prediction.

requirements, adjusting model hyperparameters, and optimizing the model to ensure accurate load forecasting.

3. Data preparation

The quality of data serves as the cornerstone of research in building load prediction. This has a direct influence on the precision of the predictions and the generalizability of the resulting models.

3.1. Current state of building energy consumption data

Automation, smart systems, and big data storage have streamlined building energy data collection. It falls into specific energy consumption data and data describing factors affecting energy consumption (energy consumption influencing variables).

Data collected on-site accurately represents the actual energy usage of buildings. However, due to issues with metering devices and data transmission, the quality of collected data is not high, presenting various problems such as anomalies, breakpoints, noise, etc., which have always been difficult to overcome. Besides poor accessibility, datasets and energy consumption monitoring platform data suffer from poor data quality (commonly containing missing and anomalous values) and low richness (few static data corresponding to buildings), leading to underutilization of data.

Research institutions or government departments have established various building energy consumption datasets. Table 2 compares the information of several publicly available building energy consumption datasets.

3.2. Data preprocessing methods

The primary purpose of preprocessing is to improve data quality,

Table 2
Publicly available building energy consumption datasets.

Dataset	Dataset Source	Description	Building Type	Survey Content	Spatial Scale
CER Smart Metering Project [52]	Irish Social Science Data Archive	Smart electricity meter customer behavior test project (CBT) containing half-hourly meter data for 5000 Irish residences and SMEs, used to assess the impact on consumer electricity usage	Residential and Small/Medium Enterprises	–	Smart meter data
Reference Energy Disaggregation Data Set (REDD) [53]	Massachusetts Institute of Technology	Used for energy disaggregation research (identifying contributions of individual appliances from a composite electrical signal), includes consumption data for 6 households over 18 days in Spring 2011.	Residential Buildings	-	Electricity meter, Appliance
Commercial Buildings Energy Consumption Survey (CBECS) [54]	U.S. Department of Energy and Environmental Protection Agency	Updated every four years since 1979, covering the entire United States. The latest survey in 2018 shows that about 6 million commercial buildings have been surveyed.	Commercial Buildings	Building information, equipment information	Regional, Single-building
National Non-Domestic Building Stock (NDBS) [55]	UK Department for Environment and Transport	Integrates real data at the level of individual buildings, detailing energy use and efficiency, geometry and materials of each non-residential building, occupant activities, and even the potential for renewable energy generation, to create a complete profile of the non-domestic buildings in England and Wales.	Non-residential buildings, including industrial buildings	Building physical characteristics, building geometric dimensions, main equipment overview	Single-building
Global Energy Forecasting Competition [56-58]	Dr. Tao Hong's team	GEFCOM held in 2012, 2014, and 2017, focusing on short-term load forecasting through hourly load and outdoor temperature	Distribution Area	Hourly meteorological data (temperature) and load data, holiday information	Distribution area
Building performance database (BPD) [59]	Lawrence Berkeley National Lab	The comprehensive dataset of energy-related data for commercial and residential buildings in the United States	Building energy consumption; Regional energy consumption	building type, location, physical characteristics.	Regional, Single-building
The Building Data Genome Project [60,61]	Clayton Miller et al.	Energy consumption data for buildings with various functions such as offices, schools, residences, and medical facilities in countries like the USA and UK	Building energy consumption	The whole building's electricity meter data	Single-building

ensuring that the prediction model can effectively learn and generalize. Relevant research indicates that data preprocessing consumes over 80 % of the workload in the entire prediction process [62], making it one of the most important tasks [63–65].

Common data preprocessing methods for HVAC system load forecasting include data cleaning, transformation, time-series processing, dimensionality reduction, and augmentation.

Data cleaning includes missing value handling and anomaly detection. These challenges arise due to factors like sensor malfunctions, communication errors, or recording mishaps, leading to incomplete or inaccurate building energy consumption records. For handling missing values, various imputation methods are available. Single-variable imputations include techniques such as mean imputation, forward filling, backward filling, linear and polynomial interpolation, Kalman filtering, and moving averages. Multivariate data gaps may be addressed using methods like K-Nearest Neighbors (KNN), Random Forest, Multiple Singular Spectral Analysis, and Matrix Factorization. Choosing the right imputation technique hinges on the extent of the data gap and the distinct features of the method itself. Cho et al. [66] used Normalized Root Mean Square Error (RMSE) as a principal metric to gauge the performance of imputation methods, analyzing both the precision and the computational demands of six different data imputation approaches. The results indicated that the impact of an imputation technique is primarily affected by the size of the missing data gap. Methods for anomaly detection include: 1) distance-based methods, such as KNN; 2) statistical methods, like box plots and standard deviation methods; 3) decision tree-based methods [67]. Chen et al. [68] proposed using wavelet analysis algorithms for data filtering during data preprocessing, proving that wavelet analysis methods have good applicability and high efficiency in handling noise data in big data.

Data transformation typically refers to data normalization/standardization. This is important for most machine learning methods, especially those sensitive to the variable scale. Additionally, feature encoding of non-numeric features (like categorical variables) is also a type of data transformation. Common feature encoding methods include One-Hot Encoding and Label Encoding. Fan et al. [65] developed the Symbolic Aggregate Approximation, transforming original time series Building Automation System (BAS) data into meaningful symbol streams to reduce data volume.

Time series analysis is a powerful tool for building performance analysis. Analyzing time series data encompassing building environment, energy consumption, and operational information, allows for tasks like identifying patterns of energy usage, detecting anomalies in building operations, and ultimately optimizing building performance. Analysis of time-series parameters often requires processing time labels and converting timestamps into multiple features to capture time periodicity. Additionally, In addition, historical data points (such as load data from previous hours or days) need to be considered to obtain the dynamic characteristics [69].

The objective of reducing data dimensionality is to minimize the dataset's dimensions while preserving as much of the original information as possible. This process not only boosts computational efficiency but also improves the effectiveness of machine learning algorithms by eliminating features that do not contribute to discrimination. Various methods for reducing data dimensionality encompass Principal Component Analysis (PCA) and Kernel Principal Component Analysis (KPCA). Xuemei et al. [70] utilized PCA and KPCA to reduce data dimensions and evaluated the performance of SVM in conjunction with PCA, SVM with KPCA, and SVM without employing any dimension reduction strategies.

3.3. Data fusion methods

Data fusion technologies integrate data from diverse sources, creating a richer and more comprehensive information view compared to relying on single sources. This data augmentation method is also applicable to situations of data insufficiency. Developing schemes to acquire data and designing appropriate information fusion strategies based on emphasized objectives become important needs [71]. Himeur et al. [72] described data fusion strategies and potential applications in building energy-saving systems from different perspectives such as data fusion levels and techniques.

Data fusion techniques can be categorized into three main types based on the processing stage: Data-level Fusion, Feature-level Fusion, and Decision-level Fusion.

Data-level Fusion is the most basic level of data fusion, where raw sensor data is directly merged to create a comprehensive dataset [69]. This is typically applicable to sensor data or image data. Since fusion usually occurs before other higher-level data processing, it retains the most original information. Li [73] addressed the challenge of improving the robustness of load prediction by leveraging a Kalman filter for data fusion. This approach integrates multi-modal data, leading to more reliable prediction results. Sha et al. [74] integrated building morphology, onsite test energy data, and simulated energy data to predict energy for the target building based on similar building data in the absence of historical data for the target building.

Feature-level fusion, distinguishing itself from data-level fusion, entails the extraction of relevant features from raw data through specific fusion algorithms. This method is generally more efficient because it processes smaller amounts of data. Himeur [75] explored the complementary features of different household appliances and proposed an effective feature extraction strategy. This technique, centered on the fusion of time-domain features, was designed to enhance the ability to discriminate between features. Wang et al. [76] applied different combinations of features, i.e., miscellaneous electrical loads, lighting, occupants count, and number of WiFi connections based on the LSTM prediction algorithm to find out which combination of features provides the most accurate prediction.

As the highest level, decision-level fusion integrates output decisions (such as classification results and prediction results) from different models. This method is common in multi-expert systems, where each system makes independent decisions, which are then merged through some strategy (such as voting, weighted averaging, etc.) to leverage the advantages of multiple independent decisions, enhancing the accuracy and reliability of the final decision. Fotopoulou et al. [77] collected data from diverse sensor modes, mapped them to an entropy semantic model, and applied multiple data fusion strategies to devise effective energy-saving actions. To enhance building energy management within BIM environments, Xiao et al. [78] proposed an ontology-based semantic retrieval method. This method facilitates the integration and retrieval of energy consumption information based on semantic relationships, improving information accessibility within BIM.

Depending on the data processing approach, data fusion algorithms are classified into statistical-based algorithms, mathematical model-based algorithms, and machine learning-based algorithms.

Statistical-based fusion algorithms do not rely heavily on a physical or theoretical understanding of the data generation process but focus on the characteristics presented by the data itself, such as mean, variance, correlation coefficients, etc. For instance, in multi-sensor systems, the weighted averaging method can be used to fuse readings from different sensors, where weights are based on the reliability or accuracy of each sensor. Huang et al. [79] fused the directly measured frozen water flow and supply/return water temperature difference with the indirectly measured chiller energy consumption and evaporation/condensation temperature, obtaining more accurate cooling load prediction results, and based on this, optimized the control of the unit. Djuric et al. [80] applied a similar approach to heat pump performance assessment, By fusing temperature and pressure measurements with power signal data, they achieved more reliable performance evaluation compared to using individual data sources. Huang et al. [81] demonstrated that when there is surplus measurement, the fusion method based on multiple sensors is superior to the model-based fusion method based on the load data of chiller units.

Mathematical model-based fusion algorithms rely on predefined mathematical models to perform data fusion. Common algorithms include the Kalman Filter [73] and its variants, maximum likelihood estimation, Bayesian methods, etc. In building energy prediction, sensors can be prone to transmission failures due to network or circuit problems. Additionally, individual sensors may not provide entirely reliable measurements. To address these limitations, the Kalman Filter, a Bayesian method, is widely used for sensor data fusion. This filter effectively combines data from multiple sensors, reducing overall uncertainty in the final measurement and leading to more accurate predictions. Kumar et al. [82] crafted an infrared (IR) proximity sensor model utilizing a probabilistic approach, which accurately assesses the uncertainty and constraints of individual sensors. By integrating information from various sensors within a Bayesian framework, they were able to generate a 3D occupancy profile of objects within a robot's workspace.

Machine learning-based fusion algorithms include decision trees, support vector machines, etc. In the era of deep learning, especially CNN and RNN have shown outstanding performance. These methods are suitable for complex data fusion tasks, especially when the relationship between models and data is difficult to describe using traditional mathematical models. These algorithms excel at handling complex data relationships but require substantial training data. Additionally, their internal decision-making processes can be less interpretable compared to model-based methods, potentially hindering the understanding of their results In the HVAC field, research on data fusion based on machine learning algorithms is still relatively scarce.

3.4. Transfer learning

Transfer learning [83] enables the application of knowledge acquired from addressing one problem to a related, yet distinct, new challenge. This essentially fast-tracks the learning process for new tasks by leveraging existing data and models. This means that features from the source task can be applied to the target task, thereby improving its learning efficiency and performance. Transfer learning is particularly valuable when there is insufficient data for the target task.

In building load forecasting, a key advantage of transfer learning is its effectiveness in solving the problem of data insufficiency. For new buildings or buildings that have not undergone long-term energy consumption monitoring, there frequently exists a shortage of ample historical data needed to train prediction models accurately. In such cases, transfer learning allows us to leverage rich data accumulated in other buildings or environments. By transferring knowledge from these related but different tasks, the load prediction capability for specific buildings or buildings with little data can be significantly enhanced. Mocanu et al. [84] utilized cross-building transfer learning for unsupervised energy consumption prediction, achieving knowledge transfer from commercial to residential buildings and from residential buildings with static electricity pricing to those with time-based electricity pricing. Zhou et al. [85] based on BiGAN for data augmentation, implemented effective load prediction for residential and commercial users under data scarcity conditions using transfer learning technology. Fang et al. [86] used LSTM to obtain temporal features and employed Domain Adversarial Neural Networks (DANN) to extract domain-invariant features for predicting cross-building energy consumption, performing prediction for office buildings with few historical measurement data under different energy consumption characteristics and climate conditions.

The geographical location of a building significantly impacts its energy demand. Climate conditions and seasonal patterns differ across regions, and these factors also affect a building's energy use. The cross-geographical transfer allows the model to adapt to energy demand patterns under different climates and environmental conditions, even with very limited data in the new region. Ribeiro et al. [87] used seasonally and trend-adjusted transfer learning to forecast energy use across four schools spread throughout Newfoundland, Canada, to enhance the precision of prediction for new buildings. Compared to a model that utilized only a month's data from the target school, the prediction accuracy using data from multiple schools increased by 11.2 %.

It is evident that applying transfer learning in building load prediction, particularly in addressing data scarcity and geographic diversity, can significantly enhance the efficacy and applicability of prediction models.

Data preprocessing converts data into a format better suited for machine learning models, essential for enhancing the models' generalizability and accuracy. Table 3 summarizes the previously discussed data preprocessing methods.

4. Feature engineering

Feature engineering aims to automatically or semi-automatically create, select, and transform features from raw data to enhance model performance. It is essential for data-driven building load prediction[90]. Extracting key information from the large volume of data is crucial for reducing model complexity, improving accuracy, and enhancing the model's generalization capability [91]. Furthermore, well-designed features can provide a more intuitive understanding and help explain the behavior and prediction outcomes of the model. In the face of high-dimensional data, feature engineering can help mitigate the curse of dimensionality.

Feature engineering typically includes feature construction, feature selection, and feature extraction. It is worth noting that feature engineering frequently entails an iterative process, applying various methods repeatedly until achieving satisfactory outcomes. Consequently, it is quite common for a variety of feature engineering strategies to be employed in tandem within a single study to improve the model's effectiveness.

4.1. Feature types classification

Input features for building load forecasting can broadly be categorized into four categories: meteorological conditions, building physical characteristics, indoor environmental information, and indoor occupants' behaviour [21].

Building air conditioning energy consumption is heavily influenced by a range of outdoor meteorological parameters. Temperature, wind speed, relative humidity, and other outdoor weather parameters are universally recognized as key factors. This is reflected in the widespread use of weather data as input features for HVAC load prediction in virtually all studies [92].

Building physical characteristics mainly include features such as the heat transfer coefficient of the building envelope, total building height, wall area, roof area, area receiving daylight, building orientation, the ratio of window area, wall area, shading coefficient, and more [17].

Building load is also significantly impacted by the number of occupants and their behaviour [32], which includes both the use of energy-consuming devices and general occupancy patterns. Due to the challenges in directly measuring these factors, studies often use proxies such as time indicators [93] (hour, day, workdays or weekends, holidays or non-holidays) to infer occupancy levels and activity patterns [94].

Buildings have similar energy use characteristics under the same day type, and under the influence of building thermal inertia, previous time step weather conditions, and energy consumption may affect current energy consumption. Therefore, historical energy consumption data frequently serves as an important input feature. Fan et al. [95] used the past 24 h of load, temperature, and relative humidity to predict the next day of building cooling load curve.

Table 3
Summary of data preprocessing methods.

