energies

MDPI

Review

A Review of State-of-the-Art AI and Data-Driven Techniques for Load Forecasting

Jian Liu, Xiaotian He, Kangji Li and Wenping Xue *

School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China; 17368620435@163.com (J.L.); 2222307018@stmail.ujs.edu.cn (X.H.); likangji@ujs.edu.cn (K.L.)
* Correspondence: xwping@ujs.edu.cn

Abstract

With the gradual penetration of new energy generation/storage, accurate and reliable load forecasting (LF) plays an increasingly important role in different energy management applications (e.g., power resource allocation, peak demand response, energy supply and demand optimization). In recent years, data-driven and artificial intelligence (AI) technologies have received considerable attention in the field of LF. This study provides a comprehensive review on the existing advanced AI and data-driven techniques used for LF tasks. First, the reviewed studies are classified from the load's spatial scale and forecasting time scale, and the research gap that this study aims to fill in the existing reviews is revealed. It was found that short-term forecasting dominates in the time scale (accounting for about 83.1%). Second, based on the summary of basic preprocessing methods, some advanced preprocessing methods are presented and analyzed. These advanced methods have greatly increased complexity compared with basic methods, while they can bring significant performance improvements such as adaptability and accuracy. Then, various LF models using the latest AI techniques, including deep learning, reinforcement learning, transfer learning, and ensemble learning, are reviewed and analyzed. These models are also summarized from several aspects, such as computational cost, interpretability, application scenarios, and so on. Finally, from the perspectives of data, techniques, and operations, a detailed discussion is given on some challenges and opportunities for LF.

Keywords: load forecasting; artificial intelligence; data-driven techniques; deep learning; ensemble learning; reinforcement learning; transfer learning

check for updates

Received: 15 June 2025
Revised: 22 July 2025
Accepted: 31 July 2025
Published: 19 August 2025

Citation: Liu, J.; He, X.; Li, K.; Xue, W. A Review of State-of-the-Art AI and Data-Driven Techniques for Load Forecasting. *Energies* **2025**, *18*, 4408. https://doi.org/10.3390/en18164408

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Currently, there is a risk of global warming and depletion of nonrenewable energy sources [1]. China aims to achieve carbon peak and carbon neutrality, with the goal of constructing a new type of power system based on renewable energy [2]. Under the framework of the new power system, the coupling between renewable energy generation, load operation, and energy dispatch has been strengthened, bringing new challenges to system accuracy and reliability [3]. On the other hand, with the diversification and complexity of the power grid usage patterns and the rapid popularization of micro-grids [4], the distribution system is becoming increasingly decentralized. Upon this background, reliable load forecasting (LF) has become increasingly important for power system scheduling and energy saving [5,6]. For example, it (1) guarantees the safe and stable operation of the power grid. LF is the foundation of dispatch. The power grid dispatching center relies on LF results to make power generation plans and peak shaving strategies, accomplishing

an energy balance between supply and demand [7]. LF with large errors may lead to voltage instability, frequency offset, and even system collapse. (2) Load forecasting helps achieve maximum accommodation of new energy. Generation of new energy is unstable, while LF provides data support for building a flexible grid response mechanism. Accurate forecasting of local loads can result in local source-load self-balance, alleviating the pressure on the main power grid [8]. (3) Load forecasting can also predict peak load periods. In the electricity market environment, predicting peak load periods avails developing better purchasing plans or pricing strategies [9]. However, many factors, such as weather conditions, human behavior, thermal performance, etc., may bring many difficulties for accurate load forecasting. Recently, data-driven techniques have been successfully applied in many fields [10–13]. Artificial intelligence (AI), which is represented by machine learning [14–17] and deep learning [18–22], has also received considerable attention. In the field of LF, application of AI and data-driven techniques is still increasing. Compared with traditional physical models [23,24] and time-series models [25], AI-based forecasting models have many advantages, such as good fitting abilities for high-dimensional nonlinear sample spaces, flexibly constructed input features, and the automatic optimization of model parameters with the aid of intelligent optimization algorithms [26]. The characteristic comparison of physical models, time-series models, and AI models is shown in Table 1.

Table 1. Comparison of three load forecasting models.

Model	Modeling Difficulty	Applicability of Time Scale	Parameter Acquisition	Interpretability
Time-series model	Easy (the principle is simple)	Short term or medium to long term	Model training with a small amount of data (easy)	Relatively strong
AI model	Easy (various universal models and mature toolkits can be used)	Short term or medium to long term	Model training with a large amount of data (relatively easy)	Weak
Physical model	Difficult (high level of multidisciplinary expertise is required)	Short term	Experimental testing or product manuals (difficult)	Strong

This paper provides a comprehensive literature review on LF models based on advanced AI and data-driven techniques. Studies published in the last 5 years (from 2020 to 2025) are searched using the databases of Web of Science® and Google Scholar®. Since this paper focuses on advanced AI models, some latest AI technologies, such as "deep learning", "reinforcement learning", "transfer learning", "ensemble learning", and "load forecasting", are chosen as the search's keywords. Using the above keywords, databases, and time range and taking into account factors such as relevance and citations, 165 papers were initially identified, for which their distribution is shown in Figure 1. These papers could be classified from two aspects, i.e., load spatial scale and forecasting time scale.

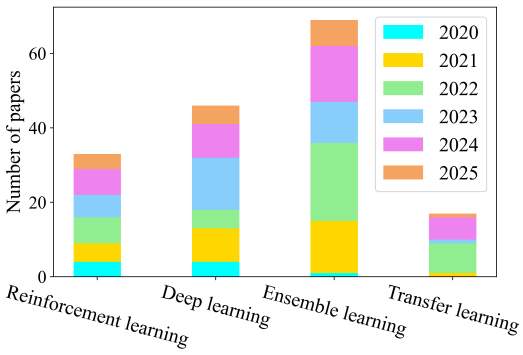


Figure 1. Distribution of identified studies on different AI methods from 2020 to 2025.

(I) *Load Spatial Scale:* Among 140 identified papers (the rest of the papers did not specify the spatial scale clearly), there are 47 papers (accounting for about 33.6%) related to city-/region-scale forecasting, 39 papers related to building-scale forecasting (such as residential, commercial, and educational buildings and so on), 30 papers related to station-scale forecasting, and 24 papers related to national-grid-scale forecasting. The statistic chart is shown in Figure 2.

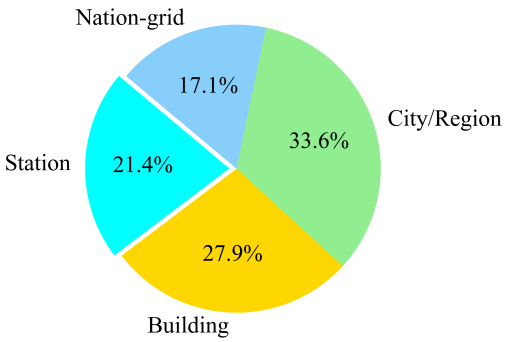


Figure 2. Statistical chart of studies on load spatial scale.

(II) *Forecasting Time Scale:* According to the time scale, LF can be categorized into short-term, medium-term, and long-term forecasting. Figure 3 provides the proportions of three time scales based on the survey of 154 papers (rest of the papers did not specify the time scale clearly). There are 128 papers (accounting for about 83.1%) focusing on short-term forecasting, 16 papers focusing on medium-term forecasting, and 10 papers focusing on long-term forecasting. It is observed that studies prefer short-term LF scales, which can provide more detailed and accurate forecasting results.

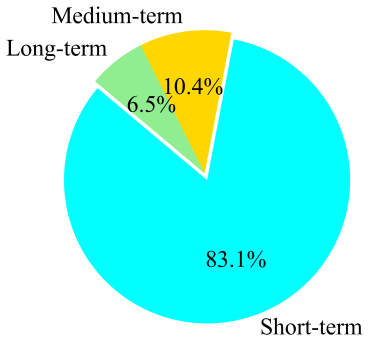


Figure 3. Statistical chart of studies on forecasting time scale.

This review aims to discuss recent research on LF models using advanced AI and data-driven technologies. In the past 5 years, many review papers on LF have been published [7,27–32]. For example, Ramokone et al. [27] conducted a chronological review of both AI-based and conventional models utilized in building energy consumption forecasting, including the model type, forecasting accuracy, and applied areas. Zhu et al. [7] presented a comprehensive overview of data-driven techniques for load forecasting in integrated energy systems. Khalil et al. [29] provided a systematic review on ML, deep learning, and statistical analysis models used in the field of building energy consumption forecasting. Han et al. [30] discussed the research status of AI-based LF for a new-type power system from nine typical LF scenarios. To comprehend the research gaps in previous reviews, Table 2 is provided, which is organized according to the main contents of our review: (I) statistics of spatial and temporal scales; (II) basic preprocessing methods; (III) advanced preprocessing methods; (IV) deep learning-based models; (V) reinforce-

ment learning-based models; (VI) transfer learning-based models; and (VII) ensemble learning-based models. From Table 2, it is observed that the majority of previous reviews did not involve advanced preprocessing methods (data preprocessing methods based on advanced AI). Further, besides deep learning-based forecasting models, other latest AI-based forecasting models were not deeply investigated. In particular, none of these reviews have considered reinforcement learning-based forecasting models. In this paper, the characteristics and applications of some advanced preprocessing methods are summarized, which may contribute to further improvements in forecasting accuracy. In addition, with the continuous development of AI, four kinds of latest AI-based models for LF are comprehensively reviewed, including their advantages, limitations, usage scenarios, and so on. The main parts of this review can be summarized as follows: (1) On the basis of summarizing traditional preprocessing methods, we present insights into the advanced AI technologies used for load data preprocessing. (2) On the basis of summarizing the characteristics of deep learning, reinforcement learning, transfer learning and ensemble learning, we comprehensively review and analyze various LF models using these latest AI technologies. (3) A detailed discussion on some challenges and opportunities of current technologies for LF is provided.

Table 2. Summary of previous review papers.

Contents	[27] (2021)	[28] (2022)	[7] (2022)	[29] (2022)	[30] (2023)	[31] (2024)	[32] (2025)
Spatial-/temporal-scale statistics	×	✓	✓	✓	×	×	✓
Basic preprocessing	×	×	×	✓	×	×	✓
Advanced preprocessing	×	×	✓	✓	×	×	×
Deep learning models	×	✓	✓	✓	✓	×	✓
Reinforcement learning models	×	×	×	×	×	×	×
Transfer learning models	×	×	✓	✓	×	×	✓
Ensemble learning models	✓	×	×	✓	✓	✓	×

2. Data Preprocessing Methods

It is well known that data processing plays an important role in time-series forecasting tasks. Apart from some commonly used methods such as normalization, outlier detection, and data filling [33], feature extraction and data decomposition are mainly addressed in this paper.

2.1. Basic Preprocessing Methods

(I) Feature extraction

Feature extraction usually includes feature selection, data-dimension reduction, and feature enhancement, which directly affect the speed and cost of model training. It is concluded that using redundant features may reduce the model's accuracy, while using features with strong correlation can improve the model's accuracy [34,35].

Three kinds of feature selection methods are commonly used: (1) machine learning-based methods such as recursive feature elimination (RFE) [36] and chi square [37]; (2) mutual information methods such as the maximal information coefficient (MIC) [38,39] and partial mutual information (PMI) [40]; and (3) statistic-based methods such as the least absolute shrinkage and selection operator (LASSO) method [41] and methods based on correlation measures. Spearman's correlation coefficient [42], Pearson's correlation coefficient [43], the distance correlation coefficient (DCC) [44], and the orthogonal maximum correlation coefficient (OMCC) [45] are commonly used to evaluate the correlation between

variables. Together with a defined threshold or ranking procedure, these correlation evaluation methods can be used to select input features that are more correlative with the output. Compared with methods based on correlation measures, LASSO is a relatively standard and effective feature selection method that performs coefficient shrinkage, often driving some coefficients to zero. In general, due to simple calculations and high operational efficiency, mutual information methods and statistic-based methods are often used for preliminary feature selection. They are suitable for scenarios with linear or monotonic relationships between the features and target variables. The machine learning-based methods are suitable for scenarios with complex coupling and nonlinear relationships between features. Considering the high computational cost, they are usually used for fine feature selection after preliminary selection. Furthermore, the above methods can also be combined to achieve better feature selection results. For example, Pirbazari et al. [46] utilized the Pearson correlation coefficient (PCC) to select features and RFE to further eliminate redundant features. Dai et al. [47] and Dong et al. [48] adopted a pre-trained random forest (RF) for ranking the importance of features and carrying out correlation analyses for feature selection.

