

# Hyperparameter Tuning of Load-Forecasting Models Using Metaheuristic Optimization Algorithms—A Systematic Review

Umme Mumtahina <sup>1,\*</sup>, Sanath Alahakoon <sup>2</sup> and Peter Wolfs <sup>1</sup>

<sup>1</sup> School of Engineering and Technology, CQ University, Rockhampton, QLD 4701, Australia; p.wolfs@cqu.edu.au

<sup>2</sup> School of Engineering and Technology, CQ University, Gladstone, QLD 4680, Australia; s.alahakoon@cqu.edu.au

\* Correspondence: u.mumtahina@cqu.edu.au

**Abstract:** Load forecasting is an integral part of the power industries. Load-forecasting techniques should minimize the percentage error while prediction future demand. This will inherently help utilities have an uninterrupted power supply. In addition to that, accurate load forecasting can result in saving large amounts of money. This article provides a systematic review based on the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) framework. This article presents a complete framework for short-term load forecasting using metaheuristic algorithms. This framework consists of three sub-layers: the data-decomposition layer, the forecasting layer, and the optimization layer. The data-decomposition layer decomposes the input data series to extract important features. The forecasting layer is used to predict the result, which involves different statistical and machine-learning models. The optimization layer optimizes the parameters of forecasting methods to improve the accuracy and stability of the forecasting model using different metaheuristic algorithms. Single models from the forecasting layer can predict the results. However, they come with their limitations, such as low accuracy, high computational burden, stuck to local minima, etc. To improve the prediction accuracy, the hyperparameters of these models need to be tuned properly. Metaheuristic algorithms can be used to tune these hyperparameters considering their interdependencies. Hybrid models combining the three-layer methods can perform better by overcoming the issues of premature convergence and trapping into a local minimum solution. A quantitative analysis of different metaheuristic algorithms and deep-learning forecasting methods is presented. Some of the most common evaluation indices that are used to evaluate the performance of the forecasting models are discussed. Furthermore, a taxonomy of different state-of-the-art articles is provided, discussing their advantages, limitations, contributions, and evaluation indices. A future direction is provided for researchers to deal with hyperparameter tuning.

**Keywords:** optimization; short-term load forecasting; metaheuristic algorithms; evaluation indices; hyperparameters

**MSC:** 37N40; 46N10; 68R01



**Citation:** Mumtahina, U.; Alahakoon, S.; Wolfs, P. Hyperparameter Tuning of Load-Forecasting Models Using Metaheuristic Optimization Algorithms—A Systematic Review. *Mathematics* **2024**, *12*, 3353. <https://doi.org/10.3390/math12213353>

Academic Editor: Sai-Ho Chung

Received: 9 September 2024

Revised: 5 October 2024

Accepted: 23 October 2024

Published: 25 October 2024



**Copyright:** © 2024 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

Electricity plays a pivotal role in every person's day-to-day life and with the population growth, electricity demand is also growing [1]. A competitive market and deregulated structure within the modern power system is introduced, which basically reshaped the 1990s monopolistic behavior of power sectors [2,3]. A balance is needed between demand and supply to maintain the resilience of the power system. However, maintaining balance is becoming challenging due to incorporating distributed renewable energy sources, energy storage, linear and nonlinear loads, etc. Therefore, electricity demand faces a change due to these characteristics, which also requires a change in electricity supply. That is why

predicting present and future load demand is necessary to maintain an uninterrupted power supply to the customer side and enhance system reliability and security [4]. Load forecasting is used to predict future demand which is one of the prime interests of scientists and researchers as it leads the way in how power system operation, energy trading, and planning happens [1]. The utility industry can have an idea of future demand using load forecasting, which helps to mitigate the difference between the generation side and the demand side. As power generation is an expensive process, load prediction helps to avoid the under-power or over-power generation cost [5]. Load-forecasting methods are classified into four categories depending on different time horizons [6–10]:

- Very short-term load forecasting (VSTLF): VSTLF is performed a few minutes to an hour ahead. The method is used for real-time prediction. If there is a fast variation in load profile, the method can be used in high-speed applications [11,12]. The method is utilized in energy prediction and operation and maintenance of power utility.
- Short-term load forecasting (STLF): Load prediction is based on 30 min to 2 weeks before using STLF. The utility industry uses this method for daily operations and scheduling the generation and transmission of electric power. If the prediction error is very small, it can save the utility industry from a deficit of generation capacity or wasting resources [13,14].
- Medium-term load forecasting (MTLF): The MTLF method is used to predict the load for a period of a month to a year. The utility industry uses this method for revenue assessment, energy trading, and outage planning [15,16].
- Long-term load forecasting (LTLF): LTLF ranges from a year up to 20 years or even several decades. The method is important for strategic planning, expansion of resources, and future investment [16–18].

As load forecasting has been a prime interest for researchers to improve the efficiency of power generation, many state-of-the-art methods have been investigated. Several challenges are hindering the accuracy of these methods. As weather is unpredictable, it is one of the main challenges faced by the researchers while developing a load-forecasting method. Metering systems such as smart and traditional systems impact load forecasting. The utility industry should develop a different forecasting algorithm for each of the metering systems to avoid a forecasting error. Data collection is also a challenging factor that affects the load-forecasting model. When designing a model, the transient behavior of the network or unexpected faults must be considered. Again, the utility company should consider an acceptable margin of error while deciding on a forecasting model.