Method	Description	Application scenarios	Advantages	Key Steps	References
Data cleaning	Remove errors or outliers from data, fill in missing values	Historical load data	Improve data quality, reduce noise impact	Missing value filling, outlier detection	[64-68]
Data transformation	Transform raw data, such as normalization or standardization	Original load data	Improve model convergence speed, reduce feature scale differences	Normalization/Standardization, feature encoding	[42,64,65]
Time-series Processing	Processing time labels	Time series load data	Capture periodicity and trends in data	Seasonal decomposition, trend smoothing, time label processing	[65,69,88]
Data dimensionality reduction	Reduce the dimensions of data while retaining key information	Multidimensional load data	Improve model training efficiency, reduce overfitting risk	PCA, KPCA	[70]
Data fusion and augmentation	Use various methods to increase the data's diversity	Multi-source data; data scarcity	Improve model's generalizability, enhance data representativeness	Kalman filter, Maximum likelihood estimation, Bayesian methods, Correlation coefficients, DT, SVM	[69,73-82,89]
Transfer learning	Migrating knowledge across domains	Insufficient data in the target task	Improve model's generalizability	Relationship knowledge transfer, data transfer	[84-87]

4.2. Feature construction

Feature construction is based on existing features in the original data, where new features are manually created through combinations, transformations, or derivations to enhance the model's ability to represent the data. Its most significant characteristic is that it relies on a deep understanding of the data and the problem, and the construction of new features is based on experience and domain knowledge. Therefore, this process requires researchers to spend time observing the raw data, considering the potential form of the problem and the data structure, and demands a high level of expertise in the relevant predictive task. Feature construction typically includes operations such as addition, subtraction, multiplication, and division of existing features, creating interaction terms (e.g., products or ratios of features), or generating lagged or cumulative features for time series data. Unlike feature extraction, new features are created through explicit mathematical or logical relationships rather than dimensionality reduction or compression of the original feature space. For example, in the field of load forecasting, based on the features "total building energy consumption" and "total building area," a new feature "energy consumption per unit area" can be constructed by division.

By manually combining, transforming, or deriving features, it is possible to capture the latent non-linear or interaction relationships in the original data, thereby enhancing the model's ability to represent complex data. Newly constructed features can help the model better adapt to different scenarios, improving its generalization capability on new data. Especially when non-linear relationships exist between features, constructing more meaningful features can significantly improve the model's predictive performance. However, when there are many data dimensions or the data is complex, feature construction often requires a great deal of time and effort. This method also demands extensive domain knowledge to create meaningful features, as poorly constructed features may introduce noise or mislead the model. Therefore, it heavily relies on the researcher's domain expertise. When the relationship between the constructed features and the original features is unclear, the model may become difficult to interpret.

Given the purpose, methods, and nature of feature construction, it is the preferred method when existing features cannot fully express the underlying patterns in the data, and domain expertise can guide the creation of meaningful and complex features.

To address the challenge of short-term prediction for metropolitan-scale electric load, Chu et al. [96] developed an integrated algorithm. This algorithm leverages a new load decomposition method that separates the electric load data into base load and weather-sensitive load. This decomposition allows the model to focus on the weather-dependent component, ultimately improving prediction accuracy. Lu et al. [40] used Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose raw data into multiple smoothed datasets, discussing the impact of data properties and sliding window length on decomposition results.

4.3. Feature selection

Feature selection identifies the most pertinent features from the original dataset, acting as a vital step in machine learning to trim down feature count, boost model precision, mitigate overfitting, and improve the model's comprehensibility. The primary approaches to feature selection are categorized into three groups: Filter Methods, which select features based on their statistical characteristics independent of any model; Wrapper Methods, which assess subsets of features based on their effectiveness as determined by a specific model's performance; and Embedded Methods, which uses algorithms that inherently perform feature selection during the learning process.

Filter methods identify optimal features by statistically evaluating the statistical relationship between each feature in the dataset and the target variable. Techniques like the Chi-squared test, ANOVA [97], and correlation coefficients [98] use univariate statistical tests to compute the statistical association of each feature with the outcome, thus those with the highest scores. Mutual Information assesses the mutual dependence between individual features and the target variable. Features with a higher degree of mutual information are considered more relevant and are selected for further analysis. Hamed Chitsaz et al. [99] applied Mutual Information to eliminate irrelevant and redundant candidate features. Their objective was to identify and select the most informative inputs for enhancing the accuracy and efficiency of the prediction process.

Wrapper methods approach feature selection as an optimization problem. They evaluate different combinations of features by training a machine learning model on each combination and measuring its performance. Commonly used Recursive Feature Elimination (RFE) progressively builds a model and removes features with lesser contributions, thereby selecting the most important feature subset. Sequential feature selection algorithms (like forward selection, backward elimination, and stepwise selection) add or remove features step-by-step based on model performance. Genetic algorithms employ strategies similar to biological evolution to search for the optimal combination of features. Li et al. [100] leveraged embedded recursive feature elimination for feature selection, subsequently predicting short-term electricity consumption in buildings.

Embedded methods assess the importance of each feature during training and prioritize those that contribute most to the prediction task. This eliminates the need for a separate feature selection step,

Table 4
Summary of different feature engineering methods.

Method	Suitable Scenarios	Advantage	Disadvantage	Type	Algorithm	Ref.	Main contribution
Feature Construction	• Existing features inadequate to express fundamental data patterns • Domain expertise available to guide creation of meaningful, complex features.	• Enhances model's generalization ability on new data • Improves model accuracy by creating more informative features • Allows capturing non-linear, interactive, and multivariate relationships	• Complex and time-consuming • Requires domain expertise to ensure meaningful feature creation • make the model difficult to interpret when the relationship with original features is not clear	combination, decomposition, transformation, or aggregation	-	[96]	• Effectively divides the electric load into base load and weather-sensitive load.
						[40]	• Uses Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose original data into multiple smooth datasets.
Feature Selection	• High-dimensional datasets requiring high model interpretability and focused analysis on key features; • Essential to reduce overfitting and simplify model without changing original features.	• Reduces the number of features and reduces overfitting • Increases model interpretability	• Over-aggressive feature selection may lead to loss of important information • Choosing the right feature selection method requires some experimentation and adjustment	Filter Methods	Chi-squared test, ANOVA, correlation coefficients, mutual information	[98]	• Uses Kendall t-rank coefficient to calculate the correlation between input variables and target variables.
						[97]	• Uses Design of Experiments and Analysis of Variance methods to identify important features.
						[99]	• Filters out irrelevant and redundant candidate features based on mutual information.
				Wrapper Methods	RFE, GA, SteepwiseFS, ReliefF algorithm	[100]	• Uses embedded RFE for feature selection.
						[112]	• Uses GA to optimize the feature set of these models.
						[113]	• Finds StepwiseFS Algorithm to be the best for feature selection.
				Embedded Methods	Lasso regression,tree-based models	[36,103,114]	• Uses Lasso Regression for feature selection.
						[103]	• Uses RF to select key features for heating load prediction.
						[101]	• Uses C5.0 Decision Tree analysis method to generate clear classification rules composed of important contextual features.
Feature Extraction	• Complex and numerous features with unclear meanings; • Effective to extract representative features through compression and mapping.	• Reduces data dimensions, lowering computational cost • Remove noise from data, and extract more informative features • Extracted to more representative features in high-dimensional and complex datasets	• Dimensionality reduction may lead to information loss, especially when high-dimensional data contains important information • Features extracted via complex mathematical transformations or neural networks, challenging to intuitively interpret their physical meaning or relationship to original data.	Linear feature extraction methods	PCA, LDA	[70]	• Analyzes the impact of PCA on load forecasting.
						[115]	• Uses RF and PCA to select feasible input features.
				Non-linear feature extraction methods	t-SNE, clustering methods, wavelet transforms, Autoencoders	[39]	• Compares feature extraction effects using GAN, Autoencoders, and 1D convolutional Autoencoders, with the same number of features.
						[108]	• Uses efficient sparse Autoencoders as a feature extraction method.
						[95]	• Uses four feature extraction methods, proving that unsupervised deep learning models have the best feature extraction effect.

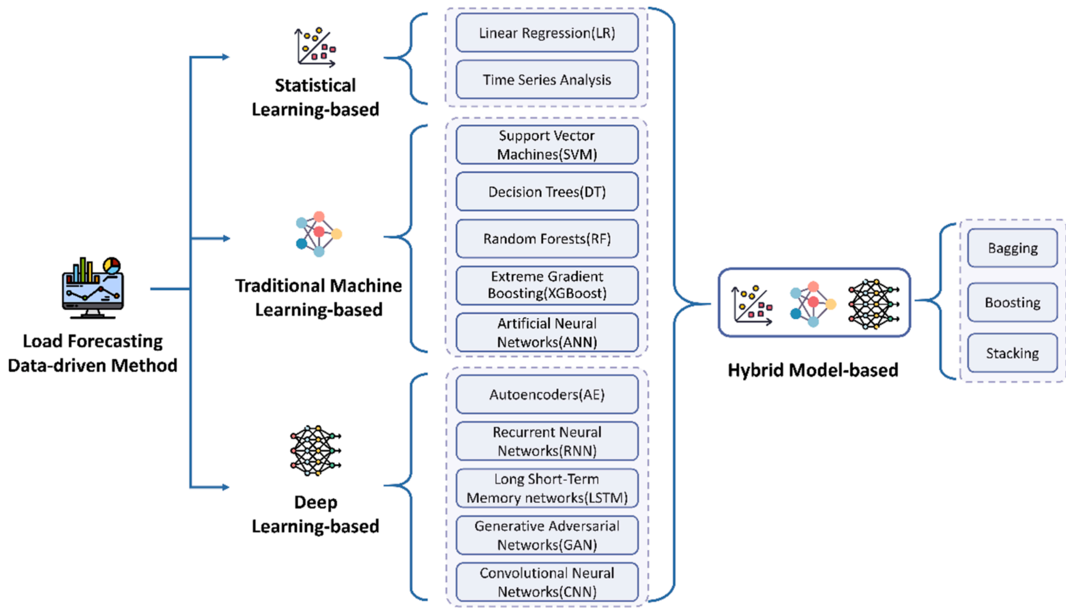


Fig. 2. Overview of load forecasting data-driven modeling.

streamlining the workflow. Lasso regression [36] is a popular choice due to its ability to eliminate irrelevant features by driving their coefficients to zero during model training. Tree-based models [101] achieve embedded selection by prioritizing features that best separate data points at each decision node. Tree-based models not only provide excellent predictive performance but also yield feature importance. Jain [102] in his research on predictive methods for multi-family housing, determined the optimal subset of exogenous features using Lasso regression. Yuan et al. [103] employed RF for feature selection in predicting heating energy consumption (top ten features).

In addition, some works have employed sensitivity analysis to pinpoint which input variables are critical factors in building energy models. Local sensitivity analysis focuses on the model's response to the output at a particular input point, assessing the extent to which the model's output changes when one or more of the input parameters are varied slightly around their baseline values. Global sensitivity analyses consider the variation of input parameters over their entire range of values and aim to understand which parameters are most influential in the change in output over all possible input values.

By utilizing Local, Morris, and Sobol methods for sensitivity analysis, Kristensen et al. [104] identified critical parameters in their dynamic building energy model and found that Floor area, U-value (roof), Infiltration rate, Heating set point, and U-value (floor) were the most influential input parameters on energy consumption. An interesting study by Hygh et al. [34] utilized the Monte Carlo framework to analyze a comprehensive set of 27 architectural design parameters. This approach allowed them to identify the most critical factors influencing building energy performance during the crucial early design stages. Ascione et al. [105] conducted a sensitivity index evaluation of 58 input parameters to optimize the input of ANN.

4.4. Feature extraction

Unlike feature selection, which retains the information of the original features, feature extraction uses mathematical methods or deep learning models to map the original features into a new lower-dimensional space, thereby reducing the dimensionality and preserving the most informative features. The newly generated features are often implicit transformations of the original ones, created through compression or mapping. These features may not have a direct mathematical relationship with the original features and are usually difficult to interpret, which is also a key distinction from feature construction. Common linear feature extraction methods, which aim to find uncorrelated and informative directions in the data, include Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for maximizing class separation. Non-linear methods, suitable for complex relationships, include techniques like t-distributed Stochastic Neighbor Embedding (t-SNE), clustering methods, wavelet transforms, Autoencoders, etc.

PCA takes the original data and transforms it into a new coordinate system through linear transformation, where the first principal component of any projection of the data has the maximum variance. LDA is a supervised learning technique for dimensionality reduction, aiming not only to find the components that best represent the data but also those that best distinguish between categories.

Li et al. [70] compared the performance of PCA and Kernel PCA for predicting building loads, finding that Kernel PCA outperforms regular PCA in performance without feature selection. PCA's application [106,107] requires extensive observational data, which is often lacking in building operational data due to project operation periods and data collection quality issues.

Autoencoders use a neural network for data encoding, attempting to capture the main information of the data using fewer encoding dimensions. Autoencoders are particularly useful in scenarios where the intrinsic structure of the data needs to be understood or when reducing the dimensionality for further processing or visualization. However, as the feature space used for energy consumption prediction is typically small [32], Autoencoders are not as commonly used as other methods for predicting building energy use. Mujeeb and Javaid [108] used an Autoencoder to compress the data and then fed this compressed data into a non-linear autoregressive network, to predict future load. To improve building energy use prediction, Luo et al. [109] used k-means to group similar weather patterns throughout the year and applied feature extraction techniques and GA to create a deep learning model that adapts its predictions based on the weather.

Table 4 presents a range of feature engineering methods suitable for building load prediction. Understanding the strengths and weaknesses of each method is crucial for selecting the most effective approach for a specific research question. It's essential to carefully choose the right feature engineering techniques after a detailed analysis of data properties, research objectives, and other relevant contextual elements. Moreover, it's noteworthy that feature engineering often involves iteratively applying multiple methods until satisfactory results are achieved [110]. To achieve accurate short-term thermal load predictions, Yan et al. [110] combined wavelet decomposition and reconstruction for in-depth data analysis, correlation analysis for feature selection, and PCA for dimensionality reduction to obtain a refined set of model inputs, potentially leading to improved prediction accuracy and efficiency. Wen [111] used the Cen-CK-means clustering, Bayesian, and BiLSTM to comprehensively consider the global spatial characteristics and local temporal characteristics of electricity consumption data to forecast short-term load in the context of large amounts of electricity consumption.

In practical applications, specific steps and methods of feature engineering vary according to data types, the nature of the problem, and the models used. An effective feature engineering process can significantly enhance a model's predictive capability and interpretability, being one of the key steps in constructing efficient machine learning models. It's important to recognize that data leakage is a critical issue to be aware of in selecting input features, as using information not available at the time of prediction can lead to artificially high prediction scores.

5. Data-driven load forecasting methods

Data-driven building load prediction encompasses a spectrum of modeling complexities, ranging from statistical learning methods to deep learning methods. Additionally, hybrid approaches combine techniques from different categories. Fig. 2. summarizes the data-driven methods commonly used in building load prediction.

5.1. Statistical learning methods

Statistical learning-based methods are characterized by low computational complexity, less data requirements, and strong interpretability. However, their capability to handle non-linear complex relationships is limited. Within statistical learning methods for building load prediction, linear regression (LR) [116] plays a prominent role. Additionally, time series analysis techniques are valuable tools for capturing patterns and trends in historical load data. To support designers in optimizing energy performance during the early design phase, Hygh et al. [34] developed a multivariate linear regression (MLR) model. This model considers twenty-seven key building parameters, enabling designers to identify factors that significantly influence energy consumption. Guo et al. [36] developed a demand-side electric load prediction model using extreme learning machines and MLR. While LR models are simple and offer quick predictions, their linear nature limits their ability to capture the complex interactions between factors like weather and occupant behavior that significantly influence HVAC loads.

Time series analysis, relying solely on the historical values of the series for future value prediction, falls into the category of univariate prediction models. Representative algorithms include autoregressive (AR) models, known for their effectiveness in capturing short-term trends, moving average (MA) models, useful for handling random fluctuations, and autoregressive integrated moving average (ARIMA) models, capable of modeling complex seasonal patterns. Leveraging ARIMA and ARMA models, Chujai et al. [88] investigated the prediction of household electricity consumption. Their work also explored identifying the most suitable forecasting period. To enhance the precision of STLF, Wang et al. [117] employed the seasonal ARIMA model, the PSO-optimized Fourier method, and a combination of both in the Northwest China Power Grid. This approach was used to refine the predictions of the seasonal ARIMA model, thereby assisting power generators and consumers in making more informed and rational planning and decision-making.

While univariate prediction models offer a simplified approach, they may not capture the full complexity of building load prediction. Multivariate models, which consider the relationships between multiple variables and the target variable, often achieve higher accuracy by incorporating a richer set of information. Multivariate models use a combination of input variables like time series features (e.g., date type, schedules), meteorological conditions (e.g., temperature), and physical building parameters (e.g., wall area, wall thermal transmittance) to predict outputs such as heating/cooling/ electricity loads [36,118].