Data-dimension reduction usually includes linear methods represented by principal component analysis (PCA), multidimensional scaling (MDS), singular value decomposition (SVD), etc. [49,50], and nonlinear methods represented by kernel principal component analysis (KPCA), isometric feature mapping (Isomap), t-distributed stochastic neighbor embedding (T-sne), etc. [51,52]. These methods are usually employed to further extract principal components within the feature space. They may help improve forecasting accuracy while interpretability is relatively low.

Some feature enhancement methods, such as sliding windows and feature differentiation, are utilized for LF [53,54]. Zhao et al. [55] used tree-based model prediction results to augment input features. These methods can effectively improve model accuracy and robustness.

(II) Data decomposition

Using data decomposition, the original time series can be decomposed into approximate stationary components. It has been proven that forecasting accuracy can be enhanced by separately modeling each component [56,57]. The frequency domain's decomposition could extract hidden information, which is beneficial for learning load response characteristics [58].

Some frequency domain decomposition methods are commonly utilized, including mode decomposition methods, such as empirical mode decomposition (EMD) [59], ensemble EMD (EEMD) [60], complete ensemble EMD (CEEMD) [61], variational mode decomposition (VMD) [62,63], improved VMD [64], complete EEMD with adaptive noise (CEEMDAN) [65], and bivariate EMD [66], and wavelet decomposition methods, such as wavelet transform (WT) [67], wavelet packet decomposition (WPD), and empirical wavelet transform (EWT) [68]. Mode decomposition methods are suitable for dealing with sequence signals with complex fluctuations and multiple frequency superpositions, while wave decomposition methods are suitable for dealing with load data with transient events or sudden changes. The seasonal-trend decomposition using the LOESS (STL) [69,70] method, which is commonly used for time series with seasonality and trend fluctuations, could also be used to achieve the related load sequence mode. Moreover, some combined methods, such as combinations of WPD and VMD [71]; VMD and fuzzy c-means (FCM) [72]; STL and VMD [73,74]; secondary VMD and wavelet packet threshold denoising (WPTD) [75]; improved CEEMDAN and VMD [76], etc., were employed to effectively extract load components by integrating the characteristics of different decomposition methods.

(III) Combined preprocessing

Combined preprocessing based on multiple data preprocessing methods has attracted a lot of attention in recent years. Zhao et al. [77] extracted high- and low-frequency components from the VMD sequence using grey relational analysis (GRA) and PCC, and they fused them to reconstruct ten new sequences. Li et al. [78] used fast EEMD to disintegrate the original time series and then adopted Pearson's coefficient to fuse highly correlated patterns; finally, they utilized hybrid feature engineering to choose input variables with high contributions. Tian et al. [79] used FCM clustering to divide time series and RFE to extract input features. Yan et al. [80] used the largest-triangle-three-bucket (LTTB) algorithm for dimension reduction, affinity propagation (AP) clustering for data decomposition, and transfer entropy to adaptively extract input features for each component. Lin et al. [81] adopted ICEEMDAN to decompose the sequence into high- and low-frequency components; then, they decomposed high-frequency components through VMD optimized by a crested porcupine optimizer (CPO), and they selected features for subsequences using a uniform information coefficient (UIC). The results demonstrated that these combined preprocessing methods could enhance the performance of LF models. In spite of the enhanced data preprocessing ability for load sequences with complex noise and fluctuations, these combined methods have significantly increased computational complexity and costs.

2.2. Advanced Data Preprocessing Methods

Recently, with the rapid development of AI, deep learning networks, such as the gated recurrent unit (GRU), convolutional neural network (CNN), and temporal convolutional network (TCN); reinforcement learning algorithms; and attention mechanisms have been applied to data preprocessing. Some representative data preprocessing methods based on these advanced AI technologies are shown in Table 3.

Table 3. Summary of data preprocessing methods using advanced AI.

Preprocessing Method	Literature (Year)	Network/Algorithm/Mechanism	Benefits of Preprocessing
Deep learning-based feature extraction	[82] (2021)	GRU	RMSE decrease: 30.52%
	[83] (2021)	GAN	RMSE decrease: 1.63%
	[44] (2021)	TCN	RMSE decrease: 41.20%; training time decrease: 11.32%
	[84] (2023)	CNN	R^2 increase: 3.85%
	[85] (2024)	GAN	RMSE decrease: 16.46%
Reinforcement learning-based feature selection	[71] (2021)	Sarsa	RMSE decrease: 0.96–10.08%
	[86] (2023)	DQN	RMSE decrease: 17.00%
	[87] (2024)	DDPG	Reliability improvement: 22.42%
	[88] (2024)	Actor–critic	RMSE decrease: 34.07%
Attention mechanism-based feature extraction	[89] (2021)	RNN, cross-attention	RMSE decrease: 18.5%
	[90] (2023)	GNN, hierarchical Attention	RMSE decrease: 6.82%
	[91] (2024)	CNN, global–local attention	RMSE decrease: 25.71%

GAN: Generative adversarial network; RMSE: root mean square error; Sarsa: state–action–reward–state–action; DQN: deep Q network; DDPG: deep deterministic policy gradient; GNN: graph neural network; RNN: recurrent neural network; R^2: coefficient of determination.

(I) Feature extraction based on deep learning

Deep learning belongs to a new field of machine learning research, and it has a structure of artificial neural networks with numerous hidden layers of multi-layer perceptions [92,93]. Feature extraction could be directly accomplished by the structural layers of deep learning networks. CNNs [59,84,94], GRUs [82], TCNs [44], generative adversarial

networks (GANs) [83,85], etc., were utilized for feature extraction. Due to the multi-layer neural network structures and strong nonlinear modeling and generalization capabilities, deep learning is more effective in extracting complex features from raw data than machine learning and statistics methods. The results in the literature showed that these deep learning-based feature extraction methods were helpful for LF models in terms of accuracy.

(II) Feature/sample selection based on reinforcement learning

Reinforcement learning achieves the optimal action based on an agent's interactions with the environment [95,96]. A simple illustration of reinforcement learning is shown in Figure 4. In the process of reinforcement learning, the agent observes the environmental state s, takes action a, receives a reward r from the feedback of the environment, and periodically completes each learning step. The main advantage of reinforcement learning lies in its ability to achieve autonomous learning through interactions with the environment, and it is particularly suitable for long-term decision optimization in complex dynamic environments. Reinforcement learning is demonstrated to be effective in the field of dynamic feature/sample selection for complex LF tasks. Liu et al. [71] established a state–action–reward–state–action (Sarsa) RL agent to dynamically select the optimal input features for a load's subsequence. Liu et al. [86] constructed a deep Q network (DQN) RL agent to optimize the optimal input features for building loads at different stages. Wang et al. [88] used the actor–critic network to dynamically choose the representative training data to reduce the impact of load data volatility. Liu et al. [87] adopted the deep deterministic policy gradient (DDPG) algorithm to select samples for wind power under extreme weather conditions, which significantly enhanced the adaptability of LF models to abnormal situations.

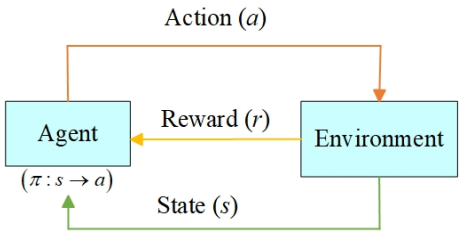


Figure 4. A simple diagram of reinforcement learning.

(III) Feature extraction based on attention mechanism

The attention mechanism can effectively extract information by reasonably adjusting the attention to information, ignoring irrelevant information, and amplifying important information [97]. The main mathematical expression of the self-attention mechanism is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

In a self-attention mechanism, Q is the query vector, which characterizes the sequence features that need to be focused on currently. K is the key vector, which quantifies the importance weights of each feature. By computing the dot product of Q and K, the correlation degree between the target sequence and other features can be obtained. V is the value vector, which carries the actual information content of each feature. Specifically, the dot product result of Q and K is normalized (e.g., by the softmax function) to generate attention weights that are then subjected to weighted summation with V to finally output the feature representation of the self-attention layer. In the LF task, the self-attention mechanism adaptively captures the time-series patterns and the influencing factors that are most relevant to the forecasting target by dynamically adjusting the attention weights of different features. Some researchers have introduced the attention mechanism into feature extraction for forecasting models. Li et al. [89] extracted load features using the attention mecha-

nism layer and recurrent neural network (RNN) layer. Yu et al. [90] developed a graph attention network to aggregate and extract the spatiotemporal features of wind power data. Yao et al. [91] incorporated the adaptive feature fusion convolution and global–local dynamic attention modules to realize effective local feature extraction at different temporal scales. The results verified that the above feature extraction methods could effectively extract deep feature information and enhance forecasting accuracy and robustness.

2.3. Comparison and Summary

A summary of the above-mentioned data preprocessing methods is given in Table 4, including the interpretability, complexity, application scenarios, and performance improvement of LF brought by them. From Table 4, the following is known: (1) In basic methods, the complexity of data decomposition is not high, while interpretability is relatively high. (2) In basic methods, the complexity and interpretability of feature extraction (including feature selection, data-dimension reduction, and feature enhancement) vary greatly depending on the specific method. (3) The advanced methods have greatly increased complexity compared to basic methods, while they may bring significant performance improvements for LF, such as adaptability and accuracy. (4) Selection of preprocessing methods should be based on their typical application scenarios.

Table 4. Summary of data preprocessing methods.

Classification	Method	Interpretability	Complexity	Application Scenarios	Performance Improvement
Basic methods	Statistic-/information-based feature selection	High	Low to medium	Linear/monotonic relationships between features and targets	Overfitting reduction, training acceleration
	ML-based feature selection	Medium	Medium to high	Complex coupling/nonlinear relationships between features	Training acceleration, improved generalization
	Data-dimension reduction	Low to medium	Medium to high	Feature redundancy	Noise reduction, training acceleration
	Feature enhancement	Low	Medium to high	Strong nonlinear data	Robustness improvement
	Mode decomposition	Medium to high	Low to medium	Series with fluctuations and multiple frequencies	Accuracy improvement, noise reduction
	Wavelet decomposition	Medium	Medium	Series with transient events or sudden changes	Noise reduction
	STL	High	Low	Series with seasonality and trend fluctuations	Robustness improvement
Advanced methods	DL-based feature extraction	Low	Very high	Big data, complex features	Accuracy improvement
	RL-based feature selection	Low	Very high	Dynamic feature selection	Adaptability improvement
	AM-based feature extraction	Medium to high	Very high	Deep/dynamic feature selection	Adaptability improvement, overfitting reduction

ML: Machine learning; DL: deep learning; RL: reinforcement learning; AM: attention mechanism.

3. Advanced AI-Based Forecasting Models

For complex LF tasks, the selection of forecasting model is critical. Classical statistical and machine learning models belong to time-series-based and AI-based models, respectively (see Section 1). In recent decades, statistical models, such as ARIMA [98] and exponential smoothing [99], and machine learning models, such as random forests [100], gradient-boosted trees [101], support vector machines [102], and extreme learning machines [103], have been thoroughly studied. These classical models are characterized by low computational costs and stable performance, which makes them widely used as benchmarks in both academic research and operational LF scenarios. Advanced forecasting models based on the latest AI usually exhibit higher accuracy and enhanced robustness than the above classical models. Four kinds of advanced AI-based LF models are reviewed in this section: deep learning-based models, reinforcement learning-based models, transfer learning-based models, and ensemble learning-based models.

3.1. Deep Learning-Based Models

Deep learning outperforms conventional machine learning because it has the following advantages [104–106]: (1) It exhibits higher flexibility and generalization capabilities; (2) it offers end-to-end learning; (3) it has stronger robustness when dealing with noise or incomplete data. Some latest deep learning models, such as the deep residual network (ResNet), TCN, and Transformer with attention mechanisms, have been widely applied in the field of LF.

(I) Deep ResNet

As one of the most successful deep learning models in recent years, ResNet can effectively solve problems, such as vanishing gradients, exploding gradients, and overfitting [107]. Chen et al. [108] proposed a trilinear ResNet with self-adaptive dropout methods and improved snapshot ensemble strategies for short-term LF. Zhang et al. [107] established deep ResNet with dual skip connections (ResNet-DSC) as the base learner of the combined forecasting model for residential loads. Hua et al. [109] proposed an ensemble framework for short-term LF, which used a CNN and long short-term memory (LSTM) for feature extraction and an attention mechanism layer for feature enhancement. Finally, the improved ResNet layer was utilized to output forecasting results, which significantly improved the accuracy of LF.