In this article, short-term load-forecasting models are considered. For load dispatching, the STLF requires more accurate forecasting results than MTLF and LTLF [19]. For energy saving, cost reduction, fine scheduling management, and security enhancement, accurate prediction of STLF is necessary [20]. The single STLF models demonstrate the behavior of overfitting or trapping into local solutions. Metaheuristic algorithms can be combined with single or hybrid STLF models to overcome these issues. Again, the emphasis is given to metaheuristic algorithms because of their ability for hyperparameter optimization, which is eventually required to obtain better prediction accuracy. This review paper has found 160 articles based on short-term load forecasting combined with metaheuristic algorithms. As a result, these algorithms play a vital role in enhancing the accuracy and stability of load-forecasting methods. This article has investigated ten review research papers that discuss the various aspects of load forecasting and different methods applied to short-term load forecasting. However, these articles did not discuss the effects of metaheuristic algorithms on load forecasting. Therefore, it is important to establish a guideline for the researchers to access the metaheuristic algorithms for hyperparameter tuning of combined models.

This paper is outlined as follows: Section 2 gives an overview of the research scope, including research gaps, challenges, and contributions. Section 3 provides a methodology for this study. Section 4 process discusses the factors that affect load forecasting. Section 5 discusses the advantages of metaheuristic algorithms on load forecasting. Section 6 discusses different evaluation indices used for prediction. Section 7 proposes a complete

framework for load forecasting using metaheuristic algorithms. Section 8 provides a generalized approach for STLF using metaheuristic algorithms. Section 9 gives an overview of the existing literature, and Section 10 and 11 discusses results and research findings and recommendations, which follow with a conclusion.

## 2. Research Scope

### 2.1. Research Gaps

As load forecasting has been one of the primary issues faced by the utility industries, several review articles have been published over the years. The authors have provided a comprehensive review of load forecasting for different generation modalities [21]. This paper explores the existing literature and discusses the common trend followed in the literature and advancement in the area.

A systematic review is presented to guide the researchers in choosing the efficient model for a particular case [22]. A comparison based on input, outputs, data size, and error type is made among the existing literature articles. It is found that for STLF, ML algorithms and time series techniques are more efficient than statistical methods.

The article summarizes the load-forecasting methods based on artificial intelligence (AI) [23]. The paper gives an overview of data processing studies as to how the data are obtained. A comparison is made between one-step and rolling forecasting methodologies. Finally, several AI-based models are briefly discussed.

Artificial intelligence-based deep-learning techniques are discussed in [24]. This article has reviewed research articles from 2015 to 2020. The study focuses on deep-learning techniques, distributed deep-learning methods, Back-Propagation (BP), and non-BP-based methods. It is found in this survey that the computation time can be reduced by depending on data aggregation.

The article provides an overview of the recent STLF methods based on Machine-Learning (ML) algorithms, especially the hybrid models [25]. The advantages and limitations of single and hybrid predictive models are briefly discussed, and a comparison is also made using different performance indices.

A review is presented which discusses the emerging deep Artificial Neural Network (ANN)-based methods [26]. The individual methods, such as CNNs, RNNs, LSTMs, GRUs, DBNs, AEs, SAEs, and SDAEs, are briefly discussed.

For microgrid load forecasting, a survey is presented in [27] focusing on the latest analytical and approximation techniques. This article surveys the existing literature focused on energy demand forecasting, price and load forecasting, and renewable generation forecasting. A brief review is presented of different models along with their methodology and applications.

Another review of microgrid load forecasting is proposed in [28], which focuses on deep-learning-based methods. This article gives a direction for researchers on which datasets to use for particular applications. It is found that the efficiency of deep-learning-based methods depends on the amount of the dataset and therefore gives an indication for using larger data storage devices and high processing devices.

Low-voltage-level load-forecasting review is presented in [29], which discusses the current trends, main applications, challenges faced in this field, and recommendations. This article encourages further research work in a low-voltage field by establishing an open community-driven dataset.

The article explores the current state-of-the-art methods for STLF at the residential level [30]. Deep-learning-based techniques are focused mainly on this paper. It also discusses the inclusion of probabilistic methods in the deep-learning methods.

Among all these recently published reviews, there is no investigation on the key hyperparameters, and their interaction among themselves is not discussed. Moreover, there is no discussion of the effect of metaheuristic algorithms on load-forecasting methodology.

### 2.2. Research Challenges

In recent years, designing a lightweight model with fewer weights and parameters, which will eventually provide reasonable accuracy, has been a challenge due to the strict choices of hyperparameters [31,32]. There are two ways to go for hyperparameter tuning: (1) the number of hyperparameters increases when a complex structure is considered, and (2) to provide a satisfying accuracy with a carefully designed model needs fewer hyperparameters, which must be tuned to the stricter range. If the model structure is known, then it is possible to tune these parameters manually with experienced engineers or knowledge from previous works. However, it can be done on a smaller scale. If the model becomes complex or very emerging, then it becomes a great deal of manual work, even for experienced professionals to properly tune the hyperparameters. Again, there is a need for a guideline for less experience professionals on how to properly tune these parameters. Therefore, this study is motivated by the emerging trend of designing and training different models of load forecasting.

