

Deep Learning Techniques for Load Forecasting



Neeraj, Pankaj Gupta, and Anuradha Tomar

Abstract Electricity load dominates energy consumption and greenhouse gas emissions. There are increasing concerns about climate change and the need to minimize energy consumption and enhance energy performance. Energy management, optimization, and planning all depend on forecasting load energy consumption. The data-driven approaches are the most popular approaches to energy forecasting. Deep learning techniques are a new category of data-driven models that have emerged in the recent years. They offer improved capabilities in managing big data, attribute extraction characteristics, and a better ability to model nonlinear phenomena. This paper examines the effectiveness and potential of deep learning-based approaches for load energy forecasting. This paper begins with a literature survey, tracked through an outline of deep learning-based concepts, methodologies, and examples. Following that, the current trends in published research were examined and how deep learning-based approaches may be utilized for forecasting and feature extraction. The study finishes with an analysis of current problems and recommendations for further research.

Keywords Load forecasting · Deep learning · Data-driven approaches · Energy consumption

P. Gupta and A. Tomar: These authors contributed equally to this work.

Neeraj (✉) · P. Gupta

Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, New Delhi, Delhi, India

e-mail: neeraj42819@gmail.com

A. Tomar

Department of Instrumentation and Control Engineering, Netaji Subhas University of Technology, New Delhi, Delhi, India

1 Introduction

1.1 Motivation

In 2018, nearly one-third of global energy consumption was accounted for by buildings and construction, and almost 40% of global CO₂ emissions. These percentages will continue to rise in the coming years. It is critical to minimize energy consumption and enhance energy efficiency in buildings and facilities to maintain sustainability. Many strategies and approaches to energy planning, management, and optimization can forecast and predict energy loads. These applications include modeling predictive controls, load demand management, load demand response, and optimization. Short- and long-term forecastings are available for scheduled maintenance, renovations, and planning. Data-driven and physics-based models are the commonly used models for load forecasting. Nowadays, data-driven models are the most commonly used energy models. These representations can also be classified into either black-box or gray-box representations. But, the physics-based models can describe the system and its components in detail. However, such models require many measured parameters to be developed and calibrated. It can be challenging to obtain the parameters needed in many cases.

On the other hand, data-driven models usage mathematical models derived from measured data. These models do not require a large number of parameters nor detailed knowledge about the building/plant or system's internal components. Many buildings/plants have smart meters and automation systems, making data access easier. These data are easily accessible and can be used to forecast the load energy.

1.2 *Compilation of Published Papers on Data-Driven Approaches for Load Forecasting*

The popularity of data-driven approaches has increased in the recent years. There have been several literature reviews published. Each study focused on a different component of energy models. A summary will be provided in this section and highlight the main points of each paper. These selections are based on the most recent advances in artificial intelligence, especially deep learning-based techniques increasing in popularity from 2015 to 2016. The author in [1] compared artificial intelligence (AI) and statistical and physical models to estimate the energy consumption. The paper suggested future research directions, including developing better accuracy models, integrating these models into building energy management systems, and collecting data for future research. The capabilities and predictions of artificial neural networks were examined in [2, 3] looked at ANN, support vector machines (SVM), and hybrid models for the load forecast of energy usage. Ahmad et al. [4] have looked into how energy models interact with building controls and operations. The procedures are still not relevant, according to [5]. Future research should reduce the computing cost and

memory requirements while retaining accuracy. Wang and Srinivasan [6] reviewed AI-based energy prediction. They are particularly interested in ensemble-based and single-point models. The author examined AI-based and traditional ways of predicting electricity [7, 8] examined time series-based forecasting strategies to estimating energy usage, highlighting popular approaches, and mixed methodologies. A full study of machine learning (ML) techniques for building energy prediction may be found in [9]. The authors suggested a few suggestions for future investigation. They devised that deep learning algorithms be studied more because they are currently understudied. Furthermore, Ahmad et al. [10] have reviewed data-driven methods for the organization and estimation of building energy. The author has studied estimating, mapping benchmarking, and describing building energy models. And also focusing on how these methods have been used for large-scale and building applications [11]. The author studied data-driven models to forecast building energy consumption [12]. A breakdown of trends was also included.

Furthermore, Runge and Zmeureanu [13] provided a thorough study of artificial neural network applications for temperature prediction. The authors also recommended that further research should be done on deep learning-based approaches. It is engrossed on how ANN models can forecast power consumption [14]. In addition, the authors noted that future research should be focused on DL-based models. Aslam et al. [15] published a review of data-driven models for energy prediction. The focus was on feature engineering and data-driven algorithms. To the best of their knowledge, there has not been a literature review paper focused on DL models for forecasting energy consumption for energy loads. However, some published papers acclaim that forthcoming research on these techniques.

1.3 The Aim of the Literature Review

Although earlier literature reviews helped review and describe the current state using various applications of load forecasting models. This review aims to summarize the main points. There are still many gaps. The review paper [16] noted that there are not many review papers that emphasize new methods for load forecasting. It is noticed in a review paper that the deep learning models are the most emerging methods for load energy forecasting [17]. The author says no current paper focuses on load forecasting using deep learning approaches. Researchers may not be able to access the previous research because there are no review papers. The review paper [18] states that a future direction of research should be to establish a roadmap for machine learning-based load forecasting models. This paper attentions on establishing an idea for deep learning-based approaches that contribute to further research direction [19]. This paper reviews how deep learning-based approaches can predict load energy consumption. This paper addresses the gaps in papers identified by the literature analysis.

1.4 Objectives and Contributions

Deep learning approaches can be used to predict the load energy. The range of applications of such methods is extensive such as energy generation, smart grid networks, electricity price forecasting, and many others [20]. These models can also be used in other areas: air pollution [21], sales estimating [22], and others like health care and business. This work will only discuss the techniques used to forecast load energy consumption because of their wide range of applications. This literature review does not include integrating fuel cells, absorption, or adsorption systems. This paper will review several publications that use DL techniques to forecast load energy. This paper is organized as follows. The second section introduces deep learning and its many categories. The third section summarizes the current research trends. Section 4 examines the research that has employed deep learning-based feature extraction approaches in their research. Section 5 looks at papers that employed deep learning-based forecasting models. The future work, results and problems are discussed in Sect. 6. Section 7 brings the evaluation to a close.