(II) TCN

For the TCN, the dilated causal convolution and residual blocks are the two most important parts. Apart from inheriting the advantage of causal convolutions in preventing future information leakage, TCN broadens the receptive field, allowing a wider range of input information [74]. Zhou et al. [74] established four TCN forecasting models for the components of two-stage load decomposition, which could fully capture long-term and short-term dependencies in the load sequence through distinct receptive fields. Su et al. [110] proposed a hybrid ensemble model for power LF, in which nine TCNs are obtained based on the cardinality minimization and augmented Lagrangian methods. Zheng et al. [111] adopted the hybrid LSTM–TCN model for short-term power LF, in which the LSTM was used to extract the temporal features, and TCN was used to build the connection between the features and output.

(III) Transformer with attention mechanism

Transformer is a deep learning model architecture developed for natural language processing (NLP) and sequence-to-sequence problems. Transformer is composed of positional encoding, multi-head attention mechanisms, stacked layers, residual connection layers and normalization layers, encoders and decoders, etc., as shown in Figure 5. The

Transformer model shows great potential in the field of LF due to its strong capabilities in feature extraction and sequence modeling. Zhao et al. [55] utilized a Transformer–ResLSTM network for short-term natural gas LF. This network replaced the original decoder structure with the LSTM network and fully connected layers, and it introduced residual connection mechanisms in the encoder and new decoder structure. Tian et al. [62] proposed a probabilistic LF model that integrated multi-layer CNNs and Transformer, noting that the Transformer could better identify the relationship between the leading and lagging terms of a time series. Gao et al. [112] established a novel Transformer-based model for short-term residential LF based on transfer learning and multi-attention fusion. The results demonstrated that the proposed model could enhance forecasting accuracy under both data-scarce and data-rich conditions.

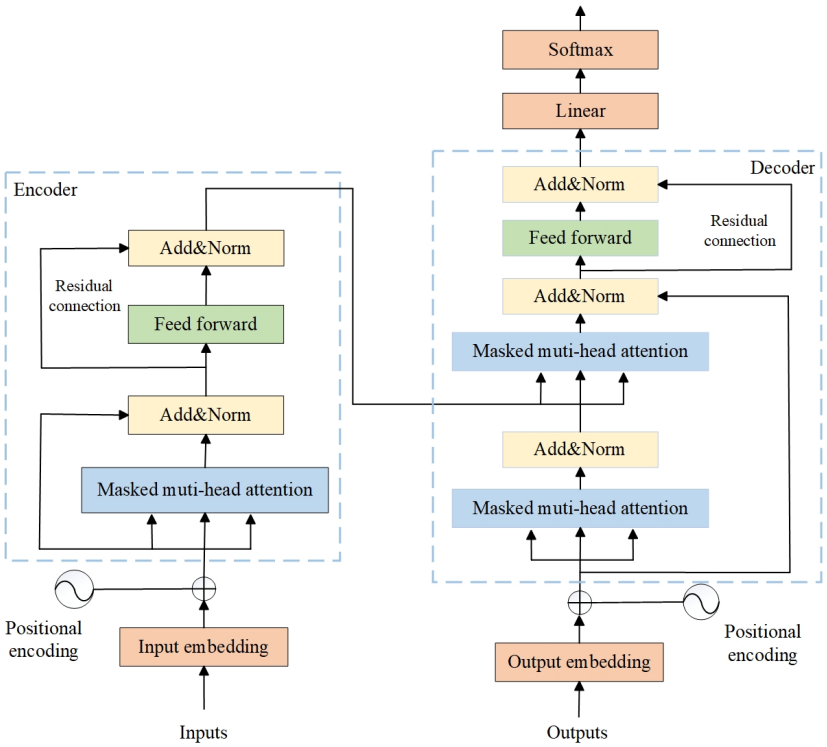


Figure 5. Structure of typical Transformer.

3.2. Reinforcement Learning-Based Models

Reinforcement learning emphasizes that, during the process of intensive interactions with a certain environment, the agent obtains the optimal decision-making policy based on trial and error [113,114]. In the field of LF, there are three main applications of reinforcement learning: direct forecasting, parameter optimization for forecasting models, and base learner construction/integration for ensemble models. Some representative RL-related papers on LF in the past 3 years are shown in Table 5.

Table 5. Representative applications of reinforcement learning in the field of LF.

Application	Literature (Year)	Reinforcement Learning Algorithm	State	Action
Direct forecasting	[115] (2021)	ADDPG	A sequence of load data	Forecasted load
	[116] (2022)	DQN	State class probabilities and historical energy consumption data	Discrete forecasted energy consumption

Table 5. *Cont.*

Application	Literature (Year)	Reinforcement Learning Algorithm	State	Action
Parameter optimization	[117] (2022)	DDPG	MAPE	Learning rate of LSTM
	[51] (2024)	REINFORCE	Hyperparameters and kernel operators for two models	Model hyperparameter values
	[118] (2025)	DDPG	Absolute value of error	Hyperparameters of GRU
Integration of base learners	[40] (2022)	Q-learning	Optimal learner label sequence in current step, absolute value of error	Optimal learner label sequence in next step
	[119] (2022)	Q-learning	Optimal learner label sequence in current step	Optimal learner label sequence in next step
	[120] (2022)	DDPG	Weights of base learners	Weight increments
	[90] (2023)	DDPG	Weights of base learners	Weight increments
	[121] (2024)	Q-learning	Weights of base learners in current step	Weights of base learners in next step
	[122] (2025)	DDPG	Predicted loads of base learners	LF value

ADDPG: Asynchronous deep deterministic policy gradient; REINFORCE: Monte Carlo policy gradient; MAPE: mean absolute percentage error.

(I) Direct forecasting

If reasonable state s and reward r are designed, the RL agent can output the optimal forecasting value a through training. Liu et al. [123] used RL models such as DDPG, recurrent deterministic policy gradients (RDPGs), and asynchronous advantage actor–critic (A3C) for directly building LF, which showed higher forecasting accuracy and longer training times compared with deep learning models. Fu et al. [116] designed an improved state based on the deep forest (DF) algorithm and adopted the DQN model for building energy consumption forecasting. Moreover, some improved RL frameworks were utilized for direct LF, such as RDPG [76] and ADDPG with adaptive early forecasting methods and reward incentive mechanisms [115]. Compared with parameter optimization and base learner integration, direct forecasting using RL remains limited, since it takes a large amount of training time and shows no obvious superiority compared with deep learning models.

(II) Parameter optimization

RL has great value in the field of optimization and decision making. Hence, using RL for model parameter optimization may lead to a more reliable forecasting strategy. Nguyen et al. [124] adopted the LSTM model for short-term electricity LF, and they used DQN to optimize the number of LSTM cells and the dimensions of the hidden state. Ge et al. [102] utilized an improved Q-learning algorithm to adjust the inertia weight of particle swarm optimization (PSO), which was combined with the least-squares support vector machine (LSSVM) power LF model. Chandrasekar et al. [51] proposed an improved RL algorithm using RNN as the policy network to dynamically adjust the hyperparameters of gradient boosting decision tree (GBDTs) and support vector regression (SVR) forecasting models. Li et al. [125] used the twin-delay deep deterministic policy gradient (TD3) to adaptively adjust the length of the time window of extreme gradient boosting (XGBoost) and CNN-LSTM LF models. Zhou [117] et al. and He et al. [118] constructed a DDPG agent to learn a policy that could dynamically tune the parameters of LF models. The results demonstrated that the above combined forecasting methods had higher accuracy and superior performance.

(III) Construction or integration of base learners

Under the framework of ensemble strategies, the RL algorithms can be applied for the construction or integration of base learners. Q-learning [126,127] was constructed as the base learner, which enhanced the diversity of the base learner library, thereby improving the performance of the ensemble forecasting model. Considering that the RL agent has the ability to adjust its policy to adapt to new observation information, RL algorithms have been widely used in the dynamic selection of base learners or the dynamic adjustment of base learners' weights. For example, in [40,65,68,119,121], the Q-learning algorithm was utilized to select the optimal base learner or adjust the base learner's weights for the ensemble forecasting model. In [128], the double DQN was employed to choose the best machine learning model in peak electricity forecasting. In [38], the DQN algorithm was adopted to dynamically tune the base learner's weights of the ensemble LF model for buildings. In [120], the DDPG network was used to realize the dynamic integration of the six base learners for heat load energy-saving predictions. The results verified that these RL-based ensemble strategies could further improve forecasting performance.

3.3. Transfer Learning-Based Models

Transfer learning can be defined as follows: Given the source domain and target domain, as well as the learning tasks, extract related knowledge from the source task to help the model accomplish the target task [129]. Transfer learning provides an effective solution for dealing with small sample problems [43,130]. Three kinds of transfer learning methods are mainly adopted in the field of LF, i.e., model transfer, instance transfer, and feature transfer. Some representative studies on LF using transfer learning are summarized in Table 6.

Table 6. Some representative studies on LF using transfer learning.

Literature (Year)	Transfer Algorithm	Transfer Method	Similarity Measurement
[131] (2022)	LSTM	Model transfer	/
[132] (2022)	LSTM	Instance transfer, model transfer	MIC
[43] (2022)	CNN-GRU	Model transfer	MMD, PCC
[133] (2022)	DANN	Feature transfer	/
[134] (2022)	MTE-LSTM	Model transfer	TSS-DC
[135] (2022)	ELM	Model transfer, feature transfer	/
[136] (2024)	K-MIFS-XGBoost	Instance transfer, model transfer	PCC
[129] (2024)	iTrAdaBoost-LSTM	Instance transfer	WD, NNS
[137] (2024)	BP, ELM, ENN, RBF, LSTM, GRU	Model transfer	PCC
[49] (2024)	CNN-LSTM	Feature transfer	MIC
[138] (2024)	DSSFA-LSTM	Model transfer	WM
[139] (2024)	CNN-GRU	Model transfer	MD

DANN: Deep adversarial neural network; MMD: maximum mean discrepancy; ELM: extreme learning machine; MTE-LSTM: multi-source transfer learning-guided ensemble LSTM; TSS-DC: two-stage source building selection strategy based on dominance comparison; WD: Wasserstein distance; NNS: nearest neighbor search; ENN: Elman neural network; RBF: radial basis function; WM: WD combined with MIC; MD: Mahalanobis distance; DSSFA: detrend singular spectrum fluctuation analysis algorithm.

(I) Model transfer

Model transfer means transferring the optimal model trained with the source domain to the target domain. That is to say that the model is first pre-trained with source domain data and then some parameters of the pre-trained model are fine-tuned to adapt to the

target task. Yang et al. [136] adopted the model transfer method to fine-tune the LF model's parameters using a small amount of load data. Xiao et al. [137] conducted a Pearson correlation analysis to determine the transferability between the source and target domains, and they proposed a Q-learning-based parameter transfer policy to optimize the transfer effectiveness for the short-term LF model. Wei et al. [138] proposed the WM algorithm, which combined Wasserstein distance (WD) and MIC, to select source domains with high similarity. Then, they trained the LF model using the data of selected domains, and they employed fine-tuning mechanisms to transfer the trained model to the target domain. Cheng et al. [139] utilized the Mahalanobis distance (MD) to measure the similarity between the source and target domains, and the model transfer method was used to construct the combined CNN-GRU model for short-term power LF. Case studies showed that the model transfer method could improve LF accuracy and efficiency.

(II) Instance transfer

Instance transfer means directly transferring some similar samples from the source domain to the target domain. It is verified that when the amount of data in the target domain is relatively small, the instance transfer method outperforms the model transfer method in terms of modeling accuracy [140]. Lu et al. [132] utilized MIC to measure the similarity between the target and source load sequence, and they transferred the source load data with similar patterns to supplement target data samples. Yang et al. [136] introduced the instance transfer method to determine the weight of holiday samples by calculating the PCC of holiday samples between the source and target domains. Wei et al. [129] proposed the instance-based multi-source transfer learning method for short-term building electricity LF. After choosing similar buildings using the WD approach, the authors picked the most helpful samples from source domains using the nearest neighbor search (NNS) approach and designed an improved TrAdaBoost algorithm for transfer and forecasting. Case studies demonstrated that the instance transfer method could improve the accuracy of LF with scarce historical data.

(II) Feature transfer

The feature transfer method maps data in the source and target domains to a unified feature space using feature transformation, thereby reducing the distance between the source and target domains. The feature transfer method may be more suitable for scenarios where the distribution difference between the two domains is large [141]. Zhou et al. [133] proposed a feature transfer method to capture related relationships between the unmasked-load and masked-load datasets for masked-load forecasting. Common feature vectors from the two datasets are extracted by adversarial training for accurate forecasting. Liu et al. [135] adopted feature transfer technology to update the weights of a multi-layer extreme learning machine model for probabilistic wind power forecasting. Lu et al. [49] used MIC combined with PCA for feature transformation. The spatial distribution of load characteristics in target and source domains was aligned by Wasserstein GANs, and the local characteristic loss was added in the transport cost function to improve training stability and the rapidity of the CNN-LSTM model. The results showed that above feature-based transfer learning enhanced the accuracy of model for small-sample LF.