### 2.3. Research Contribution

The objective of this study is to conduct an extensive review of feasible metaheuristic algorithms for hyperparameter tuning. A comparison has been made in Table 1 to better understand the contribution of this work compared to existing literature.

**Table 1.** A comparison of existing reviews and this article. DDL—Data-decomposition layer, ML—Machine learning, MH—Metaheuristic, HPT—Hyperparameter tuning, HM—Hybrid model.

Ref.	Duration	DDL	ML	MH	HPT	HM	Contribution
[24]	2018–2022	X	√	X	X	√	It provides insight into methodologies and models based on artificial intelligence at different time series. Both deep-learning and machine-learning models are provided.
[25]	2015–2020	X	√	X	X	X	Distributed deep-learning techniques with back and non-back-propagation have been discussed. It proposes the Hilbert–Schmidt Independence Criterion without back-propagation to obtain higher accuracy in load forecasting.
[26]	2000–2020	X	√	√	X	√	It provides a comparative analysis of single and hybrid machine-learning algorithms, including their advantages, disadvantages, and functions.
[27]	2015–2021	X	√	X	X	X	It provides insight only on deep-learning-based models.
[28]	2003–2020	X	√	√	X	√	A review of analytical and approximation techniques has been discussed for load forecasting in a microgrid environment.
[29]	2006–2020	X	√	X	X	X	It discusses different deep-learning models for renewable energy and load forecasting and discusses future trends in smart energy management systems.
[30]	2004–2021	X	√	X	X	X	This article discusses different regression and machine-learning models for low-voltage distribution networks. It also provides insight into probabilistic forecasting.
[31]	2013–2023	√	√	X	X	√	Different load-forecasting techniques have been examined for residential load forecasting and energy management. It also provides a recommendation for combining probabilistic methods with machine-learning models.
[33]	2013–2023	X	√	√	X	√	Different algorithms of load forecasting, such as digital twin, data mining, federated learning, etc, are examined. It provides insight into choosing a viable load-forecasting model.
This work	2014–2024	√	√	√	√	√	This review gives a roadmap for choosing a metaheuristic algorithm along with machine-learning models for hyperparameter optimization to improve the forecasting results.

The main contributions of this paper are:

- Hyperparameters of different machine-learning algorithms are discussed to provide insight into their importance and involvement in optimization.
- A comprehensive assessment of state-of-the-art articles, which include existing methods, time resolution, and evaluation matrices, is presented.
- A complete framework is given to the researchers for short-term load forecasting using optimization algorithms.
- A comparative study is presented on different decomposition methods and single deep-learning-based forecasting methods.
- The challenges faced by the industry while load forecasting are discussed briefly.
- A generalized approach is proposed using a metaheuristic algorithm.
- A brief taxonomy is on previous research articles, including their advantages, limitations, and contributions, is presented.
- A guideline has been proposed based on the research findings.

### 3. Methodology

A critical and comprehensive review of state-of-the-art academic research articles on electric forecasting is undertaken in this study. This article follows a rigorous systematic protocol outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [34] to select appropriate papers. There is a difference between traditional review and systematic review in the way of literature searches and meta-analysis findings to reduce biases [35]. The main objective of this paper is to find research articles based on short-term load forecasting that use metaheuristic algorithms for optimization. Research articles from 2013 to 2023 are selected for this purpose. In this review, four steps are followed for systematic protocol:

1. Searching through keywords: Google Scholar is a powerful tool to search for research articles using keywords. Some common keywords appear, such as “electricity demand forecasting”, “electricity load forecasting”, and “electricity prediction”. While searching for “electricity load forecasting” in everything, including abstract, title, and the rest of the content, then 49,000 results appear in the first search. To limit the search space, the “with the exact phrase” option is used from advanced search in Google Scholar. Then, the search results come down to 1040. Again, to limit the search space further, “metaheuristic”, “optimization”, “short-term”, and specific algorithm names such as “GA” and “DE” keywords have been used with electricity load forecasting separately. Another 230 articles were identified across different databases.
2. Screening: The selected research papers found from key word searching are screened by giving emphasis on electric load forecasting or prediction using metaheuristic algorithms. The screening process was carried out through the title and abstract of 1085 articles. Around 880 articles were excluded as they did not meet the inclusion criteria.
3. Extra article identification: While going through the selected papers found in Step 2, some extra articles are found from these through citation. These extra articles are also screened by following Step 2.
4. Selection of appropriate articles: The articles found from Step 2 and Step 3 are carefully investigated for their objectives, methodologies, selected models, efficiencies etc. Results and future direction are also found in this step.

By following the above four steps, 165 research articles have been found that are based on short-term load forecasting using different metaheuristic algorithms (Figure 1). Figure 2a shows the number of publications from 2013 to 2023, and their publication sites are shown in Figure 2b. These publication sites are mostly part of Elsevier, IEEE, MDPI, Springer, Hindawi, etc. The percentage shares of these publishers are shown in Figure 2c.