5.2. Traditional machine learning methods

Traditional machine learning methods are suitable for handling nonlinear relationships and often provide higher prediction accuracy compared to statistical learning-based methods. Common models include Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN) [119].

Support Vector Machines (SVM) solve nonlinear regression problems by introducing kernel functions. These functions transform data to a higher-dimensional space, enabling the efficient creation of a linear hyperplane for regression purposes. This high-dimensional mapping empowers SVMs to learn complex patterns within the data, even when those patterns are not linearly separable in the original feature space. Its higher predictive accuracy and speed make it more suitable for STLF. Amasyali et al. [120] employed SVM, using daily average sky coverage and day types to forecast the lighting load of an office building in Philadelphia, Pennsylvania. Vrablecova et al. [121] predicted loads using SVR based on smart meter data from homes across Ireland, demonstrated the applicability of SVR for STLF, and concluded that SVR was suitable for predicting aggregated loads of individual buildings and building groups. Chen et al. [122] employed SVR to forecast the hour hourly electricity demand of hotels and shopping centers, achieving errors of 4.0 % and 6.0 % respectively, which allows for optimized energy use and cost savings.

Decision trees (DT) construct decision trees based on the feature partitioning of data and generate interpretable structures. Yu et al [37] suggested the application of decision trees (DTs) to increase model interpretability, with their clear tree structure providing valuable insights compared to less interpretable methods such as regression and ANN. Hong et al. [43] crafted a DT model aimed at diminishing electrical energy usage in elementary school facilities, and combinate GA, ANN, and MRA to enhance the precision of predictions.

Random Forests (RF) employ a bagging strategy, constructing numerous trees to generate averaged predictions, thus mitigating overfitting issues. By randomly creating each decision tree with various features and datasets, and facilitating parallel training, RF models achieve superior predictive accuracy compared to individual trees. Pham et al. [123] applied RF that predicts the load of multiple buildings on an hourly basis, validating the model's effectiveness with five hourly building energy consumption datasets over a year. Wang et al. [124] employed an RF-based method for integrated prediction. RF also assigns importance to features, facilitating the selection of main features and skipping weaker ones to speed up the computational process. Lahouar and Slama [51] combined the RF model with expert feature selection to predict day-ahead building load. This analysis identified day-ahead load, day type, and temperature as the most important influencing factors.

Extreme Gradient Boosting (XGBoost), an ensemble learning algorithm, optimizes weak learners into strong ones using gradient boosting algorithms. Its built-in regularization controls model complexity and reduces overfitting, thus producing superior predictions. Unlike RF, which trains decision trees in parallel, XGBoost uses sequential boosting. This sequential strategy helps improve the accuracy of XGBoost. Focusing on long-term building load prediction, Wang et al. [125] employed XGBoost with five key factors(day type, time, holidays, and weather). Their findings suggest XGBoost's superiority over SVM, RF, and LSTM. Lu et al. [40] utilized the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method for data

Table 5
The application of various data-driven models in different scenarios.

Model Category	Reference	Model	Application phase	Time Scale	Building Type	Spatial Scale	Prediction Target	Feature	Research Purpose
Linear Regression (LR)	[34]	MLR	Design phase	LTLF	Office	Building	heating energy, cooling energy, total energy	27 design parameters	• For assessing the energy efficiency in design stages, assisting designers in decision-making during conceptual design.
	[36]	Extreme learning machine, MLR,SVR,BPNN.	Operational phase	STLF	Office	Building	40 minutes ahead heating energy demand prediction	hour, outdoor temperature, returning GSHP system water temperature, supplied GSHP system water temperature, outdoor wind speed, indoor temperature, current energy demand	• Obtaining building thermal response time for forecasting model lead time and predicting building heating load.
	[96]	LR	Operational phase	STLF	-	Metropolitan	Power load of workdays	total power load, seasonal attributes, hourly seasonal attributes	• SVM for forecasting weather-sensitive components and short-term metropolitan-scale power load.
Time-series Analysis	[88]	ARIMA, ARMA	Operational phase	STLF	Residential	Building	Electric power consumption	historical electric power	• Predicting household electricity consumption and finding the most suitable forecasting period.
	[117]	Seasonal ARIMA	Operational phase	STLF	-	Northwest China Regional Power Grid	Electricity demand	historical electric demand	• Residual correction of seasonal ARIMA predictions to enhance the accuracy of the model.
Support Vector Machine (SVM)	[120]	SVM	Operational phase	STLF	Office	Building	Daily Building lighting energy consumption	day type, average daily sky coverage	• Predicting lighting energy consumption.
	[121]	SVR	Operational phase	STLF	-	City	Power load of next half-hour	power load of last day	• Demonstrating the applicability of online SVR for STLF.
	[122]	SVR	Operational phase	STLF	Office	Building	Short-term electrical load	the outdoor temperature recorded two hours prior before the DR event	• Calculating demand response baseline.
Decision Tree (DT)	[37]	DT	Operational phase	LTLF	Residential	Building	energy use intensity (EUI)	climatic conditions building characteristics, household characteristics, household appliance energy sources	• Generating prediction models with interpretable tree-structured flowchart
	[43]	decision support model	Operational phase	LTLF	Educational	Building Group	Electric load	location, building, inhabitants.	• Reducing electrical load in primary school facilities.
Random Forest (RF)	[123]	RF	Operational phase	STLF	Office	Building	Hourly energy consumption	Hourly energy consumption	• STLF for multiple buildings on an hourly basis.
	[124]	RF	operational phase	STLF	Educational	Building	Hourly building energy prediction	meteorological, occupancy, time-related data	• Validating the feasibility of RF in STLF.
	[51]	RF	Operational phase	STLF	-	Country	Day-ahead electricity load	month number, day type, minimum temperature, maximum temperature, load before 24 h, load before 48 h	• Improving input through expert feature selection for day-ahead electricity load prediction.
Extreme Gradient Boosting (XGBoost)	[125]	XGBoost	Operational phase	LTLF	Educational	Building	Building cooling load	relative humidity, outdoor air temperature, time, weekday, holiday	• Discussing the influence of input uncertainty on prediction accuracy, enhancing model robustness.
	[40]	XGBoost	Operational phase	STLF	Industrial	Building	daily energy consumption	daily energy consumption	• STLF using CEEMDAN.
	[48]	XGBoost	Operational phase	LTLF	Residential	Building	cooling energy of air conditioners	daily energy use, daily maximum outdoor temperature, running time, rated power input	• Providing references for implementing residential building energy management for different groups.
	[41]	XGBoost	Operational phase	STLF	Residential	Building	hourly heating energy consumption	historical heating load, wind speed, temperature, relative humidity, solar radiation, time s	• Evaluating the effectiveness of XGBoost, RF, ANN, GBDT, and SVR models in the aspect of accuracy, interpretability, robustness, and efficiency; proving XGBoost as the most effective.

(*continued on next page*)

Table 5 (*continued*)

Model Category	Reference	Model	Application phase	Time Scale	Building Type	Spatial Scale	Prediction Target	Feature	Research Purpose
Artificial Neural Network (ANN)	[42]	RF, ANN	operational phase	STLF	Hotel	Building	hourly HVAC electricity load	temperature, wind speed, relative humidity, historical HVAC electricity load, numbers, rooms booked	• Comparing the performance of FBNN and RF.
	[35]	neural networks	operational phase	STLF	Bioclimatic	Building	Hourly electricity demand	day type, hour, temperature, solar radiation, state of the equipment, electric load	• STLF for bioclimatic buildings.
	[126]	Levenberg–Marquardt ANN	operational phase	STLF	Educational	Building	day-ahead load	climatic data, hour/day, operation schedule, electrical load	• Using occupancy zone electrical load to simulate occupant activities and predict cooling load.
	[49]	PSO - RBF	operational phase	LTLF	-	Country	monthly electricity consumption data	monthly electricity consumption data	• Rapidly and effectively predicting building energy consumption values based on a small amount of historical data.
	[132]	SARIMA, BPNN	operational phase	STLF	-	Region	Hourly electricity load data	hourly electricity load data for weekdays, season, period	• Applying wavelet denoising for short-term regional electricity load prediction.
	[50]	FFBPNN, RBFN, ANFIS	operational phase	LTLF	University campus	Building Group	heating energy consumption	temperature, heating load, day type, relative humidity, solar radiation, month	• Integrating various neural networks to improve prediction accuracy.
Deep learning-based methods	[127]	LSTM	operational phase	STLF	University library	Building	hourly and daily load of HVAC	historical daily load, historical cooling load	• More accurate prediction of HVAC load.
	[76]	LSTM	operational phase	STLF	Office	Building	Heat load	miscellaneous electric loads, occupant counts, WiFi counts	• Select the key features for more accurate prediction of heat gains, enabling predictive HVAC control in buildings.
	[38]	LSTM-based deep recurrent neural network models	operational phase	LTLF	Commercial and residential buildings	Building	Hourly Electric load	weather, schedule, frequency	• LTLF in residential and commercial buildings, supporting decisions of demand response strategies
	[44]	CNN-LSTM	operational phase	STLF	Residential	Building	Electric energy consumption	power load data, sensor data	• Extracting complex spatial and temporal features, enhancing the predictive performance of electrical energy consumption, and analyzing factors with the greatest impact on predicted electricity consumption.
	[45]	RICNN	operational phase	STLF	Distribution industrial complexes	Building	daily electric load	sequence information, weather information, electricity rates, sensor count, historical load	• Calibrating forecast timing and nearby hidden state vectors to enhance short-term electric load forecasting performance.
	[46]	CNN,FTS	operational phase	STLF	-	Region	Hourly Electric load	temperature time series, hourly load data, fuzzified data	• Creating an integrated model that incorporates automated feature extraction to enhance model practicality.
	[47]	RNN, CNN	operational phase	STLF	Commercial	Building	electricity demand	hourly electrical load, outdoor temperature, air pressure, humidity, wind speed (mph)	• predicting day-ahead load in commercial buildings.
	[39]	fully-connected Autoencoders, convolutional Autoencoders, GAN	operational phase	STLF	Educational	Building	cooling load	time variables, outdoor environment variables, historical cooling loads	• Improving building energy prediction performance, automating, and refining the predictive modeling process.

decomposition and used XGBoost to achieve accurate short-term predictions. Lu et al. [33] collected large-scale data from 1,325 air conditioners in Chongqing and established residential building air conditioning cooling energy consumption prediction models, including four ensemble models and two single models. The study concluded that the XGBoost model offered the most accurate predictions. Artificial Neural Networks (ANN) are inspired by the neural networks found in biological systems. They are made up of layers of artificial neurons linked together, which process data using connections that carry weights. This layered and interconnected structure allows ANNs to learn and interpret the relationships between inputs and outputs, although typically for simpler, less complex tasks. ANNs possess a degree of robustness to noise or errors within the data due to their distributed processing nature. It is worth mentioning that the ANNs referred to here are specifically shallow neural networks with a limited number of hidden layers, which restricts their ability to capture highly complex, nonlinear patterns. Ahmad et al. [42] used Feedforward Backpropagation Neural Networks (FFBPNN) and RF to forecast the hourly load of a hotel's HVAC system in Madrid, Spain, demonstrating that ANN's performance was superior to RF. Mena et al. [35] developed and assessed an ANN-based short-term electric demand prediction model for the CIESOL bioclimatic building in southeastern Spain, capable of rapidly predicting outcomes with real data. Leung et al. [126] trained an ANN model using the Levenberg-Marquardt algorithm to predict cooling system energy use in a Hong Kong university building. Their study emphasizes the importance of including energy consumption data as input for improved prediction accuracy.

5.3. Deep learning methods

With increasing data volumes and computational power, data-driven models have gradually evolved from shallow machine learning techniques to more advanced deep learning. This shift to deep learning enables direct feature extraction from data via numerous network layers, facilitating comprehensive end-to-end learning and showcasing potent capabilities in model expression. The precision of deep learning models' predictions is enhanced progressively as the amount of training data increases.

In the domain of predicting building loads, the foremost deep learning algorithms deployed are Autoencoders (AE) [39], Recurrent Neural Networks (RNN) [38], Long Short-Term Memory networks (LSTM) [51], Convolutional Neural Networks (CNN) [44], and Generative Adversarial Networks (GAN) [39]. Rahman et al. [38] employed RNN for the medium and long-term forecasting of electricity usage curves of commercial and residential buildings with an hourly resolution. Zhou et al. [127] utilized the LSTM model for forecasting the load of the HVAC system in a college library located in Guangzhou, demonstrating that LSTM outperformed ARIMA and the Back Propagation Neural Networks (BPNN). Wang et al. [76] used LSTM to predict heating load in office buildings within the United States, achieving lower prediction errors compared to the ASHRAE standard schedule (12 % to 8 % and 26 % to 16 % reduction). Kim et al. [45] developed a Recurrent Initial Convolutional Neural Network (RICNN), using a one-dimensional convolutional initial module to calibrate prediction times and values of hidden state vectors computed from nearby time steps, thereby generating an optimized network to predict short-term electric load. Sadaei et al. [46] employed Fuzzy Time Series (FTS) combined with CNN and devised a method for automatically extracting features for STLF using images created from sequence values of multivariate time series. Cai et al. [47] applied RNN and CNN to forecast the day-ahead load of commercial buildings, proving their superiority over ARIMAX wise accuracy, robustness computational efficiency, and generalization ability.

5.4. Hybrid model-based methods

Generally, data-driven building load prediction relies on a single learning algorithm, leading to instability issues that hinder real-world application. According to the No Free Lunch Theorem [128], there is no single algorithm that universally outperforms all possible scenarios and datasets, each algorithm has its characteristics and applicability, and the single algorithm typically lacks versatility and has limited predictive accuracy. Therefore, hybrid prediction models are proposed to leverage the advantages of each algorithm[129]. They are composed of a group of base models that are trained individually, and the outputs of these models are combined to provide more stable and accurate predictions. Jovanović et al. [50] used ensemble learning technology to combine Feedforward Backpropagation Neural Networks, Radial Basis Function Networks, and Adaptive Neuro-Fuzzy Inference Systems to predict the heating energy usage of a university campus. The study showed that ensemble neural networks could achieve better results.

The main hybrid models include Bagging, Boosting, and Stacking.

Bagging is a technique that reduces the variance of a model by combining multiple models. It involves bootstrap sampling of the original dataset to generate multiple sub-datasets, and then training a model on each sub-dataset. The final result is obtained by calculating the average of all individual model predictions (for regression problems) or the most occurring category (for classification problems). Bagging models have strong parallel processing capabilities and can effectively solve overfitting issues but are sensitive to noisy data. The typical algorithm for Bagging is Random Forest [51,123,124]. Wang et al. [130] explored Ensemble Bagging Trees for building load prediction. Their findings demonstrate the effectiveness of bagging in improving the accuracy of predictions compared to traditional decision trees.

Boosting combines multiple weak learners into a strong learner. It trains models sequentially, with each step attempting to correct the errors of the previous step. New models rely on the performance of previous models, focusing on samples incorrectly classified by previous models, allowing the model to gradually improve performance during iterations. Typical algorithms for Boosting include XGBoost [40,41,48,125], LightGBM, etc. Boosting models usually have good predictive performance and can provide scores for feature importance. However, compared to Bagging, Boosting models are more prone to overfitting and may have longer training times.

Stacking first trains multiple different models and then trains a new model to synthesize these models' outputs. The second layer model (meta-model) is responsible for learning the optimal way to combine the predictions from each base model. Stacking models are flexible, allowing for the selection of different types of models as base and meta models, but their training and tuning processes are complex and have lower interpretability. Fan et al. [131] employed an ensemble model for predicting peak power demand for the next day based on data mining using numerous energy consumption data from Hong Kong's tallest buildings. The model achieved a substantial improvement in accuracy compared to individual base models.