3.4. Ensemble Learning-Based Models

It is admitted that no matter how data-driven models are improved, it is hard to develop a best single model that dominates over others for all LF applications. Ensemble learning, which has been widely used in detection, simulation, estimation, etc. [142–144], paves a totally new path toward achieve better forecasting performance than any individual

base learner. Two different ensemble structures, namely, serial and parallel ensemble structures, are discussed in this section for LF scenarios.

3.4.1. Parallel Ensemble Model

For the parallel ensemble model, the training process of each base learner is independent. There are usually two basic frameworks for parallel ensemble model: bagging and stacking. In the bagging framework, multiple training sets are constructed through sample extraction with replacement [26]. Then, multiple base learners are simultaneously built using the corresponding training set, and the final forecasting results are obtained by combining the results of base learners. In [40,47], the bagging framework was adopted for ensemble forecasting modeling. Compared with the bagging framework, the stacking framework is more commonly used in LF tasks. As shown in Figure 6, base learners with different learning mechanisms are trained as Layer-0 of the stacking framework, and the meta learner (or called strong learner) is constructed as Layer-1 to combine the base learners' forecasting results.

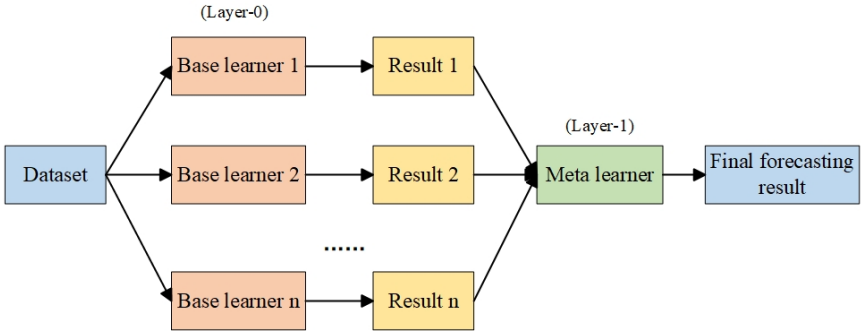


Figure 6. Parallel ensemble framework based on stacking.

For the stacking-based ensemble model, the selection of base learners and strategies for base learner combinations forms two main research aspects. To improve the forecasting performance of ensemble models, it is very important to select diverse base learners with low correlations. Selective ensemble learning was demonstrated to be effective in selecting the most suitable combination of base learners [145,146]. Combining the forecasting results of multiple base learners in a weighted manner is a kind of popular combination strategy. Weight optimization [81,147–149], weight correction [150], and weight searching [151] were utilized for ensemble models, and some encouraging forecasting results were obtained. Furthermore, some advanced meta learners are employed for stacking-based ensemble models, such as XGBoost [63], passive aggressive regression (PAR) [152], deep belief networks (DBNs) [153,154], quantized recurrent LSTM (QRLSTM) [155], extreme learning machine (ELM) [156], multiple linear regression (MLR) [157], light gradient boosting machine (LGBM) [158], etc. Li et al. [159] utilized multiple meta-learners, including RF, K-nearest neighbor (KNN), naive Bayesian (NB), and a linear discriminant (LD), and they ranked the available LF models based on accuracy. Then, the scoring–voting mechanism weighted each meta-leaner to make the final recommendations. The results verified that the meta-learning-based LF method achieved satisfactory performance. Some representative studies on parallel ensemble models for the LF tasks are summarized in Table 7.

Table 7. Some representative studies on parallel ensemble models for LF.

Literature (Year)	Basic Framework	Base Learner	Combination of Base Learners
[151] (2022)	Stacking	GRU, LSTM, TCN	Weight searching
[160] (2023)	Stacking	LSTM, GRU, TCN	Weight optimization (DE)
[64] (2024)	Stacking	TCN, ISSA-WFTS, BiLSTM-Attention	Weight optimization (ISSA)
[81] (2025)	Stacking	GRU, LSTM, BiLSTM, TCN	Weight optimization (MLP)
[46] (2021)	Stacking	BiLSTM	Meta-learner (GBDT)
[161] (2021)	Stacking	PSO, SA, ES, RS, BO-XGBoost, LGBM	Meta-learner (MLP)
[37] (2023)	Bagging	RNN, LSTM, GRU, BiLSTM	Weight optimization
[47] (2024)	Bagging	Attention-CNN-BiGRU	Average
[162] (2024)	Stacking	RNN, LSTM, GRU, BiLSTM, CNN	Meta-learner (KNN)
[163] (2025)	Stacking	TCN-LSTM, LightGBM	Meta-learner (MLR)

ISSA: Improved sparrow search algorithm; DE: differential evolution; ISSA-WFTS: weighted fuzzy time series based on improved sparrow search algorithm; MLP: multi-layer perceptron; SA: simulated annealing; ES: evolution strategy; RS: random search; BO: Bayesian optimization; LightGBM: light gradient boosting machine.

3.4.2. Serial Ensemble Model

For the serial ensemble model, the sequential process is used for base learner training. Taking the typical boosting algorithm as an example, each subsequent base model will be corrected and improved based on the previous base learner, as shown in Figure 7. Boosting and its improved algorithms are widely adopted for serial ensemble modeling. Meng et al. [164] designed a boosting-based ensemble model using multi-layer perceptrons (MLPs) and autoregressive integrated moving averages (ARIMAs) as the base learners, improving the accuracy of short-term LF models on small datasets. Zhu et al. [42] used the AdaBoost algorithm to enhance an ELM model optimized by the firefly algorithm for day-ahead industrial LF. Ma et al. [165] proposed a serial ensemble strategy for office energy consumption forecasting, incorporating the adaptive gradient boosting regression (AGBR) algorithm. Data from a real office building demonstrated that the proposed model had better performance in terms of accuracy and training time.

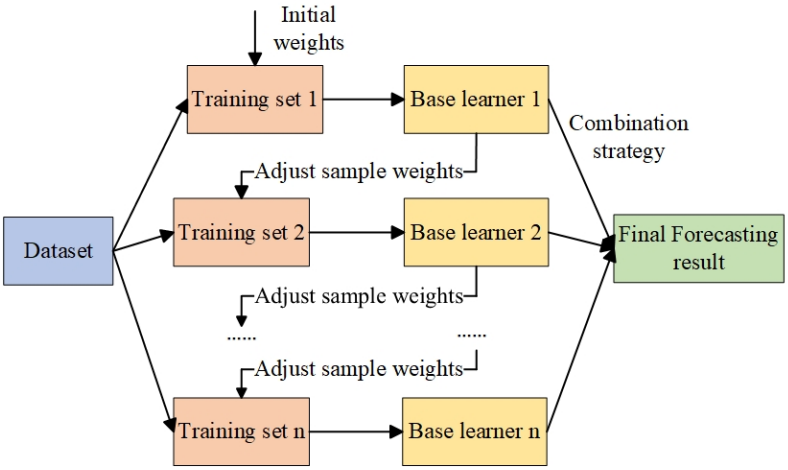


Figure 7. Serial ensemble framework based on boosting.

Apart from the boosting algorithm, the serial ensemble model based on residual correction has also attracted great attention. Introducing residual correction could further enhance the accuracy of forecasting model with fewer costs [52]. Zhao et al. [166] utilized the Gaussian process (GP) to estimate the forecasting residual of the deep ResNet when an abnormal event occurred. Comparative tests demonstrated the superiority of the proposed method for probabilistic LF against abnormal events. Ye et al. [167] proposed an error correction strategy for short-term wind power forecasting based on the swing window

method and quantile regression model. The results verified that the proposed strategy had good generalization capabilities. Zhao et al. [168] developed a global–local probabilistic LF framework using the Transformer network and residual learning method based on sparse variational Gaussian processes (SVGPs). Numerical simulations showed that the residual learning method could calibrate errors of the Transformer network, leading to enhanced forecasting accuracy. Some advanced serial ensemble methods for LF scenarios are summarized in Table 8.

Table 8. Some representative studies on a serial ensemble model for LF.

Literature (Year)	Boosting Algorithm or Correction Model	Base Learner
[129] (2024)	TrAdaBoost	LSTM
[166] (2023)	GP-based error correction	ResNet
[168] (2024)	GP-based error correction	Transformer
[169] (2024)	GRU-based error correction	GRU
[170] (2024)	GE-based error correction	RR
[85] (2024)	SVR-based similar day error correction	XGBoost

RR: Ridge regression; GE: Gaussian error.

3.5. Comparison and Summary

A summary of some advanced AI-based models for LF is given in Table 9. Their computational cost, interpretability, generalization performance, MAPE of forecasting, and typical application scenarios are compared and analyzed. From Table 9, the following is observed: (1) These advanced models can bring significant improvements on forecasting accuracy and generalization performance. Note that for multi-scenario forecasting, "MAPE" represents the average MAPE of all scenarios in the corresponding literature. (2) These models commonly have high computational costs due to their complex network architectures. Meanwhile, model interpretability is usually low. (3) Each model has its own characteristics, and forecasting performance depends on the specific scenario. Nevertheless, the Transformer-based model is quite popular due to the adopted self-attention mechanism and modular structure.

Table 9. Summary of advanced AI-based models for LF.

Advanced Model	Computational Cost	Interpretability	Generalization Capability	MAPE	Application Scenarios
Deep ResNet	High	Low to medium	Strong	1.447% [108]	High-resolution spatiotemporal data forecasting
TCN	Medium to high	Low to medium	Strong	1.123% [64]	Sequence with long-term dependency
Transformer	Very high	Medium	Very strong	1.113% [75]	Complex feature extraction, sequence with long-term dependency
DDPG-based model	Very high	Low	Medium	1.102% [122]	Sequence with sudden fluctuations or transient events
TL-based model	Medium	Low to medium	Strong	1.88% [129]	Small-sample forecasting
Parallel ensemble model	High	Low to medium	Strong	1.099% [163]	Generalized forecasting
Serial ensemble model	Medium to high	Low to medium	Strong	1.010% [52]	Non-stationary sequence

TL: Transfer learning; MAPE: mean absolute percentage error.

4. Discussion

With the increasing penetration of distributed photovoltaics, wind energy, and energy storage devices, the volatility and randomness of power generation/consumption impose new challenges on the stable and optimal operation of power grids. In addition to enhancing energy storage capacities on the main grid side, flexible scheduling and consumption of energy loads are indispensable. Under this background, accurate LF has important practical applications (e.g., power resource allocation, peak demand response, energy supply and demand optimization). Traditional energy supply–demand forecasting methods are mostly based on physical characteristics and developed through simulations with different topological structures and operation modes. Due to the increasing spatiotemporal complexity of network structures and power equipments, traditional methods are unable to meet the development needs of the current power energy system. Machine learning or deep learning can realize the approximation of complex relationships via deep nonlinear network learning with data-driven approaches. As seen from the above literature review, researchers have used advanced machine learning or deep learning methods to build various load forecasting models, achieving many beneficial results. However, facing complex supply–demand regulations at different scales, LF based on traditional machine learning or deep learning methods has the following challenges from the perspectives of data, techniques, and operations, which also indicate future directions:

(I) Data Perspective: Sample Scarcity Problem.
Supported by big data technologies, machine learning and deep learning have achieved success in the field of LF. From the perspective of data, using high-quality load characteristics as data inputs is the key to ensuring the efficiency and accuracy of these self-learning models. However, in practical scenarios, such as insufficient historical data span, differences in device operating modes, improper handling of outliers, etc., LF often faces the problem of scarce samples or limited available data. How to achieve accurate LF using a small number of samples has become one of the main challenges.

(II) Technical Perspective: Generalization Modeling Problem.
Traditional LF models for single forecasting tasks are not suitable for the requirements of new distributed energy supply–demand forecasting. From the perspective of techniques, it is quite difficult to develop a single AI model that performs the best for all forecasting scenarios. Considering the coupling and complementary relationship between distributed new energy and multiple loads, how to improve the generalization performance of the model in different forecasting scenarios has become a research trend.

(III) Operational Perspective: Model Adaptability Problem.
Current LF generally adopts the method of "ffline training, online forecasting". From the perspective of operations, the trained model lacks the ability to dynamically and adaptively adjust to the environment. Therefore, when the environment and operation parameters are significantly different from those utilized for training, accuracy of the forecasting model may decrease obviously. How to improve the environmental adaptability of a model is quite important in practical applications.