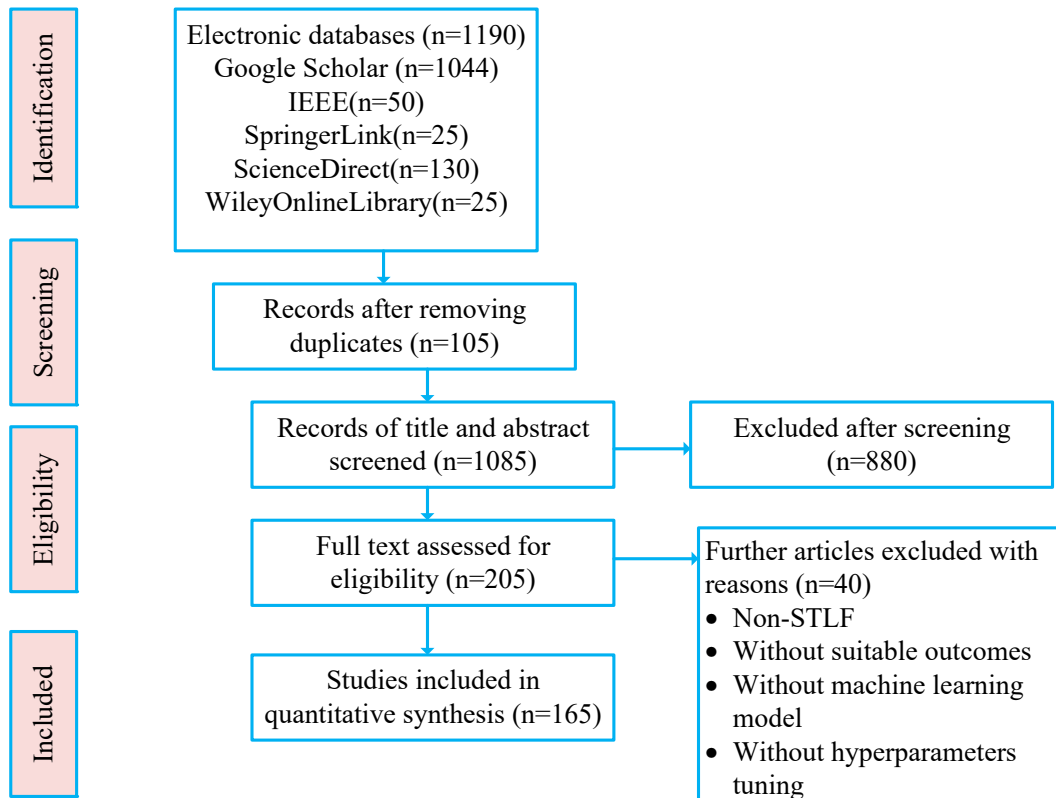


Figure 1. PRISMA flow diagram.