Each method has its unique characteristics and applicable scenarios. Statistical learning-based methods typically emphasize theoretical foundations and model interpretability, traditional machine learning-based methods have advantages in predictive performance and versatility, and deep learning-based methods exhibit outstanding performance in handling unstructured data. Hybrid model-based methods achieve superior accuracy in building load prediction compared to single algorithms by leveraging the strengths of multiple learning techniques and mitigating their weaknesses. Table 5 analyzes the application of various data-driven models in different scenarios from the perspectives of building types, time scales, forecasting objectives, load forecasting models, and feature types.

6. Proposed LLM-based frameworks for building load prediction

With the development of artificial intelligence, Large Language Models (LLMs) have also become an alternative to building load prediction. LLMs, such as GPT-3 [23], PaLM [133], Galactica [134], and

LaMA [135], are characterized by hundreds of billions of parameters or more, having been trained on extensive text datasets. The Transformer architecture is a key factor in the success of LLMs. The effectiveness of these LLMs in downstream applications far overruns the limit of small neural networks, indicating LLMs' exceptional feature representation learning and few-shot learning capabilities [24].

6.1. Motivation and impacts of load prediction with LLM

A crucial aspect of LLMs evaluation lies in testing their ability to adapt to new tasks. Typically, a variety of datasets spanning multiple tasks and domains are utilized to evaluate LLMs in few-shot or zero-shot scenarios [136]. Particularly in fields like healthcare [137–139], education [140–143], and law [144,145], the application of LLMs have sparked significant interest across various research fields. According to the experimental results in the literature [136,140,146,147], LLMs have shown outstanding capabilities in handling general tasks. Notably, GPT-4 has surpassed methods trained on specific datasets in various tasks, including language comprehension, common sense reasoning, and mathematical reasoning. Pre-trained on large-scale corpora, LLMs can extract rich knowledge and serve as domain experts.

Although building load prediction is a relatively traditional application, the introduction of LLM offers unique values and advantages, especially in addressing the complexity and dynamic changes in building energy systems.

1) **Handling complex and unstructured data:** Traditional load prediction models primarily rely on structured data, such as weather, historical energy consumption, and occupant behavior. However, data related to building energy consumption is not always complete or standardized, with much information existing in textual form (such as maintenance reports, sensor logs, and user feedback), which traditional models struggle to fully utilize. LLMs can extract valuable information from a large volume of unstructured data and transform it into usable predictive factors. By incorporating LLMs, unstructured data such as maintenance logs and user feedback can help identify equipment anomalies or changes in usage patterns, thereby enhancing the accuracy of load predictions.
2) **Enabling multi-scenario prediction:** The load patterns of different buildings vary due to differences in building structure, geographic location, climate conditions, and user behavior. Traditional models often have low predictive accuracy when sufficient historical data is lacking. LLMs, with their capabilities in contextual understanding and transfer learning, can draw experiences from different yet similar buildings or scenarios. Through knowledge transfer, predictions can be made based on the experiential data of similar buildings when historical data is limited, enhancing the model's generalization ability.
3) **Integrating multimodal data:** Energy consumption data is often composed of various modes (such as sensor readings, environmental data, and textual reports), and traditional models usually only process structured numerical data, making it difficult to effectively combine data from different modes. The multimodal data processing capability of LLMs allows them to combine text, images, and sensor outputs from various sources to create a more comprehensive energy consumption prediction model. For example, LLMs can combine textual descriptions (such as building usage reports) with real-time sensor data to provide more accurate load predictions.
4) **Simplifying system operations through natural language interaction:** LLMs can also enable natural language interactions between the load prediction system and operators. Operators can query the system in natural language to understand energy consumption

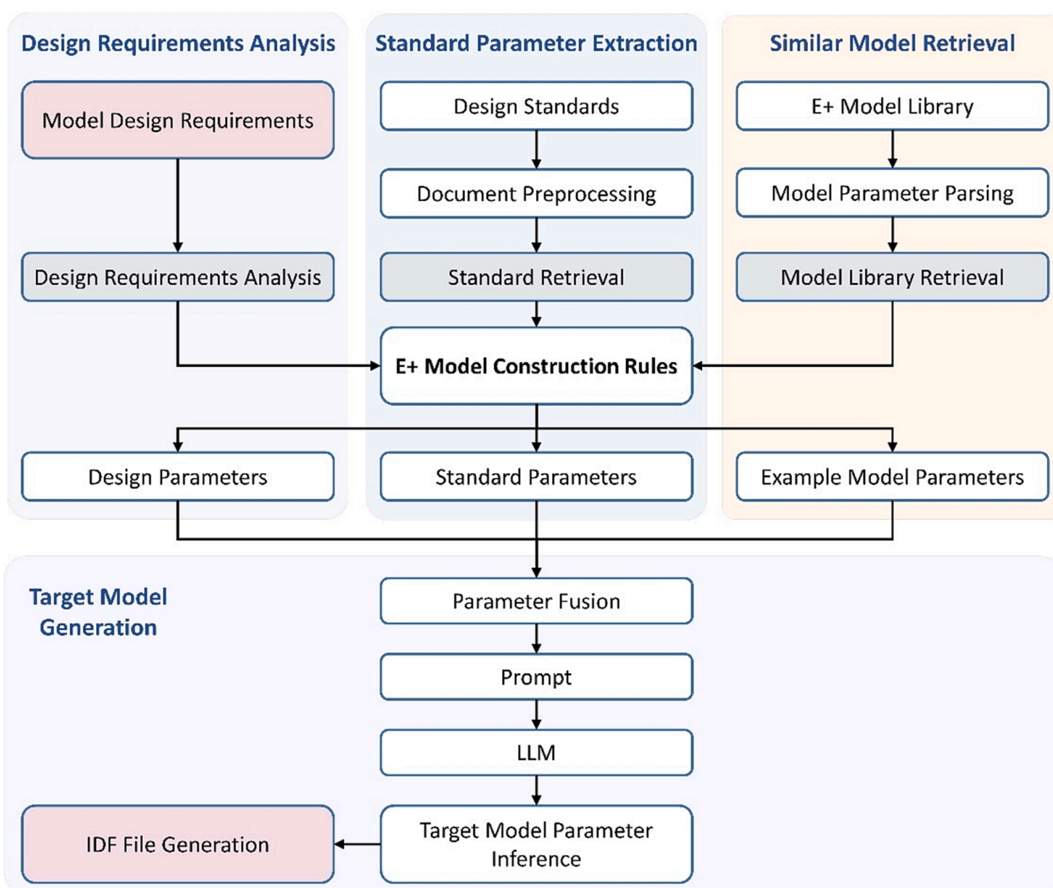


Fig. 3. Technical route for IDF files generation based on Retrieval Augmented Generation technology with LLM.

patterns, anomalies, or optimization suggestions. This makes the load prediction system more operable for non-technical experts, enhances the system's usability and acceptance, and provides building managers with more flexibility in management.

The introduction of LLMs represents not just a technological innovation but also has profound impacts on the industry. Using LLMs for load prediction can significantly enhance the accuracy of energy management and the intelligence of system operations. LLMs improve the accuracy and timeliness of predictions by efficiently processing complex and multi-source data, thereby optimizing energy purchasing and usage strategies, reducing operational costs, and enhancing economic benefits. Additionally, the application of natural language processing technology simplifies the user interface, enabling non-technical personnel to operate the system more conveniently, and enhances the system's responsiveness and flexibility. Moreover, LLMs' powerful feature representation capabilities and few-shot learning ability make it possible to build general load prediction models across different buildings and scenarios, achieve accurate predictions in data-limited environments, and reduce the cost of load prediction technology. These comprehensive impacts not only elevate the technical level of energy load prediction but are also expected to propel the entire industry towards more efficient and intelligent energy management.

While LLMs have excelled in many Natural Language Processing tasks, their application for load prediction is still in its nascent stage. For specific tasks and problems in the HVAC field, the direct use of pre-trained LLMs encounters several challenges [148]. The difficulties in domain-specific tasks frequently arise due to the diverse nature of data, the intricacy of knowledge, the distinctiveness of objectives, and the variety of constraints, such as design standards (e.g., design standards). To tackle these domain-specific challenges, current mature domain-specific methods include External Augmentation, Prompt Engineering [149,150], and Model Fine-tuning. These strategies can be employed either separately or together to enhance performance on tasks tailored to specific domains.

Fine-tuning LLMs typically demand a substantial volume of domain-specific data. Obtaining high-quality, well-labeled training data is difficult in the HVAC field. Effective fine-tuning often necessitates the involvement of domain experts, who can understand the nuances of the field and provide accurate data annotations, but such expertise may be hard to obtain or costly. More importantly, fine-tuning large language models requires powerful computing capabilities, usually relying on high-performance GPUs or TPUs. However, such hardware resources may be difficult to obtain and costly. Given these challenges, many organizations and individuals may opt for Prompt Engineering as a more economical and flexible approach. Carefully designed prompts can effectively direct the model to produce more precise and pertinent responses within a specific domain, without the need for a costly fine-tuning process. Prompt Engineering relies on a deep understanding of how the model responds and can quickly adapt to new needs and changes, making it a practical choice for using large language models within vertical domains.

Considering the actual data conditions in the field of building load prediction, this review proposes a room-scale load prediction framework based on prompt engineering methods using LLM.

6.2. IDF generation based on retrieval augmented generation technology with LLM

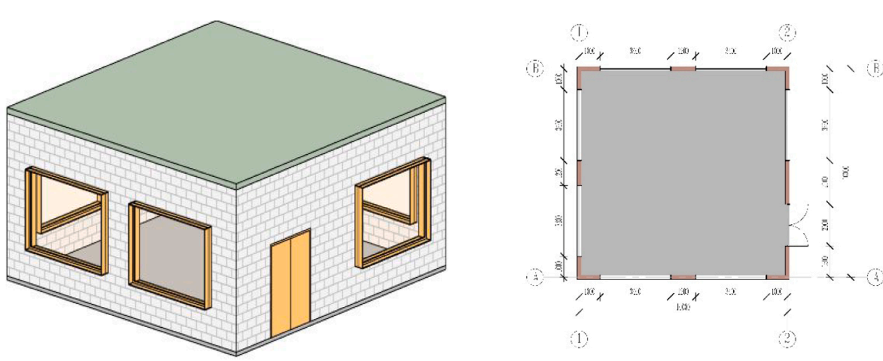
Research works in the building energy field have focused on large spatial scales, such as buildings, building groups, or regions and cities, with relatively less attention given to small spatial scales at the room level. However, building energy on small spatial scales is crucial for building design and optimization, satisfying personalized needs, and refined control. Currently, challenges in room-scale building load prediction include the arduous task of data acquisition, and scarcity of room-level energy consumption data, coupled with the demand for advanced computing resources and data processing capabilities.

Compared to actual data, simulated data is of higher quality, easier to obtain, more granular, and more flexible. In the absence of room-scale load datasets, public building parametric models can be mass-produced through parametric design and generative methods. Large room-scale load datasets can be generated using energy consumption simulation software.

Numerous defaults and approximations are required when constructing EnergyPlus models for actual buildings. The Latin Hypercube Sampling method used in previous studies still necessitates specialists to determine the initial variable set and its selection range, demanding high levels of expertise and coding skills. Therefore, this review proposes a technical route for constructing IDF files based on design standards in the HVAC field and the EnergyPlus model library, utilizing Retrieval Augmented Generation (RAG) technology with LLMs. This approach aims to simplify the EnergyPlus model generation process and enhance the correlation between the EnergyPlus model and the real model. The technical route is illustrated in Fig. 3.

The technical route consists of the following four parts:

1) Design requirements analysis
a) Input building design specifications through natural language, such as energy efficiency standards, building location, building type, window-to-wall ratio, etc.
b) Use natural language processing (NLP) techniques to analyze the input and extract key parameters.
2) Standard parameter extraction



(a) 3D view of the target building (b) Floor plan of the target building

Fig. 4. 3D view and floor plan of the target building.

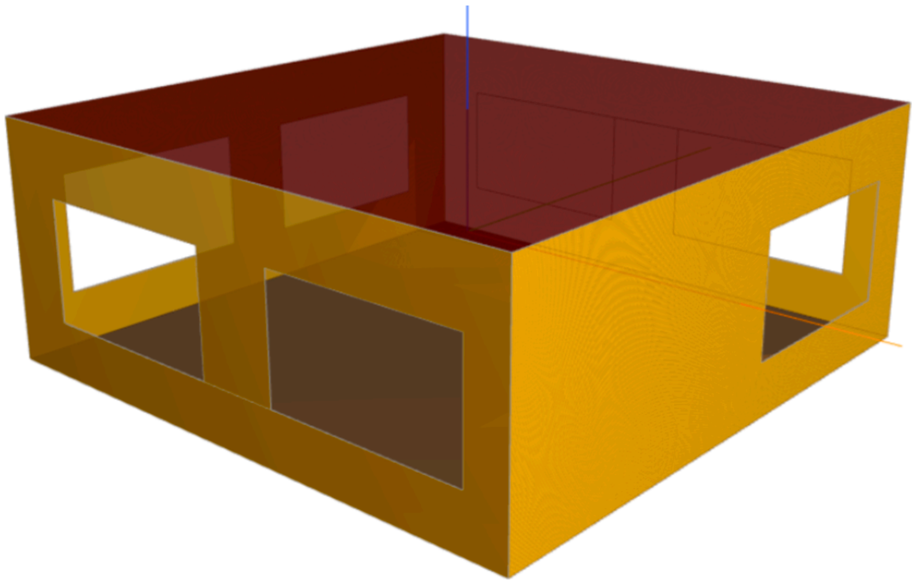


Fig. 5. GBXML information model of the target building.

The office building, located in the vibrant city of **Buffalo**, occupies an area of approximately **100 square meters**. The building features a **compactness ratio of 0.65**, reflecting its efficient use of space. With a **window-to-wall ratio of 0.229**, the structure balances natural light penetration with thermal efficiency, ensuring a comfortable indoor environment while minimizing energy loss.

The external walls are equipped with high-performance thermal insulation, offering a heat transfer coefficient of **0.87 W/(m²·K)** and a specific heat capacity of **75.4 kJ/(m²·K)**. Additionally, the ground floors exhibit strong thermal performance with a U-value of **8.35 W/(m²·K),** contributing to the building's overall energy efficiency.

The roof is insulated with an advanced system, featuring a low U-value of **0.37 W/(m²·K)** to minimize heat transfer and ensure comfortable temperatures year-round. The windows and glazed areas are also optimized for thermal performance, with a U-value of **2.35 W/(m²·K),** allowing natural light while maintaining energy efficiency.

The building is a **one story** structure designed to meet modern energy standards, offering a sustainable and comfortable workspace for its occupants, while blending seamlessly into the urban fabric of Buffalo.

Fig. 6. Requirements for target building simulation in the form of natural language.

a) Employ text mining techniques to obtain necessary design parameters from HVAC design standards, such as minimum energy efficiency standards and maximum allowable window-to-wall ratios.

b) Map user input variables and model construction variables to standard parameters, determining default values or recommended ranges for missing variables used to construct the IDF files.

Table 6
Key parameters and weights used to generate the EnergyPlus model.

Parameter	Value	weight
Building location	Buffalo	0.2
building area	100 m^2	0.05
number of floors	1	0.02
shape coefficient	0.65	0.1
window-to-wall ratio	0.229	0.15
external wall heat transfer coefficient	0.87 W/(m^2·K)	0.13
external wall heat capacity	75.4 kJ/(m^2·K)	0.05
roof heat transfer coefficient	0.37 W/(m^2·K)	0.1
window glass heat transfer coefficient	2.35 W/(m^2·K)	0.15
ground heat transfer coefficient	8.35 W/(m^2·K)	0.05

Table 7
Top 5 best matching models and similarity scores.