2 Deep Learning Techniques

This part describes the fundamental descriptions, classifications, and approaches of deep learning that are used in exploration. Autoencoders, recurrent neural networks (RNNs), and deep neural networks (DNNs) are the most commonly used deep learning approaches. However, some approaches to deep learning, such as convolution neural networks (CNNs) and Boltzmann networks, are used in fewer cases. This section summarizes the most popular deep learning approaches that have been used for load forecasting.

2.1 History, Categorization, and a General Description

Deep learning approaches are popular for load forecasting due to their ability to deal with large amounts of data and feature extraction capabilities. So that the accuracy of the model is improved due to their features. This paper will overview several techniques and approaches to deep learning approaches. The word intelligence is the ability to process the information, take as input a bunch of information, and make some informed future decision or prediction. So, the field of artificial intelligence is simply the ability of computers to take as input: much information and use that information to inform some future situations or decision making. Deep learning is simply a subset of machine learning specifically focused on using neural networks, which extract useful features and patterns in the raw data and use them. Those patterns or features inform the learning tasks.

Traditional machine learning algorithms typically operate by defining a set of rules or features in the environment in the data right. The key idea of deep learning is that these features will be learned directly from the data itself in a hierarchical manner. These types of hierarchical features, and that is the goal of deep learning compared to machine learning, are the ability to learn and extract these features to perform machine learning on them. Today, we live in a world of big data, where we have more data than ever before. Neural networks are extremely and massively parallelizable. They can benefit tremendously and have benefited tremendously from modern advances in architecture that we have experienced over the past. Source toolboxes like TensorFlow can build and deploy these algorithms, and these models have become extremely streamlined.

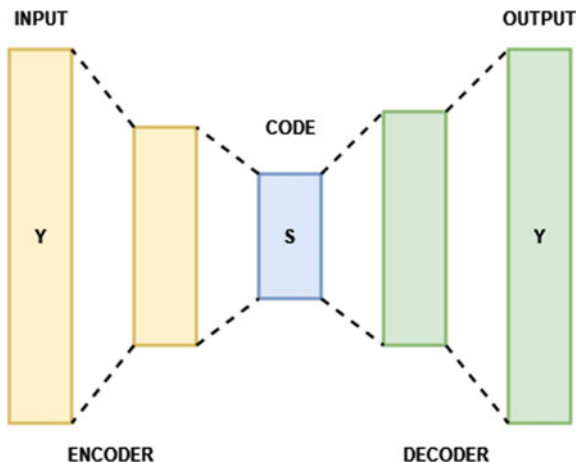
The deep learning architectures have four to five levels of nonlinear operations. First, deep learning is a way for practitioners to discover good features. This requires some engineering skills and domain expertise. Deep learning approaches do not need domain expertise; they can learn automatically using the general learning process. This is the main advantage of deep learning. Feature extraction can also be done automated. Deep learning can also easily deal with huge amounts of a dataset to make precise predictions. Nowadays, precise prediction using giant data is a growing problem, but deep learning solves such problems. These models can store and hold more information than conventional ANNs. Deep learning methods have a few drawbacks. They are not easy to train the model and contain a lot of hyperparameters. There are three main ways deep learning-based approaches were used to build power estimation.

1. Increment the number of layers concealed in a feed-forward neural organization or multi-facet discernment framework.
2. A few repetitive neural organizations like RNN, LSTM, and GRU are utilized. These intermittent neural network models can have at least one secret layer. These models can be regarded as networks with intense structures.
3. Consecutively coupling various calculations into one in general construction.

2.2 *Autoencoder*

In the case of autoencoder, the neural networks consist of multiple hidden layers. An autoencoder is composed of two sections. One is an encoder, and the other is a decoder section. An autoencoder's goal is to be able to recognize the dataset and reconstruct it using training. The encoder is the input of a hidden model, and the decoder is the output of a hidden model. Input data is represented by y , which is s equal to (y) . To make the output, the decoder extracts the hidden representation. The training aims to reduce the difference between input and output so that y equals to y' . Generally, an autoencoder is used for feature extraction in huge datasets. The structure of the autoencoder is shown in Fig. 1.

Fig. 1 Autoencoder

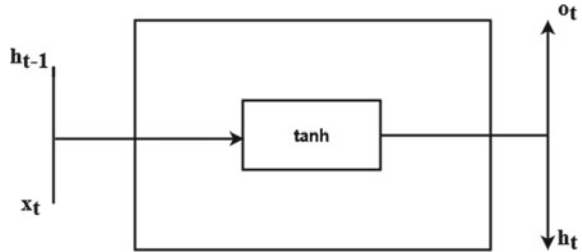


2.3 Recurrent Neural Network

Deep learning models can be used to process time-series data. Time-series data is a series of data tracked over time. Recurrent neural networks can solve the problem of feed-forward networks. The feed-forward networks such as density connected networks or convolutional neural networks. In other words, feed-forward networks do not consider the relationship between the current sample and the previous samples. The relationship between current and previous data is significant for some kinds of data, especially time-series data. The previous data predicts the following data to solve this problem. A loop is purposed to memorize the previous information [23]. RNN has a loop current connection/recurrent connection, representing the output back to itself. The RNN has a memory to remember the previous output and when the following input comes. RNN calculates a new output based on the current and previous outputs. So, recurrent neural networks can remember and memorize the previous data in the previous state. Therefore, the temporal relationship is considered to understand better how it works and can unfold on a loop in the time domain. Figure 2 shows that the input is time-series data x , and the output is data h . So first, the input data is unfolded data in the time domain. The data from the beginning of x_0, x_1, x_2 to x_t will have the output h_0, h_1, h_2 . It has only one cell, but this is the same cell at a different time. So, RNN will consider the previous input x_0 , save the output step, and then pass it to the next state. So when it has the following data x_1 , it can use x_1 and the previous output to calculate new output h_1 . Then set the state and then paste the state to the next cell. So this Fig. 2 unfolds an unknown loop that can better illustrate how RNN works.

$$S_t = F_w(S_{t-1}, X_t) \quad (1)$$

Fig. 2 RNN



This equation explains what RNNs are and how they work. X_t denotes the input at time step t , S_t denotes the state at time step t , and F_w is the recursive function.

$$S_t = \tanh(W_s S_{t-1} + W_x X_t) \tag{2}$$

A tanh function is a recursive function. W_x multiplies with the input state, while W_s multiplies with the prior state. Then, it passes through a tanh activation to get the new state. The weights are W_x and W_s . The new state S_t is multiplied with W_y to produce the output vector. It can be seen in Fig. 2, the input and output states are calculated using the previous and new state [24].