4.1. Data Perspective: Small-Sample Forecasting

For small-sample load forecasting, transfer learning provides an effective scheme. Different from traditional machine learning, transfer learning aims to generalize commonalities between different domains and tasks and enhance the learning of the target domain by acquiring knowledge from source domains and tasks. Recently, researchers have gradually introduced transfer learning into the LF field. The model transfer method is mostly em-

ployed, i.e., pre-training the model using datasets from different regions [131,136,137] and of different types (schools [134] and residences [107]) and then fine-tuning the parameters of the pre-trained model with a small amount of labeled target domain data to improve the accuracy of target task forecasting. It should be noted that the model transfer method requires high computational resources to construct pre-trained models and relies on a certain amount of target domain data for parameter adjustment, potentially leading to overfitting under extreme data scarcity. The instance transfer method offers advantages in this regard. Even with extremely limited target domain samples, favorable forecasting performance could be achieved through source domain sample weighting and reuse [132]. In 2024, Qian et al. [171] found that the instance-based boosting-type transfer method could obtain more complete target domain features compared with other transfer methods. Transferring useful samples from multiple source domains can further improve forecasting effects compared to single-source transfer methods [129,134].

The model transfer method requires relatively more target domain data and is not suitable for extreme data scarcity. The instance transfer method has a comparative advantage in this regard. Multi-source transfer has higher forecasting accuracy than single-source transfer. Currently, the optimal number of multi-source transfer lacks a reasonable selection mechanism, which can be turned into an optimization problem. Considering the respective advantages of different transfer methods, an ensemble transfer strategy combining these methods may obtain a more stable and adaptable scheme for small-sample load forecasting, which may be a potential research direction in the future.

4.2. Technical Perspective: Generalization Modeling in Different Scenarios

For short-term peak LF, bottom-up generalization modeling suffers from defects such as feature loss and reduced resolution [172], which presents another challenge for LF modeling. Tardioli et al. [173] utilized data clustering to identify typical buildings in urban building clusters and predicted the energy consumption of the target building through the energy load models of typical buildings, effectively reducing computational costs. This approach provides a new approach to problem-solving. An alternative way to improve generalization performance is integrating AI algorithms with multiple characteristics, which is also known as "hybrid forecasting". The concept of a hybrid model is rather broad, including data preprocessing-based hybrid models, optimization algorithm-based hybrid models, weighting-based hybrid models, etc. [174]. In spite of the increased complexity and computational burden, hybrid models have been shown to achieve superior forecasting performance compared with any of the individual models [175,176]. In fact, ensemble learning-based models belong to a kind of hybrid model. Fan et al. [177] applied an ensemble learning framework for the short-term forecasting of peak power loads, achieving superior performance compared with any base learner. Numerous scholars have continuously improved ensemble learning methods [122,146,151,156,161]. Note that the diversity of base learners often conflicts with modeling complexity, which indicates the importance of ensemble pruning. In 2023, Wang et al. [178] used traversal searches to identify six optimal heterogeneous base learners to form a parallel ensemble framework, reducing modeling complexity and computational cost.

The idea of the ensemble method is to create a strong learner by combining the results of weak learners so as to obtain better performance. The following should be noted: (1) It is not the case that the more base learners there are, the better the ensemble effect. For the ensemble learning strategy, accuracy, generalization, and robustness are the three main design goals. In this regard, "bias–variance–covariance decomposition" is usually considered as one of the main theoretical judgments for the effectiveness of the ensemble method. (2) Discussions on the ensemble structure are still insufficient, and the

different performance advantages of parallel, serial, and hybrid ensemble structures need to be analyzed and compared more deeply. (3) Establishing an ensemble LF framework suitable for the application of energy regulation has higher requirements in both data usage efficiency and model generalization, which is the main research trend in the field of ensemble modeling.

4.3. Operational Perspective: Online Adaptive Forecasting

Although numerous models have been proposed for LF, forecasting models often exhibit strong dependence on data and usage scenarios, with weak generalization capabilities. Current combined forecasting methods typically adopt static combinations of multiple models while lacking related research on model hyperparameter optimization and adaptation for different scenarios. Model adaptive technologies dynamically adjust behaviors and parameters by perceiving self and environmental changes, maintaining high-precision forecasting even under unexpected environmental changes. Traditional adaptive technologies often rely heavily on empirical information and prior knowledge, resulting in poor forecasting generalization performance. With the development of big data and AI, technologies such as reinforcement learning [117], quantile prediction [179], and automated feature engineering [88] enable dynamic selections of models and automatic adjustments of hyperparameters and structures, thereby accurately describing the trends in loads and enhancing forecasting accuracy, as well as generalization performance.

Due to the characteristics of obtaining the optimal policy through intensive interactions with a certain environment, reinforcement learning has potential in adaptive LF. As mentioned in Section 3.2, there are three main applications of reinforcement learning for LF tasks: direct forecasting, parameter optimization, and base learner construction/integration under the ensemble framework. However, it was found that, compared with other supervised models, direct forecasting using reinforcement learning models did not show significant superiority in terms of forecasting accuracy [117]. Hence, the latter two applications will become research trends. Among them, the rational design of states, actions, and rewards, as well as improving the training efficiency and convergence of reinforcement learning algorithms, will mainly be addressed in the future.

5. Conclusions

The volatility and randomness of power generation/consumption bring new challenges to the safe and stable operation of power grids. Therefore, research on accurate and reliable load forecasting has very important practical significance. This paper reviews studies published in the last 5 years on LF using advanced AI and data-driven technologies. The following topics are mainly included: (1) On the basis of summarizing basic preprocessing methods, some advanced preprocessing methods are reviewed, such as deep learning-based feature extraction, reinforcement learning-based feature selection, and attention mechanism-based feature extraction. In spite of greatly increased complexity compared with basic methods, these advanced methods bring significant improvements for model adaptability and accuracy. (2) Four kinds of the latest AI-based LF models are summarized and analyzed, i.e., deep learning-based models, reinforcement learning-based models, transfer learning-based models, and ensemble learning-based models. Each kind of model has its own characteristics and advantages depending on the specific forecasting scenario. (3) Existing challenges are revealed and analyzed from the perspectives of data, techniques, and operations, including small-sample forecasting, generalization modeling, and online adaptive forecasting. Some effective solutions and future directions are also proposed to address these challenges.

Author Contributions: Conceptualization, W.X. and K.L.; methodology, J.L. and W.X.; formal analysis, J.L. and W.X.; investigation, J.L. and W.X.; writing—original draft preparation, J.L., X.H. and W.X.; writing—review and editing, W.X. and K.L.; supervision, K.L.; funding acquisition, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Research and Development Program in Zhenjiang City under grant number SH2023108.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang , X.; Wang, H.; Bhandari, B.; Cheng, L. AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response. *Int. J. Precis. Eng. Manuf. Green Technol.* **2024**, *11*, 963–993. [CrossRef]
2. Mishra, S.K.; Gupta, V.K.; Kumar, R.; Swain, S.K.; Mohanta, D.K. Multi-objective optimization of economic emission load dispatch incorporating load forecasting and solar photovoltaic sources for carbon neutrality. *Electr. Power Syst. Res.* **2023**, *223*, 109700. [CrossRef]
3. Gu, L.; Wang, J.; Liu, J. A combined system based on data preprocessing and optimization algorithm for electricity load forecasting. *Comput. Ind. Eng.* **2024**, *191*, 110114. [CrossRef]
4. Eren, Y.; Küçükdemiral, İ. A comprehensive review on deep learning approaches for short-term load forecasting. *Renew. Sustain. Energy Rev.* **2024**, *189*, 114031. [CrossRef]
5. Xu, H.; Hu, F.; Liang, X.; Zhao, G.; Abugunmi, M. A framework for electricity load forecasting based on attention mechanism time series depthwise separable convolutional neural network. *Energy* **2024**, *299*, 131258. [CrossRef]
6. Mathumitha, R.; Rathika, P.; Manimala, K. Intelligent deep learning techniques for energy consumption forecasting in smart buildings: A review. *Artif. Intell. Rev.* **2024**, *57*, 35. [CrossRef]
7. Zhu, J.; Dong, H.; Zheng, W.; Li, S.; Huang, Y.; Xi, L. Review and prospect of data-driven techniques for load forecasting in integrated energy systems. *Appl. Energy* **2022**, *321*, 119269. [CrossRef]
8. Xu, H.; Fan, G.; Kuang, G.; Song, Y. Construction and application of short-term and mid-term power system load forecasting model based on hybrid deep learning. *IEEE Access* **2023**, *11*, 37494–37507. [CrossRef]
9. Lin, J.; Ma, J.; Zhu, J.; Cui, Y. Short-term load forecasting based on LSTM networks considering attention mechanism. *Int. J. Electr. Power Energy Syst.* **2022**, *137*, 107818. [CrossRef]
10. Nunekpeku, X.; Zhang, W.; Gao, J.; Adade, S.Y.S.S.; Li, H.; Chen, Q. Gel strength prediction in ultrasonicated chicken mince: Fusing near-infrared and Raman spectroscopy coupled with deep learning LSTM algorithm. *Food Control* **2025**, *168*, 110916. [CrossRef]
11. Li, K.; Xue, W.; Tan, G.; Denzer, A.S. A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms. *Build. Serv. Eng. Res. Technol.* **2020**, *41*, 108–127. [CrossRef]
12. Zhou, X.; Zhao, C.; Sun, J.; Cao, Y.; Yao, K.; Xu, M. A deep learning method for predicting lead content in oilseed rape leaves using fluorescence hyperspectral imaging. *Food Chem.* **2023**, *409*, 135251. [CrossRef]
13. Deng, J.; Ni, L.; Bai, X.; Jiang, H.; Xu, L. Simultaneous analysis of mildew degree and aflatoxin B1 of wheat by a multi-task deep learning strategy based on microwave detection technology. *LWT* **2023**, *184*, 115047. [CrossRef]
14. Chang, X.; Huang, X.; Xu, W.; Tian, X.; Wang, C.; Wang, L.; Yu, S. Monitoring of dough fermentation during Chinese steamed bread processing by near-infrared spectroscopy combined with spectra selection and supervised learning algorithm. *J. Food Process Eng.* **2021**, *44*, e13783. [CrossRef]
15. Elbeltagi, A.; Srivastava, A.; Deng, J.; Li, Z.; Raza, A.; Khadke, L.; Yu, Z.; El-Rawy, M. Forecasting vapor pressure deficit for agricultural water management using machine learning in semi-arid environments. *Agric. Water Manag.* **2023**, *283*, 108302. [CrossRef]
16. Zhang, D.; Chen, X.; Lin, Z.; Lu, M.; Yang, W.; Sun, X.; Battino, M.; Shi, J.; Huang, X.; Shi, B.; et al. Nondestructive detection of pungent and numbing compounds in spicy hotpot seasoning with hyperspectral imaging and machine learning. *Food Chem.* **2025**, *469*, 142593. [CrossRef]
17. Zhang, D.; Lin, Z.; Xuan, L.; Lu, M.; Shi, B.; Shi, J.; He, F.; Battino, M.; Zhao, L.; Zou, X. Rapid determination of geographical authenticity and pungency intensity of the red Sichuan pepper (*Zanthoxylum bungeanum*) using differential pulse voltammetry and machine learning algorithms. *Food Chem.* **2024**, *439*, 137978. [CrossRef]
18. Chen, C.; Zhu, W.; Steibel, J.; Siegford, J.; Han, J.; Norton, T. Classification of drinking and drinker-playing in pigs by a video-based deep learning method. *Biosyst. Eng.* **2020**, *196*, 1–14. [CrossRef]