No.	Model name	Similarity score
1	OfficeSmall_STD2022_Buffalo	0.6886
2	OfficeHighRise_STD2022_Buffalo	0.6777
3	OfficeMidRise_STD2022_Buffalo	0.6680
4	OfficeMedium_STD2022_NewYork	0.6127
5	OfficeHighRise_STD2022_NewYork	0.6038

3) Similar Model Retrieval
a) Parsing of IDF files from the existing EnergyPlus model library to obtain key and detailed variable parameters.
b) Search the pre-established EnergyPlus model library for cases similar to the target building, such as models with the same location or similar usage.
c) Extract detailed variable parameters from similar models to optimize variable inference.
4) Target IDF file Generation
a) Design LLM prompt templates for generating IDF file. Integrate detailed parameters obtained from design standards and the model library into the prompt, and use LLMs to infer the complete set of target EnergyPlus model parameters.
b) Generate the IDF file based on the parameters inferred by the LLM, providing a complete definition of the building's energy consumption simulation model.

Based on this technical approach, Python was utilized to write an automated script for web scraping to extract standardized parameters. The open standard collection website "Soujianzhu" (https://www.soujianzhu.cn/) was targeted to collect HVAC-related standard content, converting it into plain text format. Most standards were scanned versions without text, requiring recognition via Optical Character Recognition technology. A "fixed-size block" method was employed for text segmentation, followed by vector similarity retrieval to extract text blocks related to load calculation requirements.

For the similarity model retrieval module, given the numerous parameters in IDF files, a key variable set significantly impacting load results and a detailed variable set for constructing a complete EnergyPlus model were identified. The existing EnergyPlus model library was parsed to extract key variables and detailed variables. Key variables included: building location, building type, building area, number of floors, shape coefficient, window-to-wall ratio, external wall heat transfer coefficient, external wall heat capacity, roof heat transfer coefficient, window glass heat transfer coefficient, ground heat transfer coefficient. Detailed variables encompassed ground temperature, external wall materials, window materials, roof materials, shading materials, layer material tables, occupancy schedules, equipment schedules, lighting schedules, infiltration schedules, HVAC cooling schedules, HVAC heating schedules, temperature control schedules, humidity control schedules, and ideal HVAC system settings.

The key variables were used to retrieve historical similar EnergyPlus models from the model library. Corresponding detailed variables were utilized to enhance the prompt for input to the Large Language Model (LLM) for target model parameter inference, thereby constructing IDF files for load calculation.

To validate the proposed method, an exploratory experiment is conducted using Autodesk Revit to construct a single-zone building model as the target building. Located in Buffalo, the building is square-shaped, with 10 m in length and width, and has 7 windows, each measuring 3400 mm × 2000 mm (Fig. 4.). The experiment is based on Meta's open-source Llama-3-8B model.

GBXML (Green Building XML), an XML-based data exchange format, is used to facilitate data exchange between Building Information Modeling (BIM) and building energy performance simulation software, such as EnergyPlus. The BIM model is exported in GBXML format,

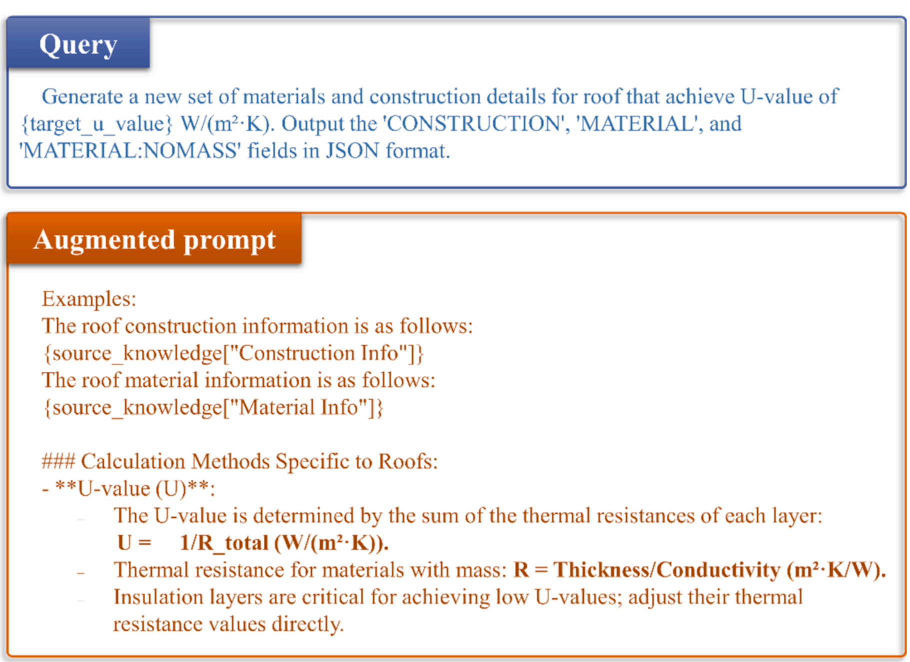


Fig. 7. Requirements for target building simulation in the form of natural language.

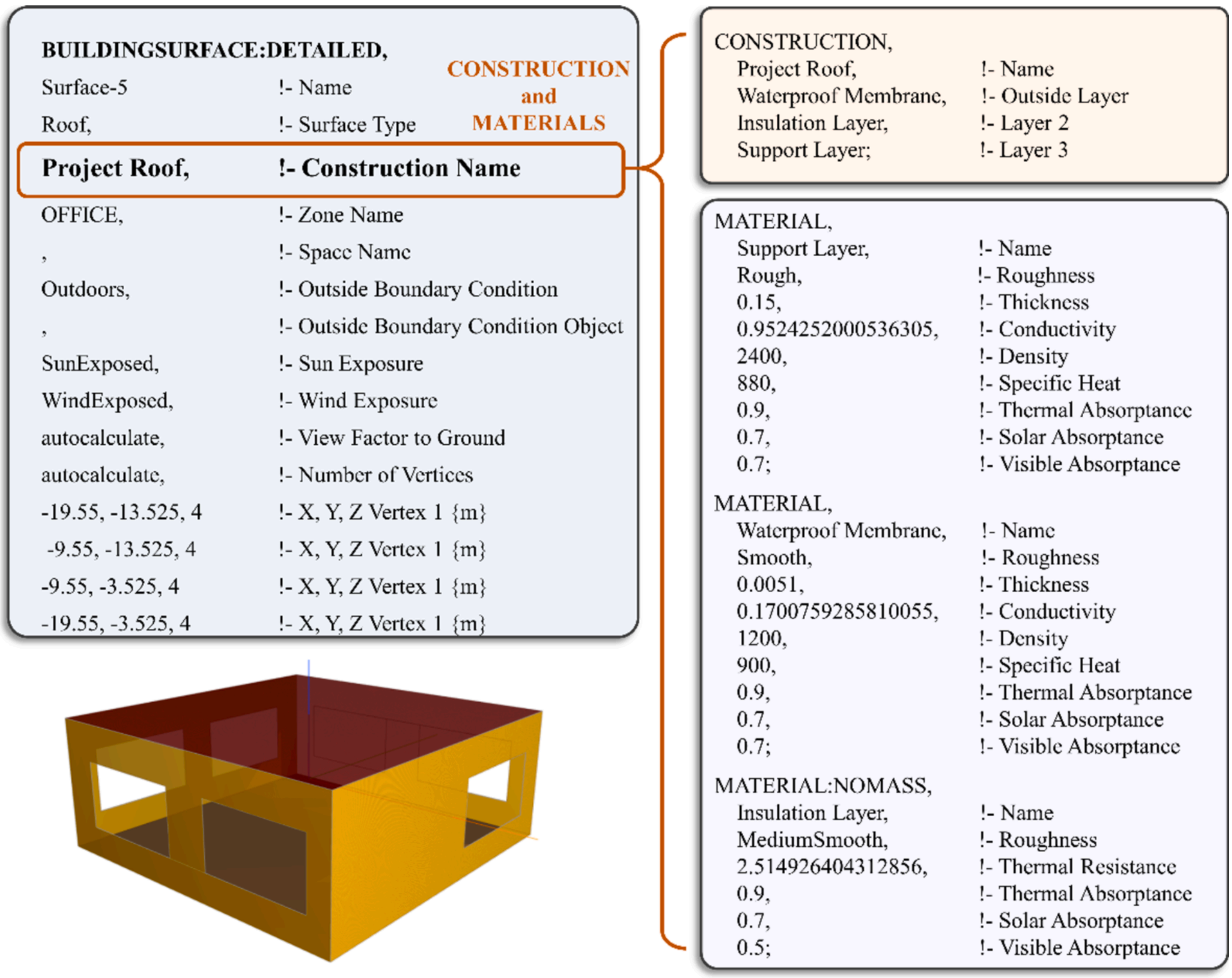


Fig. 8. The generated roof information for the target model.

retaining essential geometric and thermal zone information for further analysis in EnergyPlus (Fig. 5.).

First, the user's simulation requirements, described in natural language, are parsed. These include key variables for retrieving similar models, such as building location and thermal parameters (Fig. 6.). Since the EnergyPlus model library consists only of office buildings, building type is not considered for similarity retrieval. The key parameters and their respective weights for constructing the EnergyPlus model are defined (Table 6).

Most of the features are continuous values, so the weighted Euclidean distance is used as a measure of model similarity. The smaller the distance, the higher the similarity score, indicating greater similarity between the target model and those in the library. Before retrieval, building locations are converted to continuous values based on distance, and all variables are normalized to a range of [0, 1] to eliminate dimensional differences. The top five most similar models in the EnergyPlus library are shown in Table 7, with the closest match being "OfficeSmall_STD2022_Buffalo". Thermal parameters for the building envelope (e.g., walls, windows, roof, and floor) are generated based on this model.

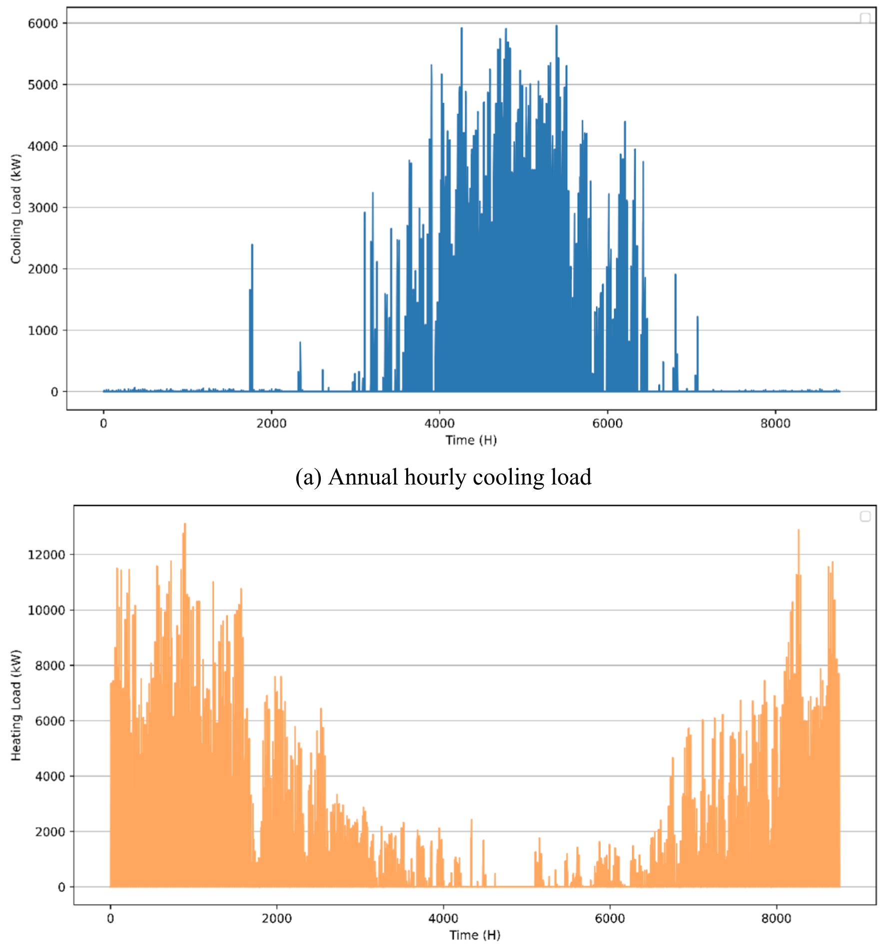
The roof is taken as an example, and prompt is constructed based on the roof materials from similar models to guide the LLM in generating a roof design that meets the target U-value. The RAG method is utilized to combine knowledge retrieved from similar models (source knowledge) with specific design objectives (target U-value), providing the LLM with sufficient contextual information. The specific steps are were follows.

1) Query definition: The task is defined for the LLM to generate new roof materials and construction details that meet the specified U-value; Output is limited to 'CONSTRUCTION', 'MATERIAL', and 'MATERIAL' fields in JSON format.
2) Incorporating information of similar model: Information from the similar model's roof construction is provided as a reference to guide LLM in adjusting and optimizing the new design.
3) Providing key points and calculation methods: Key points affecting the roof's U-value, such as waterproof layers, insulation, and structural support, are outlined; U-value calculation methods and thermal resistance calculations for each material layer are provided to guide LLM's generation process.

The prompt constructed for generating roof materials and construction details is shown in Fig. 7. The generation of roof information for the target model is completed by using this prompt as an input to the LLM (Fig. 8.), which took a total of 196 s. After calculation, the generated roof construction satisfies the target heat transfer coefficients.

The same method is applied to generate thermal parameters for other building envelope components, such as walls, windows, and floors. Prompts are constructed for each component, and a complete EnergyPlus IDF file is successfully generated, containing detailed geometric information and thermal parameters. The IDF file is used for annual energy simulations in EnergyPlus, with a focus on hourly heating and cooling loads.

The simulation results show that the total annual energy consumption is 8681 kWh, of which 2,380 kWh is for cooling and 6,301 kWh for



(a) Annual hourly cooling load

(b) Annual hourly heating load

Fig. 9. Hourly energy consumption distribution of target building.

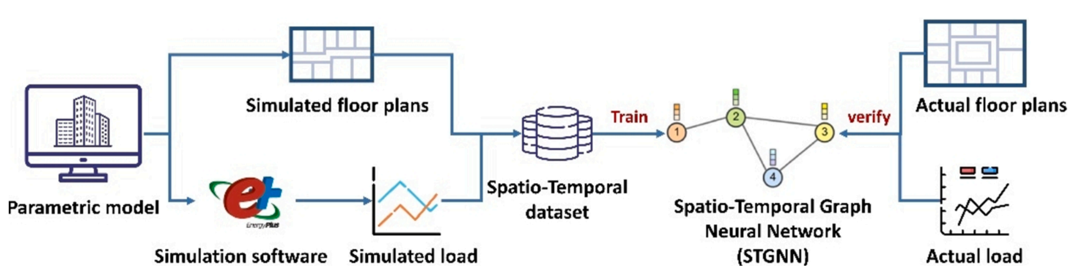


Fig. 10. Data-driven room-scale load prediction technology framework.

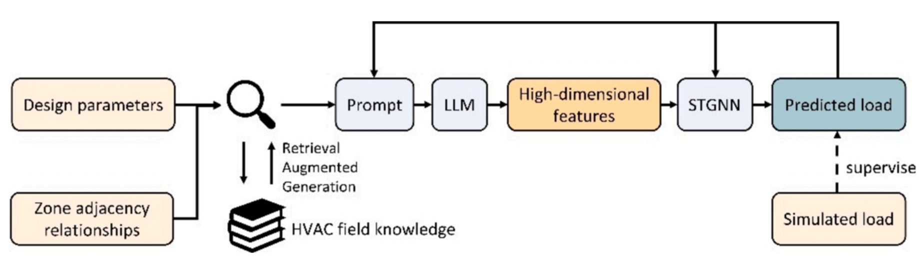


Fig. 11. Room-scale load prediction technology framework based on LLM

heating. The maximum cooling load is 5956 W (62 W/m^2), and the maximum heating load is 13112 W (137 W/m^2). As in Fig. 9., The hourly energy consumption distribution show that cooling loads are higher in the summer months (June-August) and heating loads dominate in the winter months (December-February).