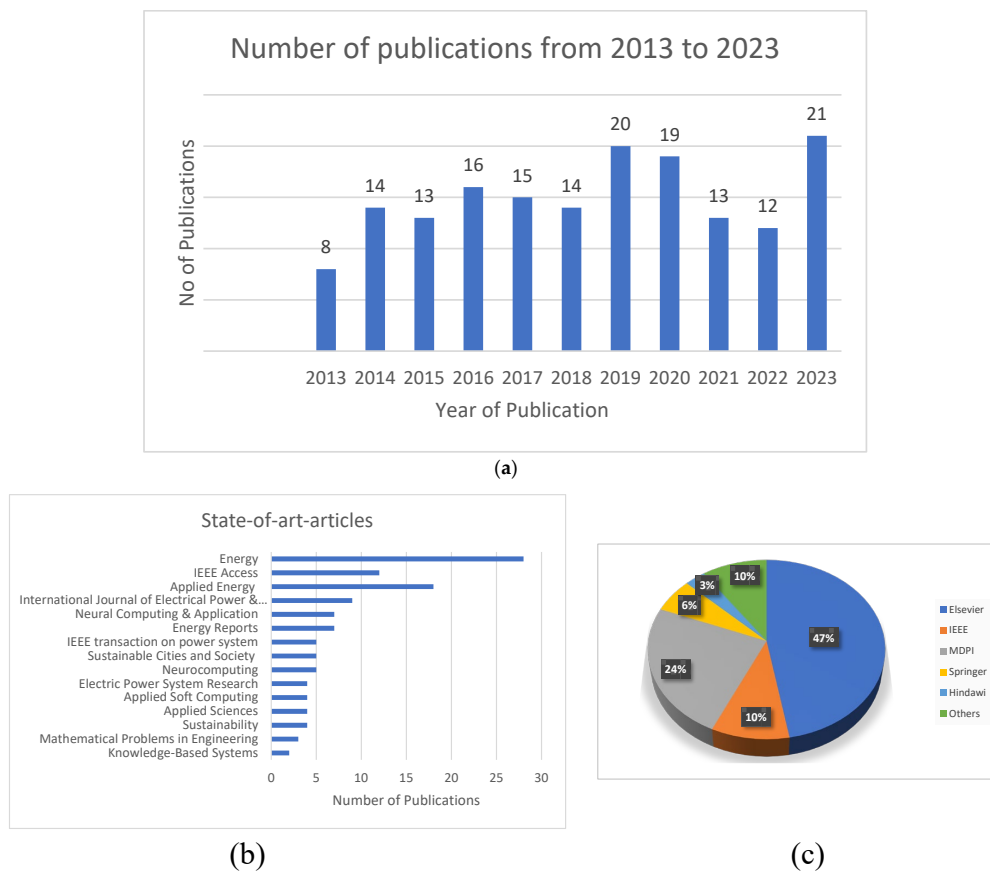
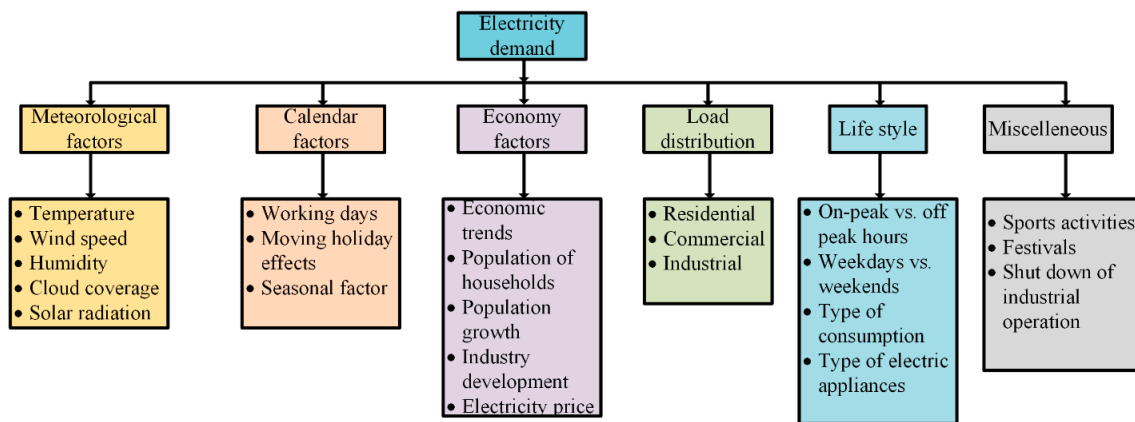


Figure 2. (a) Number of publications, (b) popular journals, and (c) publishers.



#### 4. Factors Affecting Load Forecasting

Electricity demand depends on various factors shown in Figure 3 are discussed below [36,37]:



**Figure 3.** Factors affecting the load data series.

##### 4.1. Meteorological Factors

Load forecasting is affected by weather conditions such as temperatures, humidity, wind speed, rain, snowfall etc. The weather plays an important role in the load profile. For example, during the summer and winter seasons, the usage of cooling and heating appliances goes up. Therefore, the peak demand is recorded at the coolest or warmest day relative to the demand during the days with average weather conditions. That is why weather forecast data are needed to predict the accurate load demand [38,39].

##### 4.2. Calendar Factors

In different years, there is a variation of the same month, which is known as the calendar effect. The holidays, such as special occasions that depend on moon sightings, vary each year. This is known as moving holiday effects. This effect has a great impact on residential and commercial load profiles [25,36].

##### 4.3. Economy Factors

Recession or flourishing affects the load profiles of residential and commercial consumers. Gross domestic product (GDP) and Gross national product (GNP) are used as indicators of a country's economic trend. The usage of electric appliances depends on the number of members residing in a household. Again, the population growth rate also impacts the power consumption rate. The economic development of a country depends on industrial development, which will, in turn, increase power consumption as well. The price increase or decrease in electricity price also has an impact on load forecasting [40].

##### 4.4. Load Distribution

Depending on the load type, such as residential, commercial, or industrial loads, the load profiles change. As a result, the load-forecasting method will also change.

##### 4.5. Lifestyle of Consumers

Consumer's lifestyle affects the peak and off-peak hours which depend on the time of the usage of appliances within the households. A variation in load profile is also seen during weekdays and weekends. Again, the type of electrical appliances varies from consumer to consumer. However, a similar pattern can be recognized, which has similar characteristics [41].

#### 4.6. Miscellaneous

If there is a special event such as a festival or sports event, then the usage of electricity increases. Therefore, the load profile changes within that area, which impacts the load forecasting [36].