19. Liu, J.; Abbas, I.; Noor, R.S. Development of deep learning-based variable rate agrochemical spraying system for targeted weeds control in strawberry crop. *Agronomy* **2021**, *11*, 1480. [CrossRef]
20. Wang, J.; Gao, Z.; Zhang, Y.; Zhou, J.; Wu, J.; Li, P. Real-time detection and location of potted flowers based on a ZED camera and a YOLO V4-tiny deep learning algorithm. *Horticulturae* **2021**, *8*, 21. [CrossRef]
21. Zhao, S.; Adade, S.Y.S.S.; Wang, Z.; Jiao, T.; Ouyang, Q.; Li, H.; Chen, Q. Deep learning and feature reconstruction assisted vis-NIR calibration method for on-line monitoring of key growth indicators during kombucha production. *Food Chem.* **2025**, *463*, 141411. [CrossRef]
22. Zhang, Z.; Lu, Y.; Yang, M.; Wang, G.; Zhao, Y.; Hu, Y. Optimal training strategy for high-performance detection model of multi-cultivar tea shoots based on deep learning methods. *Sci. Hortic.* **2024**, *328*, 112949. [CrossRef]
23. Piotrowski, P.; Baczyński, D.; Kopyt, M.; Gulczyński, T. Advanced ensemble methods using machine learning and deep learning for one-day-ahead forecasts of electric energy production in wind farms. *Energies* **2022**, *15*, 1252. [CrossRef]
24. Jeon, B.K.; Kim, E.J. Solar irradiance prediction using reinforcement learning pre-trained with limited historical data. *Energy Rep.* **2023**, *10*, 2513–2524. [CrossRef]
25. Jagait, R.K.; Fekri, M.N.; Grolinger, K.; Mir, S. Load forecasting under concept drift: Online ensemble learning with recurrent neural network and ARIMA. *IEEE Access* **2021**, *9*, 98992–99008. [CrossRef]
26. Zhu, Q.; Li, J.; Qiao, J.; Shi, M.; Wang, C. Application and Prospect of Artificial Intelligence Technology in Renewable Energy Forecasting. *Proc. CSEE* **2023**, *43*, 3027–3048.
27. Ramokone, A.; Popoola, O.; Awelewa, A.; Temitope, A. A review on behavioural propensity for building load and energy profile development—Model inadequacy and improved approach. *Sustain. Energy Technol. Assessments* **2021**, *45*, 101235. [CrossRef]
28. Patsakos, I.; Vrochidou, E.; Papakostas, G.A. A survey on deep learning for building load forecasting. *Math. Probl. Eng.* **2022**, *2022*, 1008491. [CrossRef]
29. Khalil, M.; McGough, A.S.; Pourmirza, Z.; Pazhoohesh, M.; Walker, S. Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption—A systematic review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105287. [CrossRef]
30. Han, F.; Wang, X.; Qiao, J.; Shi, M.; Pu, T. Review on artificial intelligence based load forecasting research for the new-type power system. *Proc. CSEE* **2023**, *43*, 8569–8591.
31. Abdel-Jaber, F.; Dirks, K.N. A review of cooling and heating loads predictions of residential buildings using data-driven techniques. *Buildings* **2024**, *14*, 752. [CrossRef]
32. Ma, H.; Yuan, A.; Wang, B.; Yang, C.; Dong, X.; Chen, L. Review and prospect of load forecasting based on deep learning. *High Volt. Eng.* **2025**, *51*, 1233–1250.
33. Liu, C.; Yu, H.; Liu, Y.; Zhang, L.; Li, D.; Zhang, J.; Li, X.; Sui, Y. Prediction of Anthocyanin Content in Purple-Leaf Lettuce Based on Spectral Features and Optimized Extreme Learning Machine Algorithm. *Agronomy* **2024**, *14*, 2915. [CrossRef]
34. Ahmed, N.; Assadi, M.; Zhang, Q. Investigating the impact of borehole field data's input parameters on the forecasting accuracy of multivariate hybrid deep learning models for heating and cooling. *Energy Build.* **2023**, *301*, 113706. [CrossRef]
35. Oprea, S.V.; Bâra, A. A stacked ensemble forecast for photovoltaic power plants combining deterministic and stochastic methods. *Appl. Soft Comput.* **2023**, *147*, 110781. [CrossRef]
36. Li, K.; Zhang, J.; Chen, X.; Xue, W. Building's hourly electrical load prediction based on data clustering and ensemble learning strategy. *Energy Build.* **2022**, *261*, 111943. [CrossRef]
37. Ibrahim, I.A.; Hossain, M. Short-term multivariate time series load data forecasting at low-voltage level using optimised deep-ensemble learning-based models. *Energy Convers. Manag.* **2023**, *296*, 117663. [CrossRef]
38. Jalali, S.M.J.; Osório, G.J.; Ahmadian, S.; Lotfi, M.; Campos, V.M.; Shafie-khah, M.; Khosravi, A.; Catalão, J.P. New hybrid deep neural architectural search-based ensemble reinforcement learning strategy for wind power forecasting. *IEEE Trans. Ind. Appl.* **2021**, *58*, 15–27. [CrossRef]
39. You, W.; Guo, D.; Wu, Y.; Li, W. Multiple Load Forecasting of Integrated Energy System Based on Sequential-Parallel Hybrid Ensemble Learning. *Energies* **2023**, *16*, 3268. [CrossRef]
40. Jalali, S.M.J.; Ahmadian, S.; Nakisa, B.; Khodayar, M.; Khosravi, A.; Nahavandi, S.; Islam, S.M.S.; Shafie-khah, M.; Catalão, J.P. Solar irradiance forecasting using a novel hybrid deep ensemble reinforcement learning algorithm. *Sustain. Energy Grids Netw.* **2022**, *32*, 100903. [CrossRef]
41. Lu, S.; Xu, Q.; Jiang, C.; Liu, Y.; Kusiak, A. Probabilistic load forecasting with a non-crossing sparse-group Lasso-quantile regression deep neural network. *Energy* **2022**, *242*, 122955. [CrossRef]
42. Zhu, Z.; Zhou, M.; Hu, F.; Wang, S.; Ma, J.; Gao, B.; Bian, K.; Lai, W. A day-ahead industrial load forecasting model using load change rate features and combining FA-ELM and the AdaBoost algorithm. *Energy Rep.* **2023**, *9*, 971–981. [CrossRef]
43. Li, C.; Li, G.; Wang, K.; Han, B. A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient integrated energy systems. *Energy* **2022**, *259*, 124967. [CrossRef]
44. Shi, H.; Wang, L.; Scherer, R.; Woźniak, M.; Zhang, P.; Wei, W. Short-term load forecasting based on adabelief optimized temporal convolutional network and gated recurrent unit hybrid neural network. *IEEE Access* **2021**, *9*, 66965–66981. [CrossRef]

45. Liu, R.; Chen, T.; Sun, G.; Muyeen, S.; Lin, S.; Mi, Y. Short-term probabilistic building load forecasting based on feature integrated artificial intelligent approach. *Electric Power Syst. Res.* **2022**, *206*, 107802. [CrossRef]
46. Pirbazari, A.M.; Sharma, E.; Chakravorty, A.; Elmenreich, W.; Rong, C. An ensemble approach for multi-step ahead energy forecasting of household communities. *IEEE Access* **2021**, *9*, 36218–36240. [CrossRef]
47. Dai, Y.; Yu, W.; Leng, M. A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting. *Energy* **2024**, *299*, 131458. [CrossRef]
48. Dong, Z.; Liu, J.; Liu, B.; Li, K.; Li, X. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. *Energy Build.* **2021**, *241*, 110929. [CrossRef]
49. Lu, J.; Liu, J.; Luo, Y.; Zeng, J. Small sample load forecasting method considering characteristic distribution similarity based on improved WGAN. *Control Theory Appl.* **2024**, *41*, 597–608.
50. Zhang, X.; Chau, T.K.; Chow, Y.H.; Fernando, T.; Iu, H.H.C. A novel sequence to sequence data modelling based CNN-LSTM algorithm for three years ahead monthly peak load forecasting. *IEEE Trans. Power Syst.* **2023**, *39*, 1932–1947. [CrossRef]
51. Chandrasekar, A.; Ajeya, K.; Vinatha, U. InFLuCs: Irradiance Forecasting Through Reinforcement Learning Tuned Cascaded Regressors. *IEEE Trans. Ind. Inform.* **2024**, *20*, 10912–10921. [CrossRef]
52. Hou, G.; Wang, J.; Fan, Y. Multistep short-term wind power forecasting model based on secondary decomposition, the kernel principal component analysis, an enhanced arithmetic optimization algorithm, and error correction. *Energy* **2024**, *286*, 129640. [CrossRef]
53. Yuan, Y.; Yang, Q.; Ren, J.; Mu, X.; Wang, Z.; Shen, Q.; Li, Y. Short-term power load forecasting based on SKDR hybrid model. *Electr. Eng.* **2024**, *107*, 5769–5785. [CrossRef]
54. Moon, J.; Park, S.; Rho, S.; Hwang, E. Robust building energy consumption forecasting using an online learning approach with R ranger. *J. Build. Eng.* **2022**, *47*, 103851. [CrossRef]
55. Zhao, M.; Guo, G.; Fan, L.; Han, L.; Yu, Q.; Wang, Z. Short-term natural gas load forecasting based on EL-VMD-Transformer-ResLSTM. *Sci. Rep.* **2024**, *14*, 20343.
56. Liu, H.; Chen, C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl. Energy* **2019**, *249*, 392–408. [CrossRef]
57. Zhao, Z.; Yun, S.; Jia, L.; Guo, J.; Meng, Y.; He, N.; Li, X.; Shi, J.; Yang, L. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105982. [CrossRef]
58. Junior, M.Y.; Freire, R.Z.; Seman, L.O.; Stefenon, S.F.; Mariani, V.C.; dos Santos Coelho, L. Optimized hybrid ensemble learning approaches applied to very short-term load forecasting. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109579.
59. Heo, S.; Nam, K.; Loy-Benitez, J.; Yoo, C. Data-driven hybrid model for forecasting wastewater influent loads based on multimodal and ensemble deep learning. *IEEE Trans. Ind. Inform.* **2020**, *17*, 6925–6934. [CrossRef]
60. Junliang, L.; Runhai, J.; Shuangkun, W.; Hui, H. An ensemble load forecasting model based on online error updating. *Proc. CSEE* **2022**, *43*, 1402–1412.
61. Zhang, Z.; Zhang, C.; Dong, Y.; Hong, W.C. Bi-directional gated recurrent unit enhanced twin support vector regression with seasonal mechanism for electric load forecasting. *Knowl.-Based Syst.* **2025**, *310*, 112943. [CrossRef]
62. Tian, Z.; Liu, W.; Jiang, W.; Wu, C. Cnns-transformer based day-ahead probabilistic load forecasting for weekends with limited data availability. *Energy* **2024**, *293*, 130666. [CrossRef]
63. Zhang, Q.; Wu, J.; Ma, Y.; Li, G.; Ma, J.; Wang, C. Short-term load forecasting method with variational mode decomposition and stacking model fusion. *Sustain. Energy Grids Netw.* **2022**, *30*, 100622. [CrossRef]
64. Cao, Z.; Wang, J.; Xia, Y. Combined electricity load-forecasting system based on weighted fuzzy time series and deep neural networks. *Eng. Appl. Artif. Intell.* **2024**, *132*, 108375. [CrossRef]
65. Wang, J.; Liu, H.; Zheng, G.; Li, Y.; Yin, S. Short-Term Load Forecasting Based on Outlier Correction, Decomposition, and Ensemble Reinforcement Learning. *Energies* **2023**, *16*, 4401. [CrossRef]
66. Yang, D.; Guo, J.E.; Sun, S.; Han, J.; Wang, S. An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. *Appl. Energy* **2022**, *306*, 117992. [CrossRef]
67. Memarzadeh, G.; Keynia, F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electr. Power Syst. Res.* **2021**, *192*, 106995. [CrossRef]
68. Liu, H.; Yu, C.; Wu, H.; Duan, Z.; Yan, G. A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting. *Energy* **2020**, *202*, 117794. [CrossRef]
69. Malhan, P.; Mittal, M. A novel ensemble model for long-term forecasting of wind and hydro power generation. *Energy Convers. Manag.* **2022**, *251*, 114983. [CrossRef]
70. Ribeiro, M.H.D.M.; da Silva, R.G.; Moreno, S.R.; Mariani, V.C.; dos Santos Coelho, L. Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107712. [CrossRef]
71. Liu, H.; Yu, C.; Yu, C. A new hybrid model based on secondary decomposition, reinforcement learning and SRU network for wind turbine gearbox oil temperature forecasting. *Measurement* **2021**, *178*, 109347. [CrossRef]