This data experiment demonstrates the feasibility of automatically generating EnergyPlus models that meet specific requirements using the proposed method. The generated models accurately reflect the building's energy performance, validating the approach of using RAG and LLM for generating EnergyPlus model. This method reduces the manual effort in model creation, improves modeling efficiency, and facilitates the efficient conversion from user-defined natural language requirements to energy simulation models. It provides a new approach to building energy analysis, lowering the requirements for domain expertise and coding skills. However, this is an exploratory experiment with a limited EnergyPlus model library for similarity retrieval. Future work could expand the model library and optimize the key variables and weight settings used for retrieval to further improve the accuracy of model generation.

While reducing the threshold of EnergyPlus model construction, the method can quickly generate a large number of room-scale load prediction results that match the actual building, providing a data base for constructing room-scale load prediction models in Section 6.3.

6.3. Room-scale load forecasting technique based on LLM

The method proposed in Section 6.2 enables the rapid supplementation of energy consumption information when the building's geometric information is complete. However, to achieve room-level load prediction, a large amount of building geometric data that matches actual buildings is still required. Parametric and generative design in architecture allows for the quick generation of numerous design schemes by easily adjusting various parameters, facilitating the mass production of building mass models. To meet the data volume requirements for constructing room-scale load prediction models, a large number of building floor plans can be generated in batches using parametric design methods. These plans can then be used to extract building geometric information, the adjacency relationships of rooms/thermal zones, and the functions of rooms.

The simulated floor plans are spatially structured data, which have four notable characteristics: high non-linearity, adjacency relation maintenance, node attribute maintenance, and sparsity/topology A good practice to represent the floor plans is undirected graph, with vertices representing spaces and edges representing relationships between spaces [151,152]. Since room-scale building load prediction requires the model to understand the spatial relationships between rooms and effectively capture the dependencies of spatio-temporal data, a deep learning method called Spatio-Temporal Graph Neural Network (STGNN) can be used [153]. A previous study has proposed ST-GNN-based aggregated load prediction method to enhance the accuracy of load predictions [154].

To approximate real usage scenarios, actual architectural drawings and load calculation results should be used to validate the model. The proposed framework for data-driven room-scale load prediction technology, as shown in Fig. 10., enables an "end-to-end" load prediction process using architectural drawings as input, enhancing the accuracy, simplifying the process, improving the practicality of the model, and lowering the professional threshold for load prediction.

For room-scale load prediction, it is crucial to integrate specific knowledge with datasets of load prediction. Utilizing the powerful feature representation learning, few-shot learning, and generalization capabilities of LLMs, more hidden information can be captured from load prediction data. Understanding and encoding complex data relationships, and building a comprehensive data representation is key. This foundation supports the accurate prediction of spatio-temporal graph neural network models. Using the high-dimensional features refined by LLM, training STGNN models for load prediction, and comparing prediction results with simulated load results, can further optimize the Prompt and STGNN network structures. The technology framework for LLM-based room-scale load prediction, as shown in Fig. 11., enables an "end-to-end" load prediction process.

The experimental validation of the technical approach in this section needs to be carried out in the future after the method in Section 6.2 is further refined and large-scale building load simulation data is generated in batches.

7. Conclusion

Analyzing building energy consumption is a pivotal aspect of demand-side management in addressing building energy requirements. Accurate prediction of building loads forms the cornerstone for both energy supply strategies and integrated multi-energy systems, marking a critical advancement in the energy-efficient control of HVAC systems and underpinning feedforward predictive control mechanisms. The aim of this review is to identify best practices at each step of the load forecasting process through a comprehensive review of data processing, feature selection, and model selection methods from previous studies, in order to provide practical solutions for energy efficiency and sustainability in the building sector.

This review starts from the perspectives of building types, time scales, prediction objectives, load prediction models, and the diversity of feature types, summarizing the research achievements of load prediction. It reviews the efficacy of machine learning algorithms within these models, scrutinizing the data preparation and processing methodologies. Special attention is paid to advanced techniques like data fusion and transfer learning to mitigate issues related to data gaps and inconsistencies. Moreover, the importance of feature engineering is highlighted, discussing strategies for extracting pivotal features from heterogeneous data sources to accurately represent the intricate dynamics of building systems. Finally, the review indicates future research directions on room-scale load prediction and the application of graph neural networks and LLMs. Specific conclusions are as follows:

(1) Data preparation for building load prediction

Building load prediction requires various types of data, including building characteristics, meteorological conditions, energy usage, and sensor outputs. Data deficiencies and inaccuracies can undermine the stability and accuracy of predictive models. Data cleaning improves data quality by eliminating errors or outliers and filling in missing values, thus reducing noise impact, especially when dealing with historical load data. Data transformation, through standardization or normalization, converts raw data, improving model convergence speed and handling differences in feature scales, which is beneficial for processing raw load data. Dimensionality reduction methods, such as PCA or KPCA, retain key information while reducing data dimensions, enhancing the training efficiency of multi-dimensional load data and lowering the risk of overfitting.

Data fusion and augmentation techniques, employing technologies like Kalman filtering, maximum likelihood estimation, and Bayesian methods, increase data diversity, thereby improving model generalization ability and data representativeness. This is particularly useful in scenarios involving multi-source data and data scarcity. However, the complexity and computational cost of data fusion are high, requiring substantial expertise and computational resources. While data augmentation strategies can mitigate data insufficiency issues, excessive augmentation might introduce noise or irrelevant information, potentially degrading model performance.

Transfer learning leverages knowledge transfer in target tasks, making it particularly suitable for building load prediction scenarios with insufficient data. It compensates for data inadequacies by utilizing extensive data from similar buildings or environments. The limitations of transfer learning lie in the potential poor transfer effectiveness due to differences between buildings and environments, and the complexity of model training and tuning, which demands extensive experimentation and adjustments to ensure effectiveness in new scenarios.

Although data fusion and transfer learning hold significant potential in building load prediction, their application faces technical challenges and limitations, necessitating further research and optimization.

(2) Feature engineering in building load prediction

Raw data used in building load prediction often lacks critical features, making feature engineering an essential step in data preparation. This process requires an understanding of the building domain, data quality, and modeling requirements. By carefully selecting and creating meaningful features, model performance can be significantly improved, enabling it to handle the complexity of building systems for accurate predictions.

Feature Construction involves creating more informative features to enhance the model's generalization ability and accuracy, especially in building load prediction scenarios that require complex features to improve prediction precision. Since feature construction is complex and time-consuming, and requires domain experts' involvement, it should be applied when there is clear data and domain knowledge to ensure the features' validity and relevance.

Feature Selection reduces the number of features to improve model accuracy, lower the risk of overfitting, and enhance model interpretability. This is suitable for scenarios where simplifying the model and improving computational efficiency are needed. Filter Methods, Wrapper Methods, and Embedded Methods can be used to select the most important features. Overly aggressive feature selection might lead to the loss of important information, so it should be performed while ensuring data integrity.

Feature Extraction includes linear and nonlinear methods, such as PCA, LDA, and autoencoders, suitable for dimensionality reduction and feature extraction in high-dimensional data. When high-dimensional data contains crucial information, feature extraction might result in information loss. Therefore, careful selection of feature extraction methods is necessary to ensure key information is retained during dimensionality reduction.

Building load prediction often involves complex nonlinear relationships between climate conditions, physical characteristics of buildings, and usage patterns. High-dimensional data can lead to high computational costs and model overfitting. Techniques like PCA can reduce data dimensions while retaining the most important information. Direct modeling with scarce or imbalanced data categories does not yield ideal prediction results. Feature engineering can remove or reduce data noise, capture nonlinear relationships within the data, and create and extract more informative features to enhance model performance. For multi-source fused data, transfer learning and data augmentation methods can be used to extract and construct more effective features by leveraging data from other similar buildings. Leveraging domain experts' knowledge during feature construction and selection can significantly improve model performance and interpretability. Regular evaluation and updating of features during model development ensure that the model consistently uses the latest and most effective feature sets. Depending on the application scenario and data characteristics, appropriate feature engineering methods should be chosen to optimize model performance and meet specific prediction needs.

Existing raw data in research often has low quality, with missing values, noise, and inconsistencies. The single source of data and lack of diversity limit the effectiveness of feature engineering. Current feature engineering methods are often tailored to specific datasets and application scenarios, lacking generalizability, making it challenging to apply them across different types of buildings and environments. During feature extraction, especially in dimensionality reduction operations (e.g., PCA), there is a risk of losing important information, which can affect model prediction performance. Balancing the retention of key information during dimensionality reduction is a significant challenge. Overly aggressive feature selection might lead to the loss of crucial information, while retaining too many features can increase model complexity and the risk of overfitting. Finding an appropriate feature selection strategy is a challenging task.

(3) Comparative analysis of building load prediction methods

Data-driven building load prediction has been a major focus of research. Statistical learning methods are suitable for stable buildings with strong linear relationships. These methods are typically simple and easy to interpret, requiring low feature dimensions. However, they perform poorly in handling nonlinear and complex systems, failing to capture intricate relationships within the data. When faced with high-dimensional and nonlinear problems, the predictive performance of statistical learning methods is limited, making them unsuitable for dynamically changing building environments. Traditional machine learning methods can handle more complex nonlinear problems. Models such as Support Vector Machines (SVM) and Random Forests (RF) are widely applied in various building systems. These methods generally require a moderate number of features but have limited generalization capabilities when data is insufficient. Deep learning methods excel at addressing large-scale, high-dimensional nonlinear problems, particularly in scenarios requiring the capture of complex dynamic features. However, they demand significant computational resources and large amounts of data, limiting their application when data is sparse or computational resources are constrained. Hybrid model-based methods combine different types of models to leverage their respective strengths, enhancing prediction accuracy and model robustness. For instance, combining XGBoost and neural networks has shown excellent performance in predicting daily energy consumption in industrial buildings. These methods are suitable for complex scenarios requiring high-precision predictions, such as multi-source data fusion in building energy management and optimization. However, these models are more complex, with more intricate construction and training processes, and require high-quality and abundant data. Fine-tuned model design and optimization are necessary.

A common drawback of existing models is their high degree of customization, making them applicable only to specific situations with weak generalization capabilities, thus challenging the construction of universal models. Furthermore, current research mostly focuses on building or regional scales, with a lack of studies on room-scale prediction, which is crucial for improving HVAC control and occupant comfort. Different models have varying requirements for the number of features, ranging from a few basic design parameters to large amounts of historical data and meteorological information. Therefore, selecting an appropriate model requires careful consideration of the building system's complexity, data characteristics, and available computational resources. Further research and optimization of these methods are needed to overcome their limitations.

This review also highlights a research gap in room-level building load predictions due to the limited availability of room-level load data and advanced data processing requirements. In the era of rapid development in large language models, this review innovatively proposes a building load prediction framework based on LLMs. The aim is to address issues such as data scarcity, data processing complexity, and high computational resource demands in room-scale load prediction. The IDF files generation technology, based on RAG with LLMs, integrates parameters from building design requirements, HVAC standard parameters, and parameters retrieved from similar EnergyPlus models to enhance prompts. Using LLMs, high-quality IDF files containing all necessary building and system parameters are generated, simplifying the modeling process and lowering the modeling threshold. In addition, the feasibility of generating EnergyPlus models based on the RAG technique and LLM is verified by constructing single-area building models. The experimental results show that it takes only 196 s to generate an IDF file by this method, which greatly reduces the manual modelling time and improves the modelling efficiency. It is reasonable to believe that a large number of room-scale load prediction results matching actual buildings can be generated quickly based on this technical route, providing a data basis for constructing room-scale load prediction models. Then, LLM-based room-scale load forecasting techniques refine high-dimensional features through LLM to capture complex data relationships and construct comprehensive data representations. The extracted high-dimensional features are used to train a Spatio-Temporal Graph Neural Network (STGNN) model for load prediction. Finally, the prediction results are compared with simulation load results to optimize prompts and the STGNN network structure, continuously improving the accuracy and robustness of the prediction model through iterative optimization. However, the experimental analysis of room-level load forecasting based on LLM is subject to further work in the future.

The proposed LLM-based building load prediction framework combines data-driven models and LLM technology to provide an "end-to-end" load prediction method. This approach enhances prediction accuracy, improves model practicality, and reduces the professional threshold for model implementation.

Based on the limitations of the existing research, future research could focus on the following:

1) **Optimization of data fusion and augmentation methods**. Optimizing data fusion and augmentation methods to reduce noise introduction while enhancing data diversity and model generalization capabilities, thereby adapting to multi-source data and data-scarce environments. Developing intelligent transfer learning algorithms that can automatically adjust model parameters based on the target task and data characteristics, improving transfer effectiveness and enhancing model adaptability across different buildings and environments.
2) **Application of multi-source data fusion and transfer learning in feature engineering.** Applying multi-source data fusion techniques and transfer learning methods to feature engineering by utilizing data from other similar buildings to extract and construct more effective features, thereby improving model performance in data-scarce or imbalanced scenarios. Combining machine learning and domain knowledge to develop standardized and generalized feature engineering frameworks that meet the needs of different building types and environments, enhancing model generalization, and creating more universally applicable building load prediction models.
3) **Construction of universal building load prediction models.** Overcoming the customization and weak generalization issues of existing models by constructing universal models suitable for different building types and environments. Researching adaptive learning and online learning technologies to update model parameters in real-time, adapting to environmental changes. Developing dynamically adjustable hybrid models that combine the strengths of different model types, enhancing generalization and prediction accuracy in data-scarce and multi-source data environments.
4) **Research on efficient model training and inference algorithms.** Researching efficient model training and inference algorithms by utilizing distributed computing and edge computing technologies to accelerate the computation process. Developing lightweight models that reduce computational costs while maintaining model performance, enhancing real-time processing capabilities in resource-constrained environments.
5) **Development of room-scale load prediction models.** Through precise data collection and modeling, developing load prediction models suitable for room scale to address the deficiencies in current room-scale prediction research, thereby enhancing HVAC system control precision and occupant comfort. Combining sensor data, indoor environment data, and user behavior data to improve the accuracy and responsiveness of prediction models.
6) **Integration of domain expertise and data-driven methods.** Integrating domain expertise in the building field with data-driven methods by incorporating knowledge of building physical characteristics and usage patterns into the model development process. Developing interpretable machine learning models to aid in understanding prediction results, thereby enhancing the practicality and credibility of models.
7) **EnergyPlus model library retrieval algorithms and LLM prompt engineering.** Developing more efficient retrieval algorithms within a larger-scale EnergyPlus model library to search for cases more similar to the target building and optimize key variable inference. Enhancing LLM prompt engineering techniques by designing more comprehensive LLM prompt templates. Utilizing LLM's powerful feature representation and inference capabilities to improve model generation accuracy and quality. Integrating natural language processing, text mining, similar model retrieval, and LLM generation technologies to build an efficient, intelligent building load prediction framework. This framework aims to achieve an "end-to-end" automated modeling process, improving overall prediction accuracy and efficiency while lowering the professional threshold.

CRediT authorship contribution statement

Yake Zhang: Writing – original draft, Visualization, Methodology, Conceptualization. **Dijun Wang:** Supervision, Investigation, Funding acquisition. **Guansong Wang:** Writing – review & editing, Formal analysis. **Peng Xu:** Writing – review & editing, Supervision. **Yihao Zhu:** Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was funded by National Natural Science Foundation of China (No. 52161135202).

Data availability

No data was used for the research described in the article.