### 5. Advantages of Metaheuristic Approaches in Load Forecasting

The load-forecasting methods can be categorized into two ways such as statistical methods and artificial intelligence-based methods [42]. Linear regression analysis method [43,44], Kalman filter [45], Box–Jenkin method [46], and Autoregressive Integrated Moving Average [47] are some examples of statistical models. Statistical models offer simpler structures and fast convergence rates. However, they suffer from lower prediction accuracy due to their linear behavior. Artificial intelligence-based models can better fit the nonlinear structure of the dataset and can work in changing environments. Although they show promising behavior, they have their limitations, such as overfitting and trapping into local minima. Therefore, the main challenges faced by forecasting methods are model generalization and hyperparameter optimization. The performance and accuracy of a prediction method can be improved by properly tuning hyperparameters and weighting coefficients [48]. Classical or metaheuristic algorithms can be used to tune these hyperparameters. Classical methods use analytical approaches to find near-optimum results. They can search through a specific subset of parameters or from predefined ranges. They are efficient in finding global solutions within the defined spaces. Again, if the fitness function contains multi-objectives, it becomes difficult for classical methods to solve the function due to its limitation on searching criteria [49]. These drawbacks can be overcome by metaheuristic algorithms, which are computationally intelligent and can solve complex multi-variable problems. One of the advantages of metaheuristic algorithms is that they can solve both linear and nonlinear problems. Again, they can be used for single or multi-objective problems [50]. Again, hyperparameters are the parameters that cannot be updated during the training process. However, the structure of the model is built on these parameters. As these parameters have a great influence on training accuracy and speed, they must be optimized carefully even before the training process [51]. An efficient optimization algorithm is needed to optimize these parameters to remove the human effort from the loop of the deep-learning process. The optimization requires large computational resources to optimize several hyperparameters together. Therefore, metaheuristic algorithms are preferred over classical methods to properly tune the hyperparameters, which will eventually improve the forecasting results.

### 6. Evaluation Criteria

To validate the accuracy of a forecasting model, several performance evaluation indices are used.

Mean square error (*MSE*): *MSE* is the square of the difference between actual and forecasted values [52,53]. The smaller the error is, the more accurate the model is. The equation for *MSE* is:

$$M_{MSE} = \frac{1}{n} \sum_{t=1}^n (x(t) - \hat{x}(t))^2 \quad (1)$$

where  $n$  is the number of iterations at time  $t$ ,  $x(t)$  is the actual value at time  $t$  and  $\hat{x}(t)$  is the predicted value.

Root mean square error (*RMSE*): *RMSE* is defined as the standard deviation of the prediction errors which is shown in the equation [54,55]. It measures the concentration of the data points around the best-fitted line.

$$M_{RMSE} = \sqrt{\frac{\sum_{t=1}^n (x(t) - \hat{x}(t))^2}{n}} \quad (2)$$



Mean absolute error (*MAE*): MAE is the average of absolute difference between actual values and estimated values as shown below equation [56]:

$$M_{MAE} = \frac{1}{n} \sum_{t=1}^n x(t) - \hat{x}(t) \quad (3)$$

Mean absolute percentage error (*MAPE*): MAPE is the percentage of the average absolute difference between actual values and estimated values, which is divided by actual value [57,58]. The equation for MAPE is:

$$M_{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x(t) - \hat{x}(t)}{x(t)} \right| \times 100\% \quad (4)$$

Symmetric mean absolute percentage error (*SMAPE*): SMAPE is calculated by the difference between actual values and estimated values divided by the sum of these values [59]. This evaluation index avoids any biases due to its symmetrical properties, which makes it non-dependent on the time horizon of the data series.

$$M_{SMAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x(t) - \hat{x}(t)}{\frac{x(t) + \hat{x}(t)}{2}} \right| \times 100\% \quad (5)$$

Normalized root mean square error (*NRMSE*): NRMSE can be calculated either by mean values or using the difference between maximum and minimum values [60].

$$M_{NRMSE} = \frac{\sqrt{\frac{\sum_{t=1}^n (x(t) - \hat{x}(t))^2}{n}}}{x_{max} - x_{min}} \quad (6)$$

where  $n$  is the number of iterations at time  $t$ ,  $x(t)$  is the actual value at time  $t$  and  $\hat{x}(t)$  is the predicted value, and  $x_{max}$  is the maximum value and  $x_{min}$  is the minimum value.