72. Ye, L.; Li, Y.; Pei, M.; Zhao, Y.; Li, Z.; Lu, P. A novel integrated method for short-term wind power forecasting based on fluctuation clustering and history matching. *Appl. Energy* **2022**, *327*, 120131. [CrossRef]
73. Ribeiro, M.H.D.M.; da Silva, R.G.; Ribeiro, G.T.; Mariani, V.C.; dos Santos Coelho, L. Cooperative ensemble learning model improves electric short-term load forecasting. *Chaos Solitons Fractals* **2023**, *166*, 112982. [CrossRef]
74. Zhou, S.; Li, Y.; Guo, Y.; Yang, X.; Shahidehpour, M.; Deng, W.; Mei, Y.; Ren, L.; Liu, Y.; Kang, T.; et al. A Load Forecasting Framework Considering Hybrid Ensemble Deep Learning with Two-Stage Load Decomposition. *IEEE Trans. Ind. Appl.* **2024**, *60*, 4568–4582. [CrossRef]
75. Zhang, M.; Xiao, G.; Lu, J.; Liu, Y.; Chen, H.; Yang, N. Robust load feature extraction based secondary VMD novel short-term load demand forecasting framework. *Electric Power Syst. Res.* **2025**, *239*, 111198. [CrossRef]
76. Sibtain, M.; Li, X.; Saleem, S.; Asad, M.S.; Tahir, T.; Apaydin, H.; et al. A multistage hybrid model ICEEMDAN-SE-VMD-RDPG for a multivariate solar irradiance forecasting. *IEEE Access* **2021**, *9*, 37334–37363. [CrossRef]
77. Zhao, X.; Sun, B.; Geng, R. A new distributed decomposition–reconstruction–ensemble learning paradigm for short-term wind power prediction. *J. Clean. Prod.* **2023**, *423*, 138676. [CrossRef]
78. Li, F.; Zheng, H.; Li, X.; Yang, F. Day-ahead city natural gas load forecasting based on decomposition-fusion technique and diversified ensemble learning model. *Appl. Energy* **2021**, *303*, 117623. [CrossRef]
79. Tian, J.; Li, K.; Xue, W. An adaptive ensemble predictive strategy for multiple scale electrical energy usages forecasting. *Sustain. Cities Soc.* **2021**, *66*, 102654. [CrossRef]
80. Yan, Q.; Lu, Z.; Liu, H.; He, X.; Zhang, X.; Guo, J. Short-term prediction of integrated energy load aggregation using a bi-directional simple recurrent unit network with feature-temporal attention mechanism ensemble learning model. *Appl. Energy* **2024**, *355*, 122159. [CrossRef]
81. Lin, Z.; Lin, T.; Li, J.; Li, C. A novel short-term multi-energy load forecasting method for integrated energy system based on two-layer joint modal decomposition and dynamic optimal ensemble learning. *Appl. Energy* **2025**, *378*, 124798. [CrossRef]
82. Sun, X.; Li, J.; Zeng, B.; Gong, D.; Lian, Z. Small-sample day-ahead power load forecasting of integrated energy system based on feature transfer learning. *Control Theory Appl.* **2021**, *38*, 63–72.
83. Wu, D.; Hur, K.; Xiao, Z. A GAN-enhanced ensemble model for energy consumption forecasting in large commercial buildings. *IEEE Access* **2021**, *9*, 158820–158830. [CrossRef]
84. Zhang, S.; Chen, R.; Cao, J.; Tan, J. A CNN and LSTM-based multi-task learning architecture for short and medium-term electricity load forecasting. *Electr. Power Syst. Res.* **2023**, *222*, 109507. [CrossRef]
85. Ye, W.; Yang, D.; Tang, C.; Wang, W.; Liu, G. Combined prediction of wind power in extreme weather based on time series adversarial generation networks. *IEEE Access* **2024**, *12*, 102660–102669. [CrossRef]
86. Liu, L.; Fu, Q.; Lu, Y.; Wang, Y.; Wu, H.; Chen, J. CorrDQN-FS: A two-stage feature selection method for energy consumption prediction via deep reinforcement learning. *J. Build. Eng.* **2023**, *80*, 108044. [CrossRef]
87. Liu, Y.; Wang, J.; Liu, L. Physics-informed reinforcement learning for probabilistic wind power forecasting under extreme events. *Appl. Energy* **2024**, *376*, 124068. [CrossRef]
88. Wang, X.; Wang, H.; Li, S.; Jin, H. A reinforcement learning-based online learning strategy for real-time short-term load forecasting. *Energy* **2024**, *305*, 132344. [CrossRef]
89. Li, A.; Xiao, F.; Zhang, C.; Fan, C. Attention-based interpretable neural network for building cooling load prediction. *Appl. Energy* **2021**, *299*, 117238. [CrossRef]
90. Yu, C.; Yan, G.; Yu, C.; Zhang, Y.; Mi, X. A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy* **2023**, *263*, 126034.
91. Yao, H.; Qu, P.; Qin, H.; Lou, Z.; Wei, X.; Song, H. Multidimensional electric power parameter time series forecasting and anomaly fluctuation analysis based on the AFFC-GLDA-RL method. *Energy* **2024**, *313*, 134180. [CrossRef]
92. Zhou, X.; Sun, J.; Tian, Y.; Lu, B.; Hang, Y.; Chen, Q. Hyperspectral technique combined with deep learning algorithm for detection of compound heavy metals in lettuce. *Food Chem.* **2020**, *321*, 126503. [CrossRef]
93. Tian, Y.; Sun, J.; Zhou, X.; Yao, K.; Tang, N. Detection of soluble solid content in apples based on hyperspectral technology combined with deep learning algorithm. *J. Food Process. Preserv.* **2022**, *46*, e16414. [CrossRef]
94. Rafi, S.H.; Deeba, S.R.; Hossain, E.; et al. A short-term load forecasting method using integrated CNN and LSTM network. *IEEE Access* **2021**, *9*, 32436–32448. [CrossRef]
95. Chen, Y.; Yu, Z.; Han, Z.; Sun, W.; He, L. A decision-making system for cotton irrigation based on reinforcement learning strategy. *Agronomy* **2023**, *14*, 11. [CrossRef]
96. Xue, W.; Jia, N.; Zhao, M. Multi-agent deep reinforcement learning based HVAC control for multi-zone buildings considering zone-energy-allocation optimization. *Energy Build.* **2025**, *329*, 115241. [CrossRef]
97. Niu, D.; Yu, M.; Sun, L.; Gao, T.; Wang, K. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Appl. Energy* **2022**, *313*, 118801. [CrossRef]

98. Moreno, S.R.; Mariani, V.C.; dos Santos Coelho, L. Hybrid multi-stage decomposition with parametric model applied to wind speed forecasting in Brazilian Northeast. *Renew. Energy* **2021**, *164*, 1508–1526. [CrossRef]
99. Yang, G.; Zheng, H.; Zhang, H.; Jia, R. Short-term load forecasting based on Holt-Winters exponential smoothing and temporal convolutional network. *Autom. Electr. Power Syst.* **2022**, *46*, 73–82.
100. Fan, G.F.; Zhang, L.Z.; Yu, M.; Hong, W.C.; Dong, S.Q. Applications of random forest in multivariable response surface for short-term load forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *139*, 108073. [CrossRef]
101. Pinto, T.; Praça, I.; Vale, Z.; Silva, J. Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing* **2021**, *423*, 747–755. [CrossRef]
102. Ge, Q.; Guo, C.; Jiang, H.; Lu, Z.; Yao, G.; Zhang, J.; Hua, Q. Industrial power load forecasting method based on reinforcement learning and PSO-LSSVM. *IEEE Trans. Cybern.* **2022**, *52*, 1112–1124. [CrossRef]
103. Zhang, T.; Tang, Z.; Wu, J.; Du, X.; Chen, K. Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. *Electric Power Syst. Res.* **2022**, *205*, 107762. [CrossRef]
104. Li, H.; Sheng, W.; Adade, S.Y.S.S.; Nunekpeku, X.; Chen, Q. Investigation of heat-induced pork batter quality detection and change mechanisms using Raman spectroscopy coupled with deep learning algorithms. *Food Chem.* **2024**, *461*, 140798. [CrossRef]
105. Xu, M.; Sun, J.; Cheng, J.; Yao, K.; Wu, X.; Zhou, X. Non-destructive prediction of total soluble solids and titratable acidity in Kyoho grape using hyperspectral imaging and deep learning algorithm. *Int. J. Food Sci. Technol.* **2023**, *58*, 9–21. [CrossRef]
106. Pan, Y.; Zhang, Y.; Wang, X.; Gao, X.X.; Hou, Z. Low-cost livestock sorting information management system based on deep learning. *Artif. Intell. Agric.* **2023**, *9*, 110–126. [CrossRef]
107. Zhang, Z.; Zhao, P.; Wang, P.; Lee, W.J. Transfer learning featured short-term combining forecasting model for residential loads with small sample sets. *IEEE Trans. Ind. Appl.* **2022**, *58*, 4279–4288. [CrossRef]
108. Chen, Q.; Zhang, W.; Zhu, K.; Zhou, D.; Dai, H.; Wu, Q. A novel trilinear deep residual network with self-adaptive Dropout method for short-term load forecasting. *Expert Syst. Appl.* **2021**, *182*, 115272. [CrossRef]
109. Hua, H.; Liu, M.; Li, Y.; Deng, S.; Wang, Q. An ensemble framework for short-term load forecasting based on parallel CNN and GRU with improved ResNet. *Electr. Power Syst. Res.* **2023**, *216*, 109057. [CrossRef]
110. Su, H.Y.; Lai, C.C. Towards Improved Load Forecasting in Smart Grids: A Robust Deep Ensemble Learning Framework. *IEEE Trans. Smart Grid* **2024**, *15*, 4292–4296. [CrossRef]
111. Zheng, G.q.; Kong, L.r.; Su, Z.e.; Hu, M.s.; Wang, G.d. Approach for short-term power load prediction utilizing the ICEEMDAN–LSTM–TCN–bagging model. *J. Electr. Eng. Technol.* **2025**, *20*, 231–243. [CrossRef]
112. Gao, S.; Liu, Y.; Wang, J.; Wang, Z.; Wenjun, X.; Yue, R.; Cui, R.; Liu, Y.; Fan, X. Short-term residential load forecasting via transfer learning and multi-attention fusion for EVs' coordinated charging. *Int. J. Electr. Power Energy Syst.* **2025**, *164*, 110349. [CrossRef]
113. Xie, F.; Guo, Z.; Li, T.; Feng, Q.; Zhao, C. Dynamic Task Planning for Multi-Arm Harvesting Robots Under Multiple Constraints Using Deep Reinforcement Learning. *Horticulturae* **2025**, *11*. [CrossRef]
114. Chen, Y.; Lin, M.; Yu, Z.; Sun, W.; Fu, W.; He, L. Enhancing cotton irrigation with distributional actor—critic reinforcement learning. *Agric. Water Manag.* **2025**, *307*, 109194. [CrossRef]
115. Zhang, W.; Chen, Q.; Yan, J.; Zhang, S.; Xu, J. A novel asynchronous deep reinforcement learning model with adaptive early forecasting method and reward incentive mechanism for short-term load forecasting. *Energy* **2021**, *236*, 121492. [CrossRef]
116. Fu, Q.; Li, K.; Chen, J.; Wang, J.; Lu, Y.; Wang, Y. Building energy consumption prediction using a deep-forest-based DQN method. *Buildings* **2022**, *12*, 131. [CrossRef]
117. Zhou, X.; Lin, W.; Kumar, R.; Cui, P.; Ma, Z. A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption. *Appl. Energy* **2022**, *306*, 118078. [CrossRef]
118. He, X.; Zhao, W.; Gao, Z.; Zhang, L.; Zhang, Q.; Li, X. Short-term load forecasting by GRU neural network and DDPG algorithm for adaptive optimization of hyperparameters. *Electr. Power Syst. Res.* **2025**, *238*, 111119. [CrossRef]
119. Kosana, V.; Teeparthi, K.; Madasthu, S.; Kumar, S. A novel reinforced online model selection using Q-learning technique for wind speed prediction. *Sustain. Energy Technol. Assessments* **2022**, *49*, 101780. [CrossRef]
120. Sun, J.; Gong, M.; Zhao, Y.; Han, C.; Jing, L.; Yang, P. A hybrid deep reinforcement learning ensemble optimization model for heat load energy-saving prediction. *J. Build. Eng.* **2022**, *58*, 105031. [CrossRef]
121. Wang, J.; Fu, J.; Chen, B. Multi-model fusion photovoltaic power generation prediction method based on reinforcement learning. *Acta Energiae Solaris Sin.* **2024**, *45*, 382–388.
122. Ren, X.; Tian, X.; Wang, K.; Yang, S.; Chen, W.; Wang, J. Enhanced load forecasting for distributed multi-energy system: A stacking ensemble learning method with deep reinforcement learning and model fusion. *Energy* **2025**, *319*, 135031. [CrossRef]
123. Liu, T.; Tan, Z.; Xu, C.; Chen, H.; Li, Z. Study on deep reinforcement learning techniques for building energy consumption forecasting. *Energy Build.* **2020**, *208*, 109675. [CrossRef]
124. Nguyen, N.A.; Dang, T.D.; Verdú, E.; Kumar Solanki, V. Short-term forecasting electricity load by long short-term memory and reinforcement learning for optimization of hyper-parameters. *Evol. Intell.* **2023**, *16*, 1729–1746. [CrossRef]