2.4 Long Short-Term Memory (LSTM)

RNN suffered a diminished and exploding gradient problem. The researcher proposed a long short-term memory model to solve the gradient management and exploding problem. It has become very successful. The long short-term memory adds multiple gates. First, they add an input gate to control if the new input is in or ignore the input and then add the forget gate. So, it can delete the trivial information. The output gate can decide to let the info impact the output at the current time step. The input gate usually outputs from zero to one. So if the output is zero, then the input will be ignored. If the gate output is one, the input will pass through to the hidden cell. So the gate is like a switch, and it is output continuously from zero to one. So it can control the part of the input that is passed to the hidden cell. So another gate is the forget gate, or if the forget gate is zero, the hidden cell's memory will clean right to zero. The last one is the output gate or the controller output that decides if the information pass to the next stage or not. The original LSTM picture is not easy to understand. The LSTM model has two paths. One is to update the memory state in the model. Another one is like the original RNN to pass the output to the next stage. For LSTM, it has two paths to pass the data to the next stage. The forget gate can control if it ignores the information. The sigma means the sigmoid activation function. The output of the sigmoid is between zero and one. So the sigmoid activation function is used as a switch of zero means turn off, and the one is turned on. It is also assigned a

value between zero. The activation is multiplied by the output with the previous state to control the portion of the information. The second state is the input gate. It can control to pass how much input information passes into the state to generate the new form. The third gate is the output gate that can control how much information is to pass to the next stage. LSTM model helps to solve the vanishing gradients problems [25].

$$i^t = \sigma(W^i[h^{t-1}, x^t] + b^i) \quad (3)$$

$$f^t = \sigma(W^f[h^{t-1}, x^t] + b^f) \quad (4)$$

$$o^t = \sigma(W^o[h^{t-1}, x^t] + b^o) \quad (5)$$

$$\bar{C}^t = \tanh(W^c[h^{t-1}, x^t] + b^c) \quad (6)$$

$$C^t = f^t C^{t-1} + i^t \bar{C}^t \quad (7)$$

$$h^t = \tanh(C^t) * o^t \quad (8)$$

i^t, f^t , and o^t are the input, forget, and output gates of the LSTM cell. W represents the recurring construction between the previously hidden and the existing layers. The hidden layers are connected to the input through the weight matrix. The cell state \bar{C} is calculated and depending on the current and previous input. C stands for the unit's internal memory. Figure 3 shows the equations that describe the behavior of all gates in the LSTM cell. As inputs, each gate accepts the hidden state and the current input x . The vectors are concatenated, and a sigmoid is applied. \bar{C} is a new potential value for the cell's state. The input gate controls the memory cell's updating. As a result, it is applied to the \bar{C} vector, which is the only one that can change the state of the cell. The forget gate determines how much of the previous state should be remembered. To get the hidden vector, this state is applied to the output gate [26] (Fig. 4).

2.5 Convolutional Neural Networks

Convolutional neural networks are a family of neural networks characterized by convolutional layers. They are particularly suitable for tasks involving data with spatial dependencies, such as images and videos. Convolution is a filtering operation applied to the data to detect certain features. This is just a matrix of numbers for a computer, with one value for each pixel. For seeing the borders, take a smaller filter matrix called kernel and perform an element-wise product between the kernel values and a portion of the image. Then sum up the results and get a single value, which indicates whether in that portion of the image borders are present or not. The kernel is then shifted by several pixels to cover another section until the whole image has been covered. The final result is a new matrix, called a feature map, whose numbers describe the borders. A convolutional layer implements several kernels,

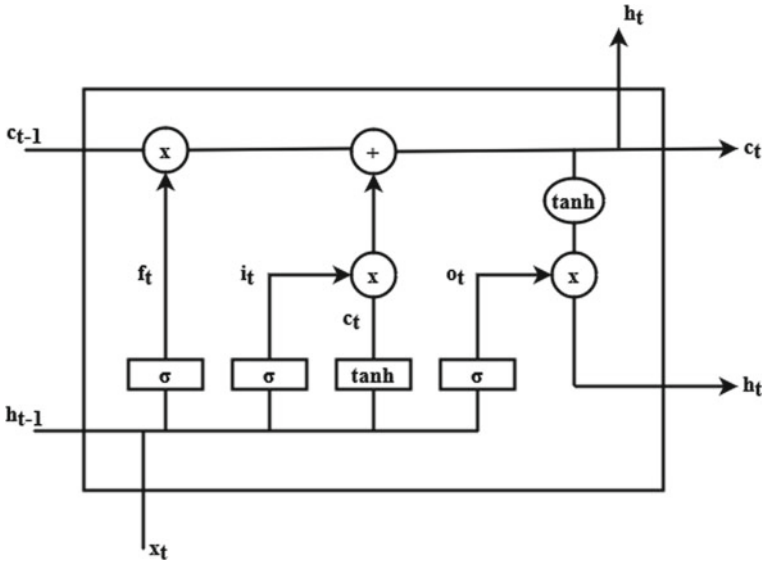


Fig. 3 LSTM

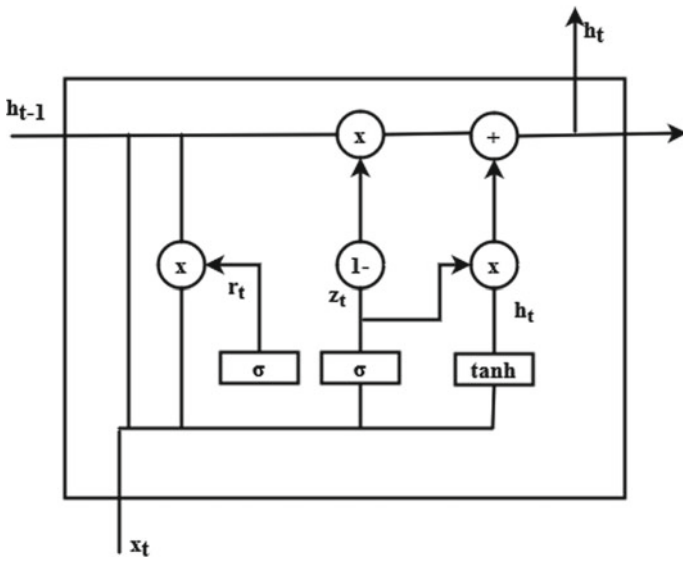


Fig. 4 GRU

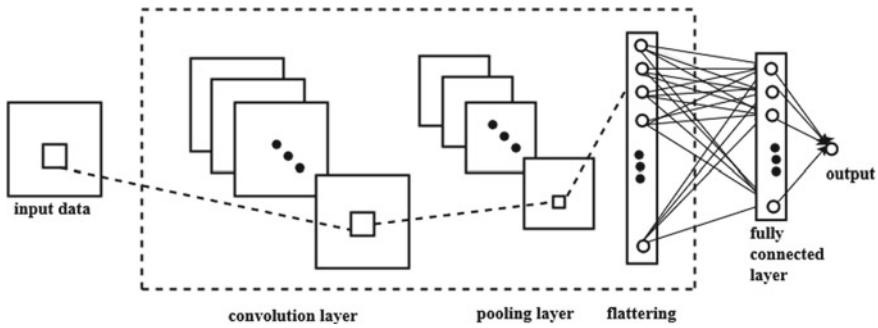


Fig. 5 CNN

each detecting a specific feature. The cool thing about convolutional layers within a neural network is that it does not have to design the kernels in advance.