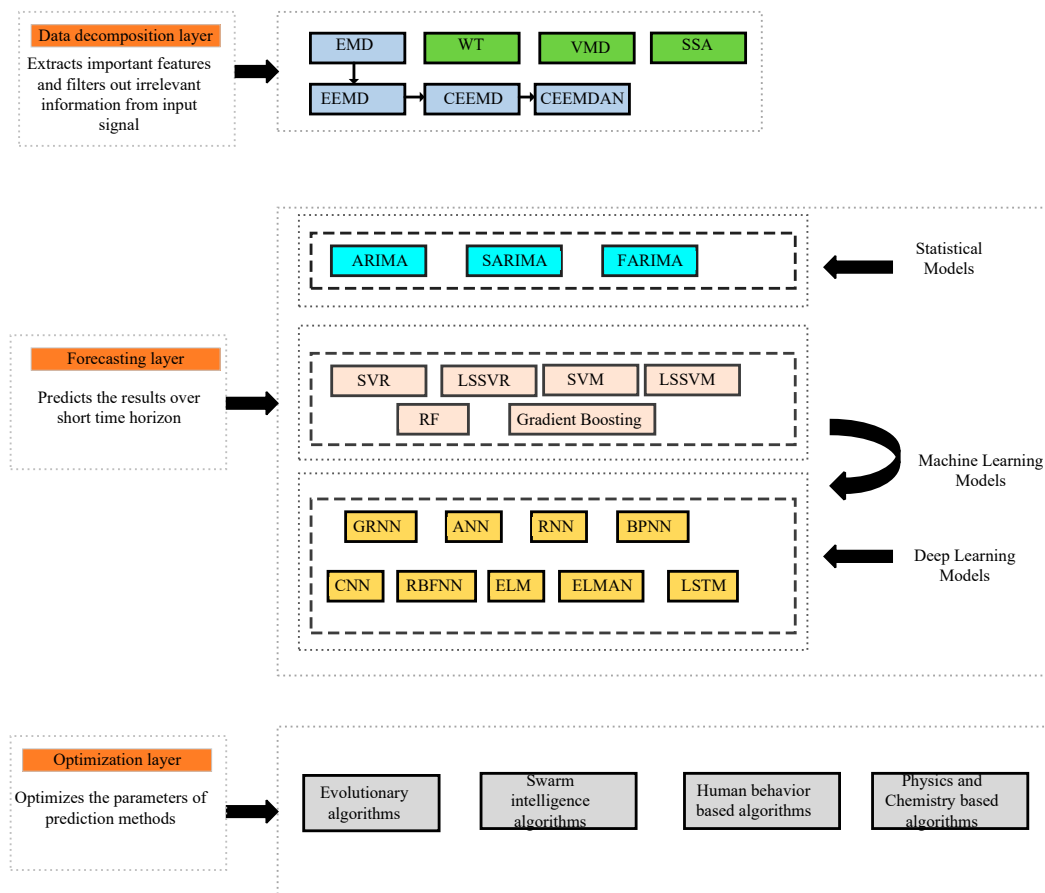
Coefficient of determination ( $R^2$ ):  $R^2$  is defined as the variation in the dependent variables that can be derived from independent variables [61].

$$M_{R^2} = 1 - \sqrt{\frac{\sum_{t=1}^n (x(t) - \hat{x}(t))^2}{\sum_{t=1}^n (x(t))^2}} \quad (7)$$

As *MSE* is the square of the difference, it is sensitive to large errors. It can be used during model training and optimization. *RMSE* is also sensitive to large errors like *MSE*. When the dataset is significantly large, then *RMSE* can be used to compare different models. *MAE* is less sensitive to large errors than *MSE* and *RMSE*. As a result, *MAE* treats all errors equally, irrespective of data size. *MAPE* is suitable for making a comparison among different datasets. However, if there is a possibility that the actual value is going to be zero or near zero, then it should be avoided as it will lead to undefined or extremely high values. *NRMSE* can be useful when there is variability in data ranges and there is a need for a standardized measure of error for multiple forecasting scenarios.  $R^2$  can be used to explain variance rather than minimize prediction errors.

## 7. A Complete Framework of Load Forecasting Using Metaheuristic Algorithms

This article proposes a short-term load-forecasting methodology using metaheuristic algorithms based on state-of-the-art articles. Any load forecasting model consists of three layers: the data-decomposition layer, forecasting layer, and optimization layer shown in Figure 4. The data-decomposition layer is used to decompose the original information from nonlinear and nonstationary datasets. The forecasting layer is used to predict the result over a selected time horizon. To improve the prediction accuracy, the optimization layer is used to optimize the parameters of the forecasting layer.



**Figure 4.** A complete framework for load forecasting.

7.1. Data-Decomposition Layer

The complexity of the forecasting layer is reduced using the data-decomposition layer. It helps input variables to follow the decomposed sequence [62]. Forecasting accuracy is also improved as the unnecessary information from the input signal is eliminated through the feature extraction process. Some of the common data-decomposition technologies that are being used in this layer are Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), Wavelet Transform (WT), and Singular Spectrum Analysis (SSA) [63].

7.1.1. Wavelet Transform (WT)

Wavelet Transform can filter out irrelevant information from the input signal. A continuous input time series is divided into different scale components by a mathematical function called wavelet [64]. It is a band pass filter that filters out the lowest level of components, ensuring the coverage of the whole spectrum. A mother wavelet, which is an oscillatory function, is translated and copied into wavelets. There are two types of Wavelet Transform: Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). If the input signal is  $x(t)$ , then CWT can be defined as follows:

$$CWT_{\psi}(a, b) = \left(1/\sqrt{|a|}\right) \int_{-\infty}^{\infty} x(t)\psi * ((t - b)/a)dt \tag{8}$$

where  $a, b$  are the scale and translation parameters, and  $\psi$  is the wavelet function.

DWT can be defined as follows:

$$DWT_x(m, n) = \left(1/\sqrt{2^m}\right) \sum_k x_k \psi * ((k - n)/2^m) \tag{9}$$

where  $m$  is scale factor, and  $n$  is the number of samples.

The identity of the input signal is carried out by the low-frequency components, which are the important part. Again, high-frequency components represent signal details that are filtered out by Wavelet Transform containing different mother wavelets and their corresponding wavelets.