125. Li, Z.; Xu, B.; Zhang, J.; Yang, J.; Guo, Y. Short-term load optimal weighted forecasting strategy based on TD3 variable length time window. *Electr. Power Constr.* **2024**, *45*, 140–148.
126. Dabbaghjamanesh, M.; Moeini, A.; Kavousi-Fard, A. Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique. *IEEE Trans. Ind. Inform.* **2020**, *17*, 4229–4237. [CrossRef]
127. Yin, S.; Liu, H. Wind power prediction based on outlier correction, ensemble reinforcement learning, and residual correction. *Energy* **2022**, *250*, 123857. [CrossRef]
128. Pannakkong, W.; Vinh, V.T.; Tuyen, N.N.M.; Buddhakulsomsiri, J. A Reinforcement Learning Approach for Ensemble Machine Learning Models in Peak Electricity Forecasting. *Energies* **2023**, *16*, 5099. [CrossRef]
129. Wei, B.; Li, K.; Zhou, S.; Xue, W.; Tan, G. An instance based multi-source transfer learning strategy for building's short-term electricity loads prediction under sparse data scenarios. *J. Build. Eng.* **2024**, *85*, 108713. [CrossRef]
130. Yang, F.; Sun, J.; Cheng, J.; Fu, L.; Wang, S.; Xu, M. Detection of starch in minced chicken meat based on hyperspectral imaging technique and transfer learning. *J. Food Process Eng.* **2023**, *46*, e14304. [CrossRef]
131. Wu, D.; Xu, Y.T.; Jenkin, M.; Wang, J.; Li, H.; Liu, X.; Dudek, G. Short-term load forecasting with deep boosting transfer regression. In Proceedings of the ICC 2022-IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; pp. 5530–5536.
132. Lu, Y.; Wang, G.; Huang, S. A short-term load forecasting model based on mixup and transfer learning. *Electr. Power Syst. Res.* **2022**, *207*, 107837. [CrossRef]
133. Zhou, Z.; Xu, Y.; Ren, C. A transfer learning method for forecasting masked-load with behind-the-meter distributed energy resources. *IEEE Trans. Smart Grid* **2022**, *13*, 4961–4964. [CrossRef]
134. Peng, C.; Tao, Y.; Chen, Z.; Zhang, Y.; Sun, X. Multi-source transfer learning guided ensemble LSTM for building multi-load forecasting. *Expert Syst. Appl.* **2022**, *202*, 117194. [CrossRef]
135. Liu, Y.; Wang, J. Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting. *Appl. Energy* **2022**, *312*, 118729. [CrossRef]
136. Yang, Q.; Lin, Y.; Kuang, S.; Wang, D. A novel short-term load forecasting approach for data-poor areas based on K-MIFS-XGBoost and transfer-learning. *Electr. Power Syst. Res.* **2024**, *229*, 110151. [CrossRef]
137. Xiao, L.; Bai, Q.; Wang, B. A dynamic multi-model transfer based short-term load forecasting. *Appl. Soft Comput.* **2024**, *159*, 111627. [CrossRef]
138. Wei, N.; Yin, C.; Yin, L.; Tan, J.; Liu, J.; Wang, S.; Qiao, W.; Zeng, F. Short-term load forecasting based on WM algorithm and transfer learning model. *Appl. Energy* **2024**, *353*, 122087. [CrossRef]
139. Cheng, M.; Zhai, J.; Ma, J.; Lu, L.; Jin, E. Transfer learning based CNN-GRU short-term power load forecasting method. *Eng. J. Wuhan Univ.* **2024**, *57*, 812–820.
140. Li, K.; Liu, Y.; Chen, L.; Xue, W. Data efficient indoor thermal comfort prediction using instance based transfer learning method. *Energy Build.* **2024**, *306*, 113920. [CrossRef]
141. Li, K.; Chen, L.; Luo, Y.; He, X. An ensemble strategy for transfer learning based human thermal comfort prediction: field experimental study. *Energy Build.* **2025**, *330*, 115344. [CrossRef]
142. Cheng, J.; Sun, J.; Yao, K.; Xu, M.; Wang, S.; Fu, L. Hyperspectral technique combined with stacking and blending ensemble learning method for detection of cadmium content in oilseed rape leaves. *J. Sci. Food Agric.* **2023**, *103*, 2690–2699. [CrossRef]
143. Yu, J.; Zhangzhong, L.; Lan, R.; Zhang, X.; Xu, L.; Li, J. Ensemble Learning Simulation Method for Hydraulic Characteristic Parameters of Emitters Driven by Limited Data. *Agronomy* **2023**, *13*, 986. [CrossRef]
144. Raza, A.; Hu, Y.; Lu, Y. Improving carbon flux estimation in tea plantation ecosystems: A machine learning ensemble approach. *Eur. J. Agron.* **2024**, *160*, 127297. [CrossRef]
145. Jin, H.; Li, Y.; Wang, B.; Yang, B.; Jin, H.; Cao, Y. Adaptive forecasting of wind power based on selective ensemble of offline global and online local learning. *Energy Convers. Manag.* **2022**, *271*, 116296. [CrossRef]
146. Shi, J.; Li, C.; Yan, X. Artificial intelligence for load forecasting: A stacking learning approach based on ensemble diversity regularization. *Energy* **2023**, *262*, 125295. [CrossRef]
147. Hu, Y.; Qu, B.; Wang, J.; Liang, J.; Wang, Y.; Yu, K.; Li, Y.; Qiao, K. Short-term load forecasting using multimodal evolutionary algorithm and random vector functional link network based ensemble learning. *Appl. Energy* **2021**, *285*, 116415. [CrossRef]
148. Sun, F.; Jin, T. A hybrid approach to multi-step, short-term wind speed forecasting using correlated features. *Renew. Energy* **2022**, *186*, 742–754. [CrossRef]
149. Yang, Q.; Tian, Z. A hybrid load forecasting system based on data augmentation and ensemble learning under limited feature availability. *Expert Syst. Appl.* **2025**, *261*, 125567. [CrossRef]
150. Che, J.; Yuan, F.; Zhu, S.; Yang, Y. An adaptive ensemble framework with representative subset based weight correction for short-term forecast of peak power load. *Appl. Energy* **2022**, *328*, 120156. [CrossRef]
151. Hadjout, D.; Torres, J.; Troncoso, A.; Sebaa, A.; Martínez-Álvarez, F. Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market. *Energy* **2022**, *243*, 123060. [CrossRef]

152. Von Krannichfeldt, L.; Wang, Y.; Zufferey, T.; Hug, G. Online ensemble approach for probabilistic wind power forecasting. *IEEE Trans. Sustain. Energy* **2021**, *13*, 1221–1233. [CrossRef]
153. Massaoudi, M.; Abu-Rub, H.; Refaat, S.S.; Trabelsi, M.; Chihi, I.; Oueslati, F.S. Enhanced deep belief network based on ensemble learning and tree-structured of Parzen estimators: An optimal photovoltaic power forecasting method. *IEEE Access* **2021**, *9*, 150330–150344. [CrossRef]
154. Fan, C.; Li, Y.; Yi, L.; Xiao, L.; Qu, X.; Ai, Z. Multi-objective LSTM ensemble model for household short-term load forecasting. *Memetic Comput.* **2022**, *14*, 115–132. [CrossRef]
155. He, Y.; Xiao, J.; An, X.; Cao, C.; Xiao, J. Short-term power load probability density forecasting based on GLRQ-Stacking ensemble learning method. *Int. J. Electr. Power Energy Syst.* **2022**, *142*, 108243. [CrossRef]
156. He, Y.; Zhang, H.; Dong, Y.; Wang, C.; Ma, P. Residential net load interval prediction based on stacking ensemble learning. *Energy* **2024**, *296*, 131134. [CrossRef]
157. Li, K.; Tian, J.; Xue, W.; Tan, G. Short-term electricity consumption prediction for buildings using data-driven swarm intelligence based ensemble model. *Energy Build.* **2021**, *231*, 110558. [CrossRef]
158. Shuyin, C.; Xinjie, W. Multivariate load forecasting in integrated energy system based on maximal information coefficient and multi-objective Stacking ensemble learning. *Electr. Power Autom. Equip.* **2022**, *42*, 32–39.
159. Li, Y.; Zhang, S.; Hu, R.; Lu, N. A meta-learning based distribution system load forecasting model selection framework. *Appl. Energy* **2021**, *294*, 116991. [CrossRef]
160. Shen, Q.; Mo, L.; Liu, G.; Zhou, J.; Zhang, Y.; Ren, P. Short-term load forecasting based on multi-scale ensemble deep learning neural network. *IEEE Access* **2023**. [CrossRef]
161. Massaoudi, M.; Refaat, S.S.; Chihi, I.; Trabelsi, M.; Oueslati, F.S.; Abu-Rub, H. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. *Energy* **2021**, *214*, 118874. [CrossRef]
162. Shojaei, T.; Mokhtar, A. Forecasting energy consumption with a novel ensemble deep learning framework. *J. Build. Eng.* **2024**, *96*, 110452. [CrossRef]
163. Gong, J.; Qu, Z.; Zhu, Z.; Xu, H.; Yang, Q. Ensemble models of TCN-LSTM-LightGBM based on ensemble learning methods for short-term electrical load forecasting. *Energy* **2025**, *318*, 134757. [CrossRef]
164. Meng, H.; Han, L.; Hou, L. An ensemble learning-based short-term load forecasting on small datasets. In Proceedings of the 2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Online, 12–15 September 2022; pp. 346–350.
165. Ma, C.; Pan, S.; Cui, T.; Liu, Y.; Cui, Y.; Wang, H.; Wan, T. Energy Consumption Prediction for Office Buildings: Performance Evaluation and Application of Ensemble Machine Learning Techniques. *J. Build. Eng.* **2025**, *102*, 112021. [CrossRef]
166. Zhao, P.; Cao, D.; Wang, Y.; Chen, Z.; Hu, W. Gaussian process-aided transfer learning for probabilistic load forecasting against anomalous events. *IEEE Trans. Power Syst.* **2023**, *38*, 2962–2965. [CrossRef]
167. Ye, L.; Dai, B.; Li, Z.; Pei, M.; Zhao, Y.; Lu, P. An ensemble method for short-term wind power prediction considering error correction strategy. *Appl. Energy* **2022**, *322*, 119475. [CrossRef]
168. Zhao, P.; Hu, W.; Cao, D.; Zhang, Z.; Huang, Y.; Dai, L.; Chen, Z. Probabilistic multienergy load forecasting based on hybrid attention-enabled transformer network and gaussian process-aided residual learning. *IEEE Trans. Ind. Inform.* **2024**, *20*, 8379–8393. [CrossRef]
169. Fan, C.; Nie, S.; Xiao, L.; Yi, L.; Wu, Y.; Li, G. A multi-stage ensemble model for power load forecasting based on decomposition, error factors, and multi-objective optimization algorithm. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109620. [CrossRef]
170. Fan, C.; Nie, S.; Xiao, L.; Yi, L.; Li, G. Short-term industrial load forecasting based on error correction and hybrid ensemble learning. *Energy Build.* **2024**, *313*, 114261. [CrossRef]
171. Qian, F.; Ruan, Y.; Lu, H.; Meng, H.; Xu, T. Enhancing source domain availability through data and feature transfer learning for building power load forecasting. In *Building Simulation*; Springer: Berlin/Heidelberg, Germany, 2024; Volume 17, pp. 625–638.
172. Moghadam, S.T.; Delmastro, C.; Corgnati, S.P.; Lombardi, P. Urban energy planning procedure for sustainable development in the built environment: A review of available spatial approaches. *J. Clean. Prod.* **2017**, *165*, 811–827. [CrossRef]
173. Tardioli, G.; Kerrigan, R.; Oates, M.; O'Donnell, J.; Finn, D.P. Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. *Build. Environ.* **2018**, *140*, 90–106. [CrossRef]
174. Feng, C.; Cui, M.; Hodge, B.M.; Zhang, J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* **2017**, *190*, 1245–1257. [CrossRef]
175. Carneiro, T.C.; Rocha, P.A.; Carvalho, P.C.; Fernández-Ramírez, L.M. Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. *Appl. Energy* **2022**, *314*, 118936. [CrossRef]
176. Lu, S.; Bao, T. Short-term electricity load forecasting based on NeuralProphet and CNN-LSTM. *IEEE Access* **2024**, *12*, 76870–76879. [CrossRef]

177. Fan, C.; Xiao, F.; Wang, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl. Energy* **2014**, *127*, 1–10. [CrossRef]
178. Wang, Z.; Liang, Z.; Zeng, R.; Yuan, H.; Srinivasan, R.S. Identifying the optimal heterogeneous ensemble learning model for building energy prediction using the exhaustive search method. *Energy Build.* **2023**, *281*, 112763. [CrossRef]
179. Zhang, Y.; Wen, H.; Wu, Q.; Ai, Q. Optimal adaptive prediction intervals for electricity load forecasting in distribution systems via reinforcement learning. *IEEE Trans. Smart Grid* **2022**, *14*, 3259–3270. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.