During training, the network decides the important features and adapts the kernel to detect them. The parameters to set in this stage are the number of kernels to train, the kernel size, and the convolution dimension such as 1D, 2D, and 3D convolution. The difference between 1D, 2D, and 3D convolutions is: the convolution dimension sets the number of axes on which the kernel moves. In a 1D convolution, the kernel moves along one axis; in a 2D convolution along two axes, and so on. Convolutions with different dimensions discover features in those dimensions. The data dimension does not necessarily bind the dimension of the applied convolution. For example, black and white images are 2D objects, while color images are 3D objects because of the additional color channel. In both cases, if we are interested in 2D features, like borders, a 2D convolution moving along the width and height of the image will do the job. The same holds for time series. There are two dimensions, values and time. If we want to discover a 1D feature such as upward trends, we can apply a 1D convolution. Therefore, the dimension of the convolution is determined by the dimension of the feature to discover, not by the object's dimension. A CNN is represented in Fig. 5. Allude to reference [27] for a point by point depiction of a CNN's overseeing conditions and merits.

2.6 Deep Belief Networks

Deep belief networks (DBNs) is a type of deep neural network created from [28]. DBNs can be described as a range of algorithms that combine probabilities with unsupervised learning to produce outputs. The restricted Boltzmann machines (RBM) are fundamental to the DBN. It can then be configured to exhibit desirable properties [29]. The visible layer or input layer is the first layer of RBM. The hidden layer is the second. The RBM is illustrated in Fig. 6.

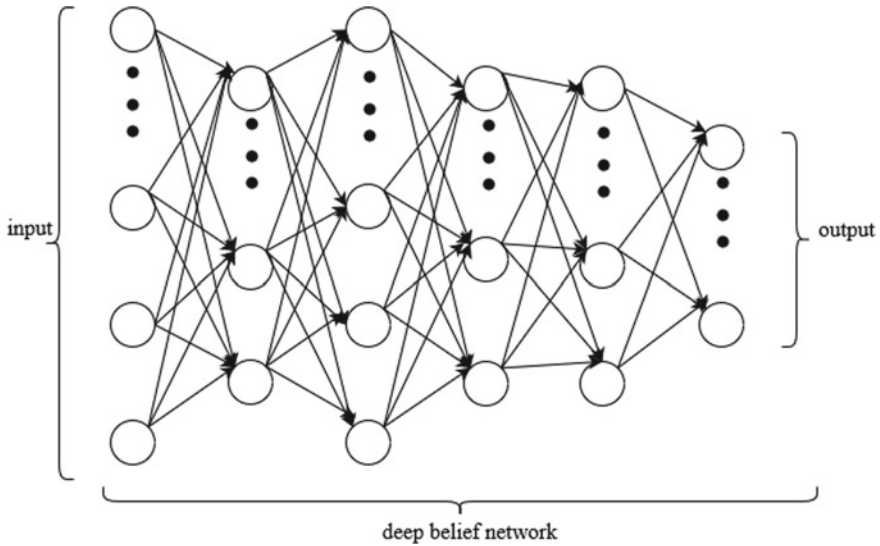


Fig. 6 DBN

Figure 6 shows an example of a DBN. Although stacking multiple RBMs together can produce large models, it may prove cumbersome to train such large models. Refer to references [30] for more information about the governing equations, merits, and limitations of RBMs, DBNs, and their potential benefits.

2.7 Deep Feed-forward Neural Networks

Deep feed-forward neural networks (DFFNs) are another popular technique for forecasting energy in buildings. These models differ from the standard feed-forward neural networks (FFNNs) because they have multiple hidden layers. To extract more information from the data, additional layers are added. Research has shown that there are many other deep learning-based structures [31] (Fig. 7).

3 Trends in the Present

From 2000 to 2021, the publications are looked. This section examines the trends that have been observed in the published data.

3.1 Level of Building Application

Data-driven models need to be validated on testbeds before being used in real-world applications. There are four levels to these testbeds. The forecasting model may be modified by the building level data and time steps data. According to the analysis, the studies were broken into four categories: district level, buildings, sub-meter level, and component level. The use of data from existing systems for large-scale installations and district heating/cooling systems may explain the tendency to focus on whole building cases and districts.

3.2 Qualities of Data

In each case study, the data size varies to the length and amount of data. DL-based approaches are being suggested to solve the problem related to large amounts of data to handle the big data. According to the observed breakdown of published work, 18% used less than six months, 23% used 6 months to 1 year, 57% used more than 1 year, and 2% did not justify their data size. This review also examined data types. Three types of data are commonly used in the published research papers: energy plus, real data, and target data. According to the findings, 93% of the case studies were applied to real data. Following that were 4% for experimental data and 3% for the target data.

3.3 Output Variables

The DL-based model's energy usage was applied to the forecast. The sub-meter and components are target variables like electric heating and cooling demand, etc.