References

[1] Birol DF. World Energy Outlook 2022 n.d.
[2] A. Chong, W. Xu, S. Chao, N.-T. Ngo, Continuous-time Bayesian calibration of energy models using BIM and energy data, Energ. Build. 194 (2019) 177–190, https://doi.org/10.1016/j.enbuild.2019.04.017.
[3] M. Krarti, An overview of artificial intelligence-based methods for building energy systems, J. Sol. Energy Eng. 125 (2003) 331–342, https://doi.org/10.1115/1.1592186.
[4] D.C. Hittle, The building loads analysis and system thermodynamics (BLAST) program. Version 2. 0. Users manual. Volume I. Final report. Army Construction Engineering Research Lab., Champaign, IL (USA); 1979.
[5] D.A. York, C.C. Cappiello, DOE-2 engineers manual (Version 2. 1A). Lawrence Berkeley Lab., CA (USA); Los Alamos National Lab., NM (USA); 1981.
[6] K.W.D.K.C. Dahanayake, C.L. Chow, Studying the potential of energy saving through vertical greenery systems: Using EnergyPlus simulation program, Energ. Buildings 138 (2017) 47–59, https://doi.org/10.1016/j.enbuild.2016.12.002.
[7] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C. O. Pedersen, et al., EnergyPlus: creating a new-generation building energy simulation program, Energ. Buildings 33 (2001) 319–331, https://doi.org/10.1016/S0378-7788(00)00114-6.
[8] N. Alibabaei, A.S. Fung, K. Raahemifar, Development of Matlab-TRNSYS co-simulator for applying predictive strategy planning models on residential house HVAC system, Energ. Buildings 128 (2016) 81–98, https://doi.org/10.1016/j.enbuild.2016.05.084.
[9] P.A. Strachan, G. Kokogiannakis, I.A. Macdonald, History and development of validation with the ESP-r simulation program, Build. Environ. 43 (2008) 601–609, https://doi.org/10.1016/j.buildenv.2006.06.025.
[10] D. Yan, J. Xia, W. Tang, F. Song, X. Zhang, Y. Jiang, DeST — An integrated building simulation toolkit Part I: Fundamentals, Build. Simul. 1 (2008) 95–110, https://doi.org/10.1007/s12273-008-8118-8.
[11] X. Zhang, J. Xia, Z. Jiang, J. Huang, R. Qin, Y. Zhang, et al., DeST—An integrated building simulation toolkit Part II: Applications, Build. Simul. 1 (2008) 193–209, https://doi.org/10.1007/s12273-008-8124-x.

[12] H. Elmqvist, A Structured Model Language for Large Continuous Systems. Lund University, n.d.
[13] J. Runge, R. Zmeureanu, Forecasting energy use in buildings using artificial neural networks: a review, Energies 12 (2019) 3254, https://doi.org/10.3390/en12173254.
[14] G. Tardioli, R. Kerrigan, M. Oates, J. O'Donnell, D. Finn, Data driven approaches for prediction of building energy consumption at urban level, Energy Procedia 78 (2015) 3378–3383, https://doi.org/10.1016/j.egypro.2015.11.754.
[15] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, et al., A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renew. Sustain. Energy Rev. 33 (2014) 102–109, https://doi.org/10.1016/j.rser.2014.01.069.
[16] C. Deb, F. Zhang, J. Yang, S.E. Lee, K.W. Shah, A review on time series forecasting techniques for building energy consumption, Renew. Sustain. Energy Rev. 74 (2017) 902–924, https://doi.org/10.1016/j.rser.2017.02.085.
[17] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205, https://doi.org/10.1016/j.rser.2017.04.095.
[18] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, et al., A review of data-driven approaches for prediction and classification of building energy consumption, Renew. Sustain. Energy Rev. 82 (2018) 1027–1047, https://doi.org/10.1016/j.rser.2017.09.108.
[19] M. Bourdeau, X.Q. Zhai, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: a review of data-driven techniques, Sustain. Cities Soc. 48 (2019) 101533, https://doi.org/10.1016/j.scs.2019.101533.
[20] T.M. Mitchell, Machine Learning, McGraw-Hill, Nachdr. New York, 2013.
[21] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, et al., A review of machine learning in building load prediction, Appl. Energy 285 (2021) 116452, https://doi.org/10.1016/j.apenergy.2021.116452.
[22] J. Devlin, M.-W. Chang, K. Lee, K.B.E.R.T. Toutanova, Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics; 2019, p. 4171–86. https://doi.org/10.18653/v1/N19-1423.
[23] T.B. Brown, B. Mann, N. Ryder, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners 2020. https://doi.org/10.48550/arXiv.2005.14165.
[24] S. Mirchandani, F. Xia, P. Florence, B. Ichter, D. Driess, M.G. Arenas, et al., Language Models as General Pattern Machines 2023. https://doi.org/10.48550/arXiv.2307.04721.
[25] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, et al., Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online: Association for Computational Linguistics; 2020, p. 8342–60. https://doi.org/10.18653/v1/2020.acl-main.740.
[26] T. Li, S. Shetty, Kamath A, Jaiswal A, Jiang X, Ding Y, et al. CancerGPT: Few-shot Drug Pair Synergy Prediction using Large Pre-trained Language Models 2023. https://doi.org/10.48550/arXiv.2304.10946.
[27] Y. Gu, Z. Zhang, X. Wang, Z. Liu, M. Sun, Train No Evil: Selective Masking for Task-Guided Pre-Training. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online: Association for Computational Linguistics; 2020, p. 6966–74. https://doi.org/10.18653/v1/2020.emnlp-main.566.
[28] M.Q. Raza, A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, Renew. Sustain. Energy Rev. 50 (2015) 1352–1372, https://doi.org/10.1016/j.rser.2015.04.065.
[29] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renew. Sustain. Energy Rev. 73 (2017) 1104–1122, https://doi.org/10.1016/j.rser.2017.02.023.
[30] C. Kuster, Y. Rezgui, M. Mourshed, Electrical load forecasting models: a critical systematic review, Sustain. Cities Soc. 35 (2017) 257–270, https://doi.org/10.1016/j.scs.2017.08.009.
[31] T. Ahmad, H. Chen, Y. Guo, J. Wang, A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: a review, Energ. Buildings 165 (2018) 301–320, https://doi.org/10.1016/j.enbuild.2018.01.017.
[32] Y. Sun, F. Haghighat, B.C.M. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, Energ. Buildings 221 (2020) 110022, https://doi.org/10.1016/j.enbuild.2020.110022.
[33] J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, L. Xi, Review and prospect of data-driven techniques for load forecasting in integrated energy systems, Appl. Energy 321 (2022) 119269, https://doi.org/10.1016/j.apenergy.2022.119269.
[34] J.S. Hygh, J.F. DeCarolis, D.B. Hill, R.S. Ranji, Multivariate regression as an energy assessment tool in early building design, Build. Environ. 57 (2012) 165–175, https://doi.org/10.1016/j.buildenv.2012.04.021.
[35] R. Mena, F. Rodríguez, M. Castilla, M.R. Arahal, A prediction model based on neural networks for the energy consumption of a bioclimatic building, Energ. Buildings 82 (2014) 142–155, https://doi.org/10.1016/j.enbuild.2014.06.052.
[36] Y. Guo, J. Wang, H. Chen, G. Li, J. Liu, C. Xu, et al., Machine learning-based thermal response time ahead energy demand prediction for building heating systems, Appl. Energy 221 (2018) 16–27, https://doi.org/10.1016/j.apenergy.2018.03.125.
[37] Z. Yu, F. Haghighat, B.C.M. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energ. Build. 42 (2010) 1637–1646, https://doi.org/10.1016/j.enbuild.2010.04.006.
[38] A. Rahman, V. Srikumar, A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Appl. Energy 212 (2018) 372–385, https://doi.org/10.1016/j.apenergy.2017.12.051.
[39] C. Fan, Y. Sun, Y. Zhao, M. Song, J. Wang, Deep learning-based feature engineering methods for improved building energy prediction, Appl. Energy 240 (2019) 35–45, https://doi.org/10.1016/j.apenergy.2019.02.052.
[40] H. Lu, F. Cheng, X. Ma, G. Hu, Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower, Energy 203 (2020) 117756, https://doi.org/10.1016/j.energy.2020.117756.
[41] R. Wang, S. Lu, Q. Li, Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings, Sustain. Cities Soc. 49 (2019) 101623, https://doi.org/10.1016/j.scs.2019.101623.
[42] M.W. Ahmad, M. Mourshed, Y. Rezgui, Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption, Energ. Build. 147 (2017) 77–89, https://doi.org/10.1016/j.enbuild.2017.04.038.
[43] T. Hong, C. Koo, K. Jeong, A decision support model for reducing electric energy consumption in elementary school facilities, Appl. Energy 95 (2012) 253–266, https://doi.org/10.1016/j.apenergy.2012.02.052.
[44] T.-Y. Kim, S.-B. Cho, Predicting residential energy consumption using CNN-LSTM neural networks, Energy 182 (2019) 72–81, https://doi.org/10.1016/j.energy.2019.05.230.
[45] J. Kim, J. Moon, E. Hwang, P. Kang, Recurrent inception convolution neural network for multi short-term load forecasting, Energ. Buildings 194 (2019) 328–341, https://doi.org/10.1016/j.enbuild.2019.04.034.
[46] H.J. Sadaei, E. De Lima, P.C. Silva, F.G. Guimarães, M.H. Lee, Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series, Energy 175 (2019) 365–377, https://doi.org/10.1016/j.energy.2019.03.081.
[47] M. Cai, M. Pipattanasomporn, S. Rahman, Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, Appl. Energy 236 (2019) 1078–1088, https://doi.org/10.1016/j.apenergy.2018.12.042.
[48] L. Yan, M. Liu, A simplified prediction model for energy use of air conditioner in residential buildings based on monitoring data from the cloud platform, Sustain. Cities Soc. 60 (2020) 102194, https://doi.org/10.1016/j.scs.2020.102194.
[49] Y. Zhang, Q. Chen, Prediction of building energy consumption based on PSO - RBF neural network, in: 2014 IEEE International Conference on System Science and Engineering (ICSSE), 2014, pp. 60–63, https://doi.org/10.1109/ICSSE.2014.6887905.
[50] R.Ž. Jovanović, A.A. Sretenović, B.D. Živković, Ensemble of various neural networks for prediction of heating energy consumption, Energ. Buildings 94 (2015) 189–199, https://doi.org/10.1016/j.enbuild.2015.02.052.
[51] A. Lahouar, B.H. Slama, Day-ahead load forecast using random forest and expert input selection, Energ. Conver. Manage. 103 (2015) 1040–1051, https://doi.org/10.1016/j.enconman.2015.07.041.
[52] M. Chaouch, Clustering-based improvement of nonparametric functional time series forecasting: application to intra-day household-level load curves, IEEE Trans. Smart Grid 5 (2014) 411–419, https://doi.org/10.1109/TSG.2013.2277171.
[53] J.Z. Kolter, M.J. Johnson, REDD : A Public Data Set for Energy Disaggregation Research. Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability 2011.
[54] Commercial Buildings Energy Consumption Survey 2016.
[55] N.D. Mortimer, A. Ashley, M. Elsayed, M.D. Kelly, J.H.R. Rix, Developing a database of energy use in the UK non-domestic building stock, Energy Policy 27 (1999) 451–468, https://doi.org/10.1016/S0301-4215(99)00044-0.
[56] T. Hong, P. Pinson, S. Fan, Global energy forecasting competition 2012, Int. J. Forecast. 30 (2014) 357–363, https://doi.org/10.1016/j.ijforecast.2013.07.001.
[57] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, R.J. Hyndman, Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond, Int. J. Forecast. 32 (2016) 896–913, https://doi.org/10.1016/j.ijforecast.2016.02.001.
[58] T. Hong, J. Xie, J. Black, Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting, Int. J. Forecast. 35 (2019) 1389–1399, https://doi.org/10.1016/j.ijforecast.2019.02.006.
[59] Building Performance Database (BPD). EnergyGov n.d. https://www.energy.gov/eere/buildings/building-performance-database-bpd (accessed January 18, 2024).
[60] C. Miller, F. Meggers, The Building Data Genome Project: An open, public data set from non-residential building electrical meters, Energy Procedia 122 (2017) 439–444, https://doi.org/10.1016/j.egypro.2017.07.400.
[61] C. Miller, A. Kathirgamanathan, B. Picchetti, P. Arjunan, J.Y. Park, Z. Nagy, et al., The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition, Sci. Data 7 (2020) 368, https://doi.org/10.1038/s41597-020-00712-x.
[62] A.T. Bogatu, Cost-effective data wrangling in data lakes [Thesis], The University of Manchester, Manchester, UK, 2020 2020..
[63] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Autom. Constr. 50 (2015) 81–90, https://doi.org/10.1016/j.autcon.2014.12.006.
[64] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energ. Buildings 75 (2014) 109–118, https://doi.org/10.1016/j.enbuild.2014.02.005.