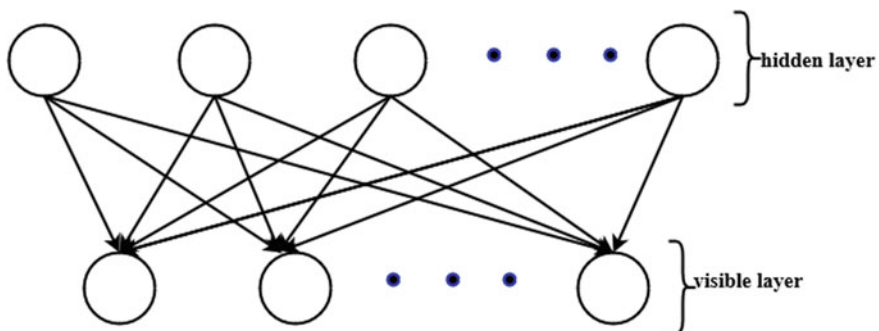


Fig. 7 Deep feed-forward neural network

3.4 *Input Styles*

The characteristics or regressors utilized as inputs into the forecasting model are inputs. All energy-based models require that you choose the correct input data. Data-driven models require the selection of appropriate input data. The poor choice of input variables may cause poor forecasting performance. The most commonly used features were: environmental data, such as outdoor temperature, and historical data, such as past energy use. Nowadays, it is not easy to find out which attributes are the most crucial. These may depend upon the various case study conditions such as weather, place, and type of structure. So, many feature extraction techniques were introduced in published research. Although a thorough examination of feature selection may be beneficial, the focus of this paper will be on DL-based approaches for variable selection.

3.5 *Granularity of Time*

Forecasting models have two main types: forecast horizon and resolution. The forecast horizon is the projected time. The term “resolution” mainly relates to the data’s time step. These two temporal granularities are applicable in various ways to forecast models. For hourly time step data, a forecast horizon can be used to estimate a horizon of 24h ahead. The models’ resolutions were 1% annually, 0% monthly, and 3% weekly. There are three types of prediction horizons: medium, long-term, and short-term [31]. It is important to note that the classifications mentioned above are not set in stone and may differ from those published.

4 **Feature Extraction Applications Using Deep Learning**

Feature selection is a process that reduces the size of an initial dataset into more manageable segments. Big datasets take a lot of computational time to process the model. So, computational time can be reduced by choosing the appropriate attribute. This can improve accuracy, reduce overfitting risks, and reduce computational resources for forecasting-based models. Recently, DL-based approaches are widely used for feature extractions and load forecasting. Due to its fast computing speed and simplicity of construction, this model has grown in popularity. Many studies have compared their efficacy to that of other data-driven models. In [32], compared four feature extraction techniques and forecasting models. This paper examined four different feature extraction methods:

- (i) Technical, which selected the model on the basis of technical expertise.
- (ii) Analytical calculated actual data from a response variable that could be used as a criterion.

- (iii) Architectural, in which the time series was transformed.
- (iv) Autoencoder.

Variables were selected from various models to forecast the energy load with different time horizons. Although, DL-based models are the best estimating performance models, the researchers conclude. The DL-based model is compared with data-driven models such as autoencoder and machine learning approaches in [33]. In the case of a retail facility, these models were utilized to forecast the total energy use. The model was used with a horizon of 60 min and 30-minute intervals. The autoencoder and machine learning approaches provide lower estimating errors. In [34], compared different feature selection methods. The performance of each method was compared with feed-forward neural network (FFNN), support vector regression (SVR), and random forest models. The models have trained with the resolution of 15 min and ahead horizon. The observation showed that the prediction error was reduced in 33% of the AE coupling clusters. However, the predicting inaccuracy was either maintained or significantly increased in 1/3 of the clusterings. In [35], the autoencoder model is compared with support vector machine (SVM) and FFNN. In this paper, forecasting was conducted on the office building energy load with the resolution of 5 min. They targeted the heating load and cooling load. Chitalia et al. [36] compared the various data-driven models for estimating the energy load of a commercial building with the resolution of 24 h ahead. This research found that combining DL feature extraction with estimating models results in a high-performing estimating model. DL feature removal methods to anticipate building energy use are still being developed. More study is required to compare these models across different case studies and applications.

5 Application Summary at the Load Level

To predict, the energy loads of the whole plant are the plant-level applications. The published research is classified into the following categories: educational, industrial, domestic, combined, etc. Combined states to publications that have applied their findings to various case studies involving various plant kinds.

This section contains publications that employed a DL-based approach to conduct a study on power loads. According to the analysis, all of the paper in this area came from educational places. There were also several case studies involving educational structures. Within this research, there are many case studies discussed. Cooling and electricity usages are the essential target load in the used research. There are still several gaps in understanding DL forecasting models for educational buildings, such as heating and lighting. This discrepancy might be due to the difficulties in obtaining data for specific loads [37]. On the other hand, the energy loads of heating and lighting can account for a significant portion of an industrial or educational building. They use 30% for heating and 15% for lighting [38]. As a result, future work may benefit from looking into these possibilities. In some papers, cooling loads were examined by

DL-based predicting models [39, 40]. Increasing the predicting efficiency for cooling loads uses an autoencoder for variable extractions [35, 45]. The author discussed in [28] the performance of the RNN, LSTM, and GRU-based models to predict the cooling loads using various approaches such as direct and recursive approaches. This paper showed that for the RNN model, the direct approach was more reliable. In [40], twelve predicting models are compared for the cooling load applications. According to this paper, the LSTM model gives more accurate results. In [41], various university campus heating load was predicted using the RNN model. According to their work, the RNN models worked better than the other machine learning-based approaches for medium and long-term estimation. They show that to get better results to predict the thermal energy loads, RNN models are used. More study is needed to corroborate the previous work based on different case studies. Marino et al. [42] show how GRU-based models may be used to predict energy usage. After examining several strategies for inferring missing data, GRU forecasting models were tested. For LSTM models that output power load predictions in educational buildings, see references [31]. The authors of the reference publication [43] evaluate the efficacy of several deep learning models.

6 Results and Discussion

Because deep learning approaches and techniques can manage vast volumes of data and give superior results, they have seen rapid expansion in the recent years. There is substantial research on their application in load energy forecasting. According to the observations, most of the DL-based models are used to predict the full power loads. They also target the energy loads for the entire plant. Most DL techniques have been used in the LSTM and deep feed-forward neural networks. When comparing forecasting performance with other ML-based methods, it was found that DL-based methods typically lead to better performance than ML-based ones. In some cases, however, it did not. Similar observations were also observed when the models were used as forecasting models. Despite the strong outcomes, there are still considerable hurdles to be overcome.

6.1 Challenges

Although the use of DL-based methods is still in their initial stages for forecasting the energy loads, several new and thrilling tasks remain. The most significant challenges that have been observed are divided into two categories:

1. The difficulties that the research community is confronted
2. The DL-based methods are faced the technical challenges

The following are the main challenges which the researchers face:

1. Most papers used unpublished proprietary datasets. This point was raised in a review study on data-driven models [38, 44]. Because of the extensive usage of proprietary data, it is not easy to produce the results, conduct comparison, and expand on the work of others.
2. As a result of the developing number of publishers, there is no standard for foreseeing model data in each diary composition.
3. Inadequate descriptions of the components/or methods used in their research. It was found that some papers did not specify their forecast horizons or hyperparameter tuning approach.
4. Many performance metrics can be applied to each publication. The most common performance metric in research is the mean absolute percentage error, found in references [45]. However, it is not always employed in the study. Occasionally, the author will utilize other metrics or change the measurements.
5. The issue is further complicated by using unclear terms in research.

A few significant challenges have been identified in the research on DL-based models. It is challenging to develop and test DL-based models without guidelines. Creating, applying, and comparing such models are much more challenging because there are no guidelines. According to the findings, the majority of articles had changed their hyperparameters by trial and error. Building various models and ensuring repeatability may be easier with an automated method and guideline. The models can improve forecasting performance at multiple levels, but they have a trade-off: increased complexity of the model and longer training times than typical machine learning approaches. Future researchers would benefit from the establishment of guidelines for DL modeling. This will provide them with a standard set of criteria that they can use to compare and build on models. This could allow for more generalizations more quickly.

6.2 Data Collection and Results

6.2.1 Data Collection

Data is collected from the Delhi power plant from January 2011 to December 2020 with hourly resolution. Before processing, any analysis dataset is normalized using a min-max scalar. Figure 8 represents the power consumption data before normalization, and Fig. 9 represents the data after normalization using a min-max scalar. Normalized data provides equal weights to each attribute. Because they are more significant numbers, no one attribute influences model performance in one way.

6.2.2 Results

From Figs. 10, 11, 12, it can be observed that the forecasted load performance is identical to the actual load consumption. All deep learning models provide accurate

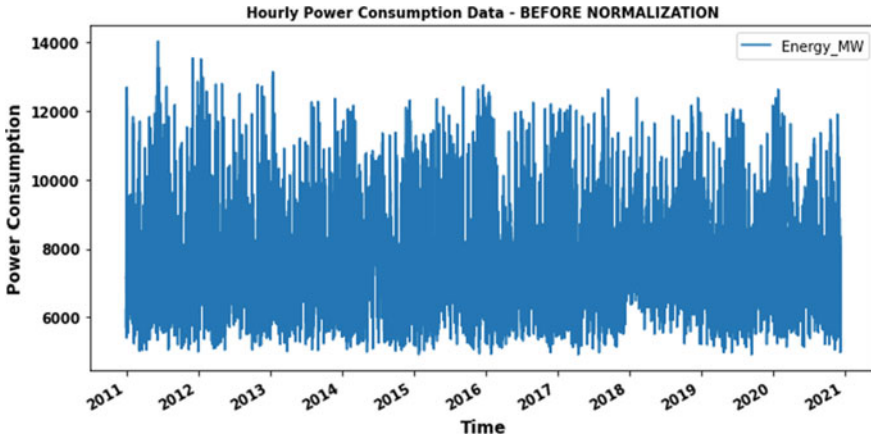


Fig. 8 Hourly power consumption data—before normalization

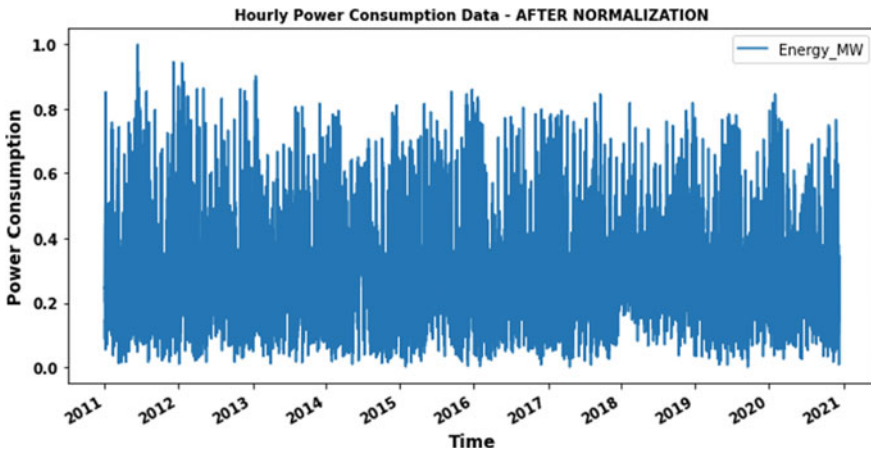


Fig. 9 Hourly power consumption data—after normalization

results as compared to the actual values. It can be seen that where the behavior of the power consumption load is unstable. The predicted results are that diverging from the actual values is moderately important. However, the power load consumption curves are repeated in all cases, as shown in Fig. 13. If the power load consumption graph is irregular or stable, the forecasted values fluctuate, especially when the power consumption load appears irregular.