[65] C. Fan, F. Xiao, H. Madsen, D. Wang, Temporal knowledge discovery in big BAS data for building energy management, Energ. Buildings 109 (2015) 75–89, https://doi.org/10.1016/j.enbuild.2015.09.060.
[66] B. Cho, T. Dayrit, Y. Gao, Z. Wang, T. Hong, A. Sim, et al., Effective missing value imputation methods for building monitoring data, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, Atlanta, GA, USA, 2020, pp. 2866–2875, https://doi.org/10.1109/BigData50022.2020.9378230.
[67] J. Kim, H. Naganathan, S.-Y. Moon, W.K.O. Chong, S.T. Ariaratnam, Applications of clustering and isolation forest techniques in real-time building energy-consumption data: application to LEED certified buildings, J. Energy Eng. 143 (2017) 04017052, https://doi.org/10.1061/(ASCE)EY.1943-7897.0000479.
[68] Y. Chen, X. Hao, G. Zhang, S. Wang, Flow meter fault isolation in building central chilling systems using wavelet analysis, Energ. Conver. Manage. 47 (2006) 1700–1710, https://doi.org/10.1016/j.enconman.2005.10.023.
[69] F. Guarino, D. Croce, I. Tinnirello, M. Cellura, Data fusion analysis applied to different climate change models: an application to the energy consumptions of a building office, Energ. Buildings 196 (2019) 240–254, https://doi.org/10.1016/j.enbuild.2019.05.002.
[70] L. Xuemei, D. Lixing, X. Lv Jinhu, L.J. Gang, A novel hybrid approach of KPCA and SVM for building cooling load prediction, in: 2010 Third International Conference on Knowledge Discovery and Data Mining, Phuket, IEEE, Phuket, 2010, pp. 522–526, https://doi.org/10.1109/WKDD.2010.137.
[71] X. Deng, Y. Jiang, L.T. Yang, M. Lin, L. Yi, M. Wang, Data fusion based coverage optimization in heterogeneous sensor networks: a survey, Inform Fusion 52 (2019) 90–105, https://doi.org/10.1016/j.inffus.2018.11.020.
[72] Y. Himeur, A. Alsalemi, A. Al-Kababji, F. Bensaali, A. Amira, Data fusion strategies for energy efficiency in buildings: overview, challenges and novel orientations, Inform Fusion 64 (2020) 99–120, https://doi.org/10.1016/j.inffus.2020.07.003.
[73] X. Li, J. Wen, System identification and data fusion for on-line adaptive energy forecasting in virtual and real commercial buildings, Energ. Buildings 129 (2016) 227–237, https://doi.org/10.1016/j.enbuild.2016.08.014.
[74] H. Huajing, H. Sha, P. Xu, M. Lin, C. Peng, Q. Dou, A Framework of hybrid building energy forecasting model. Volume 10: Sustainable Energy Solutions for Changing the World: Part II, 2021.
[75] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Efficient Multi-Descriptor Fusion for Non-Intrusive Appliance Recognition. 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain: IEEE; 2020, p. 1–5. https://doi.org/10.1109/ISCAS45731.2020.9180493.
[76] Z. Wang, T. Hong, M.A. Piette, Data fusion in predicting internal heat gains for office buildings through a deep learning approach, Appl. Energy 240 (2019) 386–398, https://doi.org/10.1016/j.apenergy.2019.02.066.
[77] E. Fotopoulou, A. Zafeiropoulos, F. Terroso-Sáenz, U. Şimşek, A. González-Vidal, G. Tsiolis, et al., Providing personalized energy management and awareness services for energy efficiency in smart buildings, Sens 17 (2017) 2054, https://doi.org/10.3390/s17092054.
[78] Y.-Q. Xiao, Z.-Z. Hu, J.-R. Lin, in: Ontology-Based Semantic Retrieval Method of Energy Consumption Management, Springer International Publishing, Cham, 2019, pp. 231–238, https://doi.org/10.1007/978-3-030-00220-6_28.
[79] G. Huang, Y. Sun, P. Li, Fusion of redundant measurements for enhancing the reliability of total cooling load based chiller sequencing control, Autom. Constr. 20 (2011) 789–798, https://doi.org/10.1016/j.autcon.2011.02.001.
[80] N. Djuric, G. Huang, V. Novakovic, Data fusion heat pump performance estimation, Energ. Buildings 43 (2011) 621–630, https://doi.org/10.1016/j.enbuild.2010.11.003.
[81] G. Huang, Y. Sun, S. Wang, Building instantaneous cooling load fused measurement: multiple-sensor-based fusion versus chiller-model-based fusion, Build. Serv. Eng. Res. Technol. 34 (2013) 177–194, https://doi.org/10.1177/0143624411432651.
[82] M. Kumar, D.P. Garg, R.A. Zachery, Intelligent sensor modeling and data fusion via neural network and maximum likelihood estimation, Am. Soc. Mech. Eng. Digital Collection (2008) 1759–1768, https://doi.org/10.1115/IMECE2005-80972.
[83] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[84] E. Mocanu, P.H. Nguyen, W.L. Kling, M. Gibescu, Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning, Energ. Buildings 116 (2016) 646–655, https://doi.org/10.1016/j.enbuild.2016.01.030.
[85] D. Zhou, S. Ma, J. Hao, D. Han, D. Huang, S. Yan, et al., An electricity load forecasting model for Integrated Energy System based on BiGAN and transfer learning, Energy Rep. 6 (2020) 3446–3461, https://doi.org/10.1016/j.egyr.2020.12.010.
[86] X. Fang, G. Gong, G. Li, L. Chun, W. Li, P. Peng, A hybrid deep transfer learning strategy for short term cross-building energy prediction, Energy 215 (2021) 119208, https://doi.org/10.1016/j.energy.2020.119208.
[87] M. Ribeiro, K. Grolinger, H.F. ElYamany, W.A. Higashino, M.A.M. Capretz, Transfer learning with seasonal and trend adjustment for cross-building energy forecasting, Energ. Buildings 165 (2018) 352–363, https://doi.org/10.1016/j.enbuild.2018.01.034.
[88] P. Chujai, N. Kerdprasop, K. Kerdprasop, Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models. Hong Kong 2013.
[89] G. Huang, S. Wang, F. Xiao, Y. Sun, A data fusion scheme for building automation systems of building central chilling plants, Autom. Constr. 18 (2009) 302–309, https://doi.org/10.1016/j.autcon.2008.09.003.
[90] W. Wang, D. Tran, M. Feiszli, What Makes Training Multi-Modal Classification Networks Hard? 2020. https://doi.org/10.48550/arXiv.1905.12681.
[91] L. Zhang, J. Wen, A systematic feature selection procedure for short-term data-driven building energy forecasting model development, Energ. Buildings 183 (2019) 428–442, https://doi.org/10.1016/j.enbuild.2018.11.010.
[92] C. Wang, X. Li, H. Li, Role of input features in developing data-driven models for building thermal demand forecast, Energ. Buildings 277 (2022) 112593, https://doi.org/10.1016/j.enbuild.2022.112593.
[93] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, S. Ploix, A prediction system for home appliance usage, Energ. Buildings 67 (2013) 668–679, https://doi.org/10.1016/j.enbuild.2013.02.008.
[94] S. D'Oca, T. Hong, A data-mining approach to discover patterns of window opening and closing behavior in offices, Build. Environ. 82 (2014) 726–739, https://doi.org/10.1016/j.buildenv.2014.10.021.
[95] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Appl. Energy 195 (2017) 222–233, https://doi.org/10.1016/j.apenergy.2017.03.064.
[96] Y. Chu, P. Xu, M. Li, Z. Chen, Z. Chen, Y. Chen, et al., Short-term metropolitan-scale electric load forecasting based on load decomposition and ensemble algorithms, Energ. Buildings 225 (2020) 110343, https://doi.org/10.1016/j.enbuild.2020.110343.
[97] S. Sholahudin, H. Han, Simplified dynamic neural network model to predict heating load of a building using Taguchi method, Energy 115 (2016) 1672–1678, https://doi.org/10.1016/j.energy.2016.03.057.
[98] L.-G. Maltais, L. Gosselin, Predictability analysis of domestic hot water consumption with neural networks: From single units to large residential buildings, Energy 229 (2021) 120658, https://doi.org/10.1016/j.energy.2021.120658.
[99] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood, N. Amjady, Short-term electricity load forecasting of buildings in microgrids, Energ. Buildings 99 (2015) 50–60, https://doi.org/10.1016/j.enbuild.2015.04.011.
[100] K. Li, J. Tian, W. Xue, G. Tan, Short-term electricity consumption prediction for buildings using data-driven swarm intelligence based ensemble model, Energ. Buildings 231 (2021) 110558, https://doi.org/10.1016/j.enbuild.2020.110558.
[101] Y.-H. Hsiao, Household electricity demand forecast based on context information and user daily schedule analysis from meter data, IEEE Trans Ind Informat 11 (2015) 33–43, https://doi.org/10.1109/TII.2014.2363584.
[102] R.K. Jain, K.M. Smith, P.J. Culligan, J.E. Taylor, Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy, Appl. Energy 123 (2014) 168–178, https://doi.org/10.1016/j.apenergy.2014.02.057.
[103] P. Yuan, L. Duanmu, Z. Wang, Coal consumption prediction model of space heating with feature selection for rural residences in severe cold area in China, Sustain. Cities Soc. 50 (2019) 101643, https://doi.org/10.1016/j.scs.2019.101643.
[104] M.H. Kristensen, S. Petersen, Choosing the appropriate sensitivity analysis method for building energy model-based investigations, Energ. Buildings 130 (2016) 166–176, https://doi.org/10.1016/j.enbuild.2016.08.038.
[105] F. Ascione, N. Bianco, C. De Stasio, G.M. Mauro, G.P. Vanoli, Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach, Energy 118 (2017) 999–1017, https://doi.org/10.1016/j.energy.2016.10.126.
[106] T. Chen, Y.-C. Wang, Long-term load forecasting by a collaborative fuzzy-neural approach, Int. J. Electr. Power Energy Syst. 43 (2012) 454–464, https://doi.org/10.1016/j.ijepes.2012.05.072.
[107] K. Li, C. Hu, G. Liu, W. Xue, Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis, Energ. Buildings 108 (2015) 106–113, https://doi.org/10.1016/j.enbuild.2015.09.002.
[108] S. Mujeeb, N. Javaid, ESAENARX and DE-RELM: Novel schemes for big data predictive analytics of electricity load and price, Sustain. Cities Soc. 51 (2019) 101642, https://doi.org/10.1016/j.scs.2019.101642.
[109] X.J. Luo, L.O. Oyedele, A.O. Ajayi, O.O. Akinade, H.A. Owolabi, A. Ahmed, Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings, Renew. Sustain. Energy Rev. 131 (2020) 109980, https://doi.org/10.1016/j.rser.2020.109980.
[110] Y. Ding, Q. Zhang, T. Yuan, K. Yang, Model input selection for building heating load prediction: A case study for an office building in Tianjin, Energ. Buildings 159 (2018) 254–270, https://doi.org/10.1016/j.enbuild.2017.11.002.
[111] M. Wen, Z. Yu, W. Li, S. Luo, Y. Zhong, C. Changqing, Short-term load forecasting based on feature mining and deep learning of big data of user electricity consumption, AIP Adv. 13 (2023) 125315, https://doi.org/10.1063/5.0176239.
[112] Z. Li, D. Friedrich, G.P. Harrison, Demand forecasting for a mixed-use building using agent-schedule information with a data-driven model, Energies 13 (2020) 780, https://doi.org/10.3390/en13040780.
[113] E. Ceperic, V. Ceperic, A. Baric, A strategy for short-term load forecasting by support vector regression machines, IEEE Trans. Power Syst. 28 (2013) 4356–4364, https://doi.org/10.1109/TPWRS.2013.2269803.
[114] Z. Chen, J. Freihaut, B. Lin, C.D. Wang, Inverse energy model development via high-dimensional data analysis and sub-metering priority in building data monitoring, Energ. Buildings 172 (2018) 116–124, https://doi.org/10.1016/j.enbuild.2018.04.061.
[115] Z. Dong, J. Liu, B. Liu, K. Li, X. Li, Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification, Energ. Buildings 241 (2021) 110929, https://doi.org/10.1016/j.enbuild.2021.110929.

[116] M.T. Hagan, S.M. Behr, The time series approach to short term load forecasting, IEEE Trans. Power Syst. 2 (1987) 785–791, https://doi.org/10.1109/TPWRS.1987.4335210.
[117] Y. Wang, J. Wang, G. Zhao, Y. Dong, Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China, Energy Policy 48 (2012) 284–294, https://doi.org/10.1016/j.enpol.2012.05.026.
[118] H. Do, K.S. Cetin, Evaluation of the causes and impact of outliers on residential building energy use prediction using inverse modeling, Build. Environ. 138 (2018) 194–206, https://doi.org/10.1016/j.buildenv.2018.04.039.
[119] Y. Chen, M. Guo, Z. Chen, Z. Chen, Y. Ji, Physical energy and data-driven models in building energy prediction: A review, Energy Rep. 8 (2022) 2656–2671, https://doi.org/10.1016/j.egyr.2022.01.162.
[120] K. Amasyali, N. El-Gohary, Building lighting energy consumption prediction for supporting energy data analytics, Procedia Eng. 145 (2016) 511–517, https://doi.org/10.1016/j.proeng.2016.04.036.
[121] P. Vrablecová, A. Bou Ezzeddine, V. Rozinajová, S. Šárik, A.K. Sangaiah, Smart grid load forecasting using online support vector regression, Comput. Electr. Eng. 65 (2018) 102–117, https://doi.org/10.1016/j.compeleceng.2017.07.006.
[122] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, et al., Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings, Appl. Energy 195 (2017) 659–670, https://doi.org/10.1016/j.apenergy.2017.03.034.
[123] A.-D. Pham, N.-T. Ngo, T.T. Ha Truong, N.-T. Huynh, N.-S. Truong, Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability, J. Clean. Prod. 260 (2020) 121082, https://doi.org/10.1016/j.jclepro.2020.121082.
[124] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, Random Forest based hourly building energy prediction, Energ. Buildings 171 (2018) 11–25, https://doi.org/10.1016/j.enbuild.2018.04.008.
[125] Z. Wang, T. Hong, M.A. Piette, Building thermal load prediction through shallow machine learning and deep learning, Appl. Energy 263 (2020) 114683, https://doi.org/10.1016/j.apenergy.2020.114683.
[126] M.C. Leung, N.C.F. Tse, L.L. Lai, T.T. Chow, The use of occupancy space electrical power demand in building cooling load prediction, Energ. Buildings 55 (2012) 151–163, https://doi.org/10.1016/j.enbuild.2012.08.032.
[127] C. Zhou, Z. Fang, X. Xu, X. Zhang, Y. Ding, X. Jiang, et al., Using long short-term memory networks to predict energy consumption of air-conditioning systems, Sustain. Cities Soc. 55 (2020) 102000, https://doi.org/10.1016/j.scs.2019.102000.
[128] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans Evol Computat 1 (1997) 67–82, https://doi.org/10.1109/4235.585893.
[129] J.-S. Chou, D.-K. Bui, Modeling heating and cooling loads by artificial intelligence for energy-efficient building design, Energ. Buildings 82 (2014) 437–446, https://doi.org/10.1016/j.enbuild.2014.07.036.
[130] Z. Wang, Y. Wang, R.S. Srinivasan, A novel ensemble learning approach to support building energy use prediction, Energ. Buildings 159 (2018) 109–122, https://doi.org/10.1016/j.enbuild.2017.10.085.
[131] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Appl. Energy 127 (2014) 1–10, https://doi.org/10.1016/j.apenergy.2014.04.016.
[132] J. Wang, J. Wang, Y. Li, S. Zhu, J. Zhao, Techniques of applying wavelet de-noising into a combined model for short-term load forecasting, Int. J. Electr. Power Energy Syst. 62 (2014) 816–824, https://doi.org/10.1016/j.ijepes.2014.05.038.
[133] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, et al., PaLM: Scaling Language Modeling with Pathways 2022. https://doi.org/10.48550/arXiv.2204.02311.
[134] R. Taylor, M. Kardas, G. Cucurull, T. Scialom, A. Hartshorn, E. Saravia, et al., Galactica: A Large Language Model for Science 2022. https://doi.org/10.48550/ARXIV.2211.09085.
[135] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, et al. LLaMA: Open and Efficient Foundation Language Models 2023. https://doi.org/10.48550/ARXIV.2302.13971.
[136] C. Qin, A. Zhang, Zhang Z, Chen J, Yasunaga M, Yang D. Is ChatGPT a General-Purpose Natural Language Processing Task Solver? 2023. https://doi.org/10.48550/arXiv.2302.06476.
[137] O. Nov, N. Singh, D. Mann, Putting ChatGPT's medical advice to the (turing) test: survey study, JMIR Med Educ 9 (2023), https://doi.org/10.2196/46939 e46939.
[138] S. Chen, B.H. Kann, M.B. Foote, H.J. Aerts, G.K. Savova, R.H. Mak, et al., The utility of ChatGPT for cancer treatment information, Health Informatics (2023), https://doi.org/10.1101/2023.03.16.23287316.
[139] Y. Li, Z. Li, K. Zhang, R. Dan, S. Jiang, ChatDoctor: A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge 2023. https://doi.org/10.48550/ARXIV.2303.14070.
[140] OpenAI. GPT-4 Technical Report 2023. https://doi.org/10.48550/arXiv.2303.08774.
[141] G. Kortemeyer, Could an Artificial-Intelligence agent pass an introductory physics course? PhysRevPhysEducRes 19 (2023) 010132 https://doi.org/10.1103/PhysRevPhysEducRes.19.010132.
[142] S. Bordt, U. von Luxburg, ChatGPT Participates in a Computer Science Exam 2023. https://doi.org/10.48550/ARXIV.2303.09461.
[143] K. Malinka, M. Peresíni, A. Firc, O. Hujnák, F. Janus, On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree? Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1, Turku Finland: ACM; 2023, p. 47–53. https://doi.org/10.1145/3587102.3588827.
[144] A. Blair-Stanek, N. Holzenberger, B. Van Durme, Can GPT-3 Perform Statutory Reasoning? Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law, ACM, Braga Portugal, 2023, pp. 22–31.
[145] F. Yu, L. Quartey, F. Schilder, Legal Prompting: Teaching a Language Model to Think Like a Lawyer 2022. https://doi.org/10.48550/ARXIV.2212.01326.
[146] S. Bubeck, V. Chandrasekaran, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4 2023. https://doi.org/10.48550/arXiv.2303.12712.
[147] J. Ye, X. Chen, Xu N, Zu C, Shao Z, Liu S, et al. A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models 2023. https://doi.org/10.48550/arXiv.2303.10420.
[148] C. Ling, X. Zhao, J. Lu, C. Deng, C. Zheng, J. Wang, et al., Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey 2023.
[149] X.L. Li, Prefix-Tuning: Optimizing Continuous Prompts for Generation. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online: Association for Computational Linguistics; 2021, p. 4582–97. https://doi.org/10.18653/v1/2021.acl-long.353.
[150] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, et al., LoRA: Low-Rank Adaptation of Large Language Models 2021. https://doi.org/10.48550/arXiv.2106.09685.
[151] W. Pedrycz, S.-M. Chen, editors. Deep Learning: Concepts and Architectures. vol. 866. Cham: Springer International Publishing; 2020. https://doi.org/10.1007/978-3-030-31756-0.
[152] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, et al., Graph learning: a survey, IEEE Trans. Artif. Intell. 2 (2021) 109–127, https://doi.org/10.1109/TAI.2021.3076021.
[153] G. Jin, Y. Liang, Y. Fang, J. Huang, J. Zhang, Y. Zheng, Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey 2023.
[154] S. Eandi, A. Cini, S. Lukovic, C. Alippi, Spatio-temporal graph neural networks for aggregate load forecasting, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, Padua, Italy, 2022, pp. 1–8, https://doi.org/10.1109/IJCNN55064.2022.9892780.