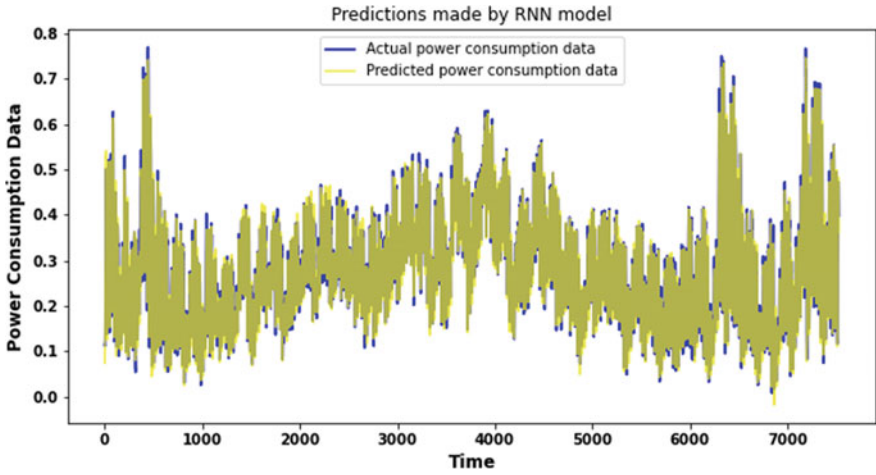


Fig. 10 Prediction made by RNN model

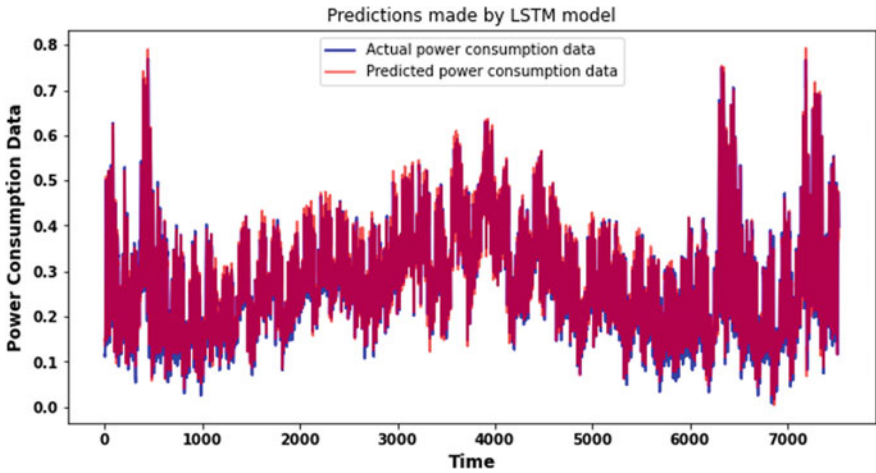


Fig. 11 Prediction made by LSTM model

6.3 Future Research Prospects

The potential of coming ways for DL-based methods in energy load forecasting contains:

1. The improvement of DL approaches across a scope of area types for load forecasting focuses on comparison-based papers.
2. The applications of DL models in research papers have not been much discussed.
3. Different case studies have been analyzed using DL gray-box models.

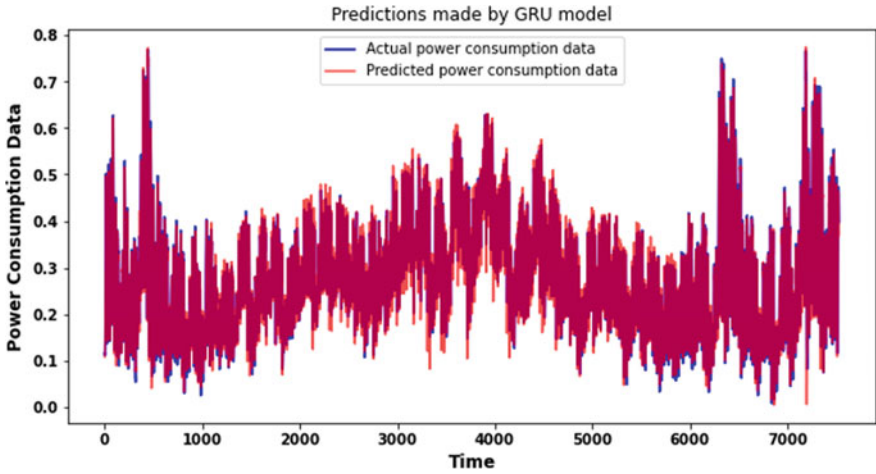


Fig. 12 Prediction made by GRU model

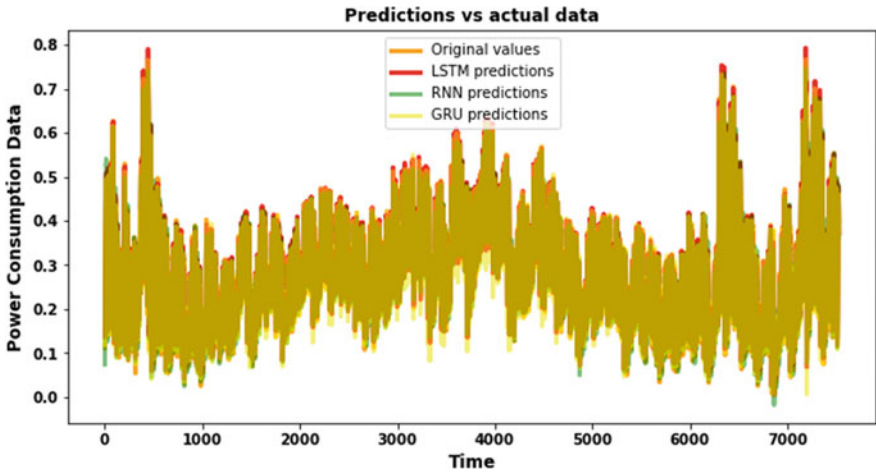


Fig. 13 Predicted versus actual

4. Analyze the sensitivity of DL models and their uncertainty.
5. The selection of hyperparameters for the DL model proposal establishes the guidelines.
6. The production of mountable DL-based models can be immediately evolved and tuned for use in various areas for load estimation.
7. The growth of strong models can provide accurate predictions even in the case of sensor fiascos, variations of the process, and other unforeseen events.

8. Implementation of innovative deep learning-based approaches in real-world applications, such as predictive model controllers, and demand-side management scheduling optimization.

7 Conclusion

This paper reviewed deep learning approaches that can be used to estimate the load energy consumption. Firstly, the concept and characteristics of deep learning approaches were discussed. After that, most widely used and vital methods of deep learning were presented. The basic overview and types for deep learning-based models are provided first, followed by an overview of some of the most popular methodologies. Following that, this report included a summary of current trends based on published studies. After that, attributes extraction and load forecasting using deep learning techniques were studied. At last, this paper discusses some issues related to such type of model using deep learning approaches. According to our review, deep learning strategies have proved to generate more significant performance outcomes when used to feature extraction when compared to other methods, according to our consideration. Furthermore, comparable effects were reported when the deep learning approaches were used as prediction models. Because there are few comparison-based studies among DL-based approaches, determining which one produced the most promising outcomes is challenging. However, the current results are encouraging, and future research should build on the existing part of the information. Despite the significant growth of the papers and case studies in the recent years, there are still several obstacles and tasks to be completed. The applications of the deep learning approaches are not typically used for case studies and target attributes. But the comparison of the deep learning approaches across various case studies and implementation of DL-based models is the actual application of such works. Many applications for energy management and improvement rely heavily on forecasting models. Predictive control, demand response management, fault detection, and optimization models are examples of such applications. The conversation and discoveries of this paper might assist the researchers with concluding which profound learning-based models are utilized for load determining.

References

1. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
2. Zhao H-X, Magoulès F (2012) A review on the prediction of building energy consumption. *Renew Sustain Energy Rev* 16(6):3586–3592
3. Kumar R, Aggarwal R, Sharma J (2013) Energy analysis of a building using artificial neural network: a review. *Energy Build* 65:352–358
4. Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, Saidur R (2014) A review on applications of ANN and SVM for building electrical energy consumption fore-

- casting. *Renew Sustain Energy Rev* 33:102–109
5. Wang Z, Srinivasan RS (2017) A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. *Renew Sustain Energy Rev* 75:796–808
 6. Wang Z, Srinivasan RS (2015) A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. In: 2015 Winter simulation conference (WSC). IEEE, New York, pp 3438–3448
 7. Deb C, Zhang F, Yang J, Lee SE, Shah KW (2017) A review on time series forecasting techniques for building energy consumption. *Renew Sustain Energy Rev* 74:902–924
 8. Amasyali K, El-Gohary NM (2018) A review of data-driven building energy consumption prediction studies. *Renew Sustain Energy Rev* 81:1192–1205
 9. Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, Han M, Zhao X (2018) A review of data-driven approaches for prediction and classification of building energy consumption. *Renew Sustain Energy Rev* 82:1027–1047
 10. Ahmad T, Chen H, Guo Y, Wang J (2018) A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: a review. *Energy Build* 165:301–320
 11. Bourdeau M, Qiang Zhai X, Nefzaoui E, Guo X, Chatellier P (2019) Modeling and forecasting building energy consumption: a review of data-driven techniques. *Sustain Cities Soc* 48:101533
 12. Mohandes SR, Zhang X, Mahdiyar A (2019) A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing* 340:55–75
 13. Runge J, Zmeureanu R (2019) Forecasting energy use in buildings using artificial neural networks: a review. *Energies* 12(17):3254
 14. Wang H, Lei Z, Zhang X, Zhou B, Peng J (2019) A review of deep learning for renewable energy forecasting. *Energy Convers Manage* 198:111799
 15. Aslam Z, Javaid N, Ahmad A, Ahmed A, Gulfam SM (2020) A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids. *Energies* 13(21):5599
 16. Marcjasz G (2020) Forecasting electricity prices using deep neural networks: a robust hyperparameter selection scheme. *Energies* 13(18):4605
 17. Tao Q, Liu F, Li Y, Sidorov D (2019) Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional GRU. *IEEE Access* 7:76690–76698
 18. Runge J, Zmeureanu R (2021) A review of deep learning techniques for forecasting energy use in buildings. *Energies* 14(3):608
 19. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge
 20. Wang H, Raj B (2017) On the origin of deep learning. *arXiv preprint arXiv:1702.07800*
 21. Hong T, Fan S (2016) Probabilistic electric load forecasting: a tutorial review. *Int J Forecast* 32(3):914–938
 22. Fan C, Xiao F, Zhao Y (2017) A short-term building cooling load prediction method using deep learning algorithms. *Appl energy* 195:222–233
 23. Fan C, Wang J, Gang W, Li S (2019) Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. *Appl Energy* 236:700–710
 24. Mishra S, Palanisamy P (2018) Multi-time-horizon solar forecasting using recurrent neural network. In: 2018 IEEE energy conversion congress and exposition (ECCE). IEEE, New York, pp 18–24
 25. Xiaoqiao H, Zhang C, Li Q, Yonghang T, Gao B, Shi J (2020) A comparison of hour-ahead solar irradiance forecasting models based on LSTM network. *Math Prob Eng* 2020:1–15
 26. Srivastava S, Lessmann S (2018) A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy* 162:232–247
 27. Son M, Moon J, Jung S, Hwang E (2018) A short-term load forecasting scheme based on auto-encoder and random forest. In: International conference on applied physics, system science and computers. Springer, Berlin, pp 138–144
 28. Rahman A, Srikumar V, Smith AD (2018) Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl Energy* 212:372–385

29. Kim J, Moon J, Hwang E, Kang P (2019) Recurrent inception convolution neural network for multi short-term load forecasting. *Energy Build* 194:328–341
30. Kim T-Y, Cho S-B (2019) Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* 182:72–81
31. Somu N, Gauthama Raman MR, Ramamritham K (2020) A hybrid model for building energy consumption forecasting using long short term memory networks. *Appl Energy* 261:114131
32. Shi Z, Li H, Cao Q, Ren H, Fan B (2020) An image mosaic method based on convolutional neural network semantic features extraction. *J Sign Process Syst* 92(4):435–444
33. He W (2017) Load forecasting via deep neural networks. *Proc Comput Sci* 122:308–314
34. Wang J, Chen X, Zhang F, Chen F, Xin Y (2021) Building load forecasting using deep neural network with efficient feature fusion. *J Mod Power Syst Clean Energy* 9(1):160–169
35. Kong Z, Zhang C, Lv H, Xiong F, Fu Z (2020) Multimodal feature extraction and fusion deep neural networks for short-term load forecasting. *IEEE Access* 8:185373–185383
36. Chitalia G, Pipattanasomporn M, Garg V, Rahman S (2020) Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Appl Energy* 278:115410
37. Zhang G, Tian C, Li C, Zhang JJ, Zuo W (2020) Accurate forecasting of building energy consumption via a novel ensemble deep learning method considering the cyclic feature. *Energy* 201:117531
38. Fan C, Sun Y, Zhao Y, Song M, Wang J (2019) Deep learning-based feature engineering methods for improved building energy prediction. *Appl energy* 240:35–45
39. Laib O, Khadir MT, Mihaylova L (2019) Toward efficient energy systems based on natural gas consumption prediction with LSTM recurrent neural networks. *Energy* 177:530–542
40. Wang Z, Hong T, Piette MA (2020) Building thermal load prediction through shallow machine learning and deep learning. *Appl Energy* 263:114683
41. Yang J, Tan KK, Santamouris M, Lee SE (2019) Building energy consumption raw data forecasting using data cleaning and deep recurrent neural networks. *Buildings* 9(9):204
42. Marino DL, Amarasinghe K, Manic M (2016) Building energy load forecasting using deep neural networks. In: *IECON 2016–42nd annual conference of the IEEE Industrial Electronics Society*. IEEE, New York, pp 7046–7051
43. Nichiforov C, Stamatescu G, Stamatescu I, Calofir V, Fagarasan I, Iliescu SS (2018) Deep learning techniques for load forecasting in large commercial buildings. In: *2018 22nd international conference on system theory, control and computing (ICSTCC)*. IEEE, New York, pp 492–497
44. Su H, Zio E, Zhang J, Xu M, Li X, Zhang Z (2019) A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-RNN model. *Energy* 178:585–597
45. Xue P, Jiang Y, Zhou Z, Chen X, Fang X, Liu J (2019) Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. *Energy* 188:116085