### 7.1.2. Empirical Mode Decomposition (EMD) and Its Different Versions

Empirical Model Decomposition (EMD), which was first proposed in [65], can select the appropriate features based on the input signal characteristics. One of the drawbacks of Wavelet Transform is that it requires preselection of wavelet functions. This drawback can be overcome by EMD as it does not depend on the preselection of functions, which eventually reduces human intervention. EMD is considered to be a sieving process that decomposes the input signal into a series of oscillation functions, which are known as intrinsic mode functions (IMFs). This trend term was used in [66], which decomposes the input signal into five IMFs. The equation for EMD is given as follows [67]:

$$f(t) = \sum_{k=1}^N f_k(t) + r(t) \tag{10}$$

where input signal  $f(t)$  is decomposed into  $N + 1$  number of IMFs  $f_k(t)$  and  $r(t)$  is the residual noise.

There are some obvious limitations of EMD, such as mode aliasing, false mode detection, and end effect. Therefore, researchers have focused on improving EMD for different load-forecasting scenarios.

#### Ensemble Empirical Mode Decomposition (EEMD)

EEMD uses white Gaussian noise to the input signal and was first discussed in [68]. White Gaussian noise has uniform frequency characteristics, which can change the distribution of extreme signal points. EEMD can effectively eliminate the mode-aliasing effect of EMD. The equation for EEMD is the same as Equation (10). The only difference is that white noise is added each time [69].

$$f(t) = \sum_{k=1}^N f_{kj}(t) + f_j(t) \tag{11}$$

where  $j$  is the number of added noise.

However, computational time is increased due to the multiple decomposition processes and residual noise in reconstruction is seen due to adding white noise.

#### Complete Ensemble Empirical Mode Decomposition (CEEMD)

A complete ensemble empirical mode decomposition (CEEMD) is suggested to extract the residual noise from the data mixture by adding positive and negative white noise to the original data, which is of the same amplitude [70]. Though this method has used the same magnitude of noise similar to EEMD, the residual noise is effectively removed from IMFs through CEEMD.

The IMF generated by CEEMD is denoted by [71]:

$$f_k(t) = \frac{1}{2N} \sum_{k=1}^N f_{kj}^+(t) + f_{kj}^-(t) \tag{12}$$

where  $f_{kj}^+(t)$  is the  $k$ th IMF generated in  $j$ th positive trial and  $f_{kj}^-(t)$  is the  $k$ th IMF generated in  $j$ th negative trial.

Though the denoising effect is achieved by CEEMD, the computation burden is still high.

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was first proposed in [72], where adaptive white noise of opposite signs is added to each stage of the decomposition process. After the decomposition process, the output signal matches with the input sequence which represents no reconstruction error. The denoising effect is achieved as well as the computational cost problem [73].

A comparative analysis of EMD and its derivatives is shown in Table 2.

**Table 2.** Shows a comparative analysis of EMD and its derivatives.

Methods	Process	Advantages	Disadvantages
EMD	Decomposes the original signal into a series of IMFs	<ul style="list-style-type: none"> <li>Does not depend on the preselection of functions.</li> <li>Reduces human intervention.</li> </ul>	<ul style="list-style-type: none"> <li>Mode aliasing</li> <li>False mode detection</li> <li>End effect</li> </ul>
EEMD	Adds white Gaussian noise to the original signal	<ul style="list-style-type: none"> <li>Modal aliasing effect is eliminated.</li> </ul>	<ul style="list-style-type: none"> <li>Noising effect</li> <li>Computational burden</li> </ul>
CEEMD	Adds positive and negative white noise	<ul style="list-style-type: none"> <li>Denoising effect is achieved.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden</li> </ul>
CEEMDAN	Adds adaptive white noise	<ul style="list-style-type: none"> <li>Computation efficiency is increased.</li> </ul>	<ul style="list-style-type: none"> <li>Still have some residue.</li> </ul>

7.1.3. Variational Mode Decomposition (VMD)

The Variational Mode Decomposition method is based on the frequency domain, which was first proposed in [74]. This decomposition method is used to construct and solve variational problems. It is an adaptive, non-recursive, quasi-orthogonal decomposition method that consists of Weiner filtering, Hilbert transformation, and multiplier [62]. Weiner filtering is used to achieve the denoising effect, the marginal spectrum problem is solved by Hilbert transformation, and multipliers are used to solve an unconstrained problem. The input signal  $f(t)$  is decomposed into a  $k$  number of sub-modes  $u_k$  which contains specific sparsity components. The center frequency is  $\omega_k$ . The goal of the VMD process is to find the optimal solution for  $k$  modes, which makes the smallest bandwidth [67].

$$\min_{u_k, \omega_k} \left\{ \sum_{k=1}^K \left\| \sigma_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \text{ s.t. } \sum_{k=1}^K u_k = f(t) \tag{13}$$

where  $\delta(t)$  is the Dirac distribution function,  $\left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t)$  is the Hilbert transformation,  $e^{-j\omega_k t}$  is the exponential term to adjust the frequency spectrum.

One of the drawbacks of VMD is the mode-aliasing effect. Therefore, parameters such as the number of modes  $k$  must be selected carefully.

7.1.4. Singular Spectrum Analysis (SSA)

Singular spectrum analysis (SSA), a non-parametric method, consists of classical time series analysis, multivariate statistics geometry, and signal processing [75,76]. SSA consists of decomposition and reconstruction stages. In the decomposition stage, input data are decomposed, and the original signal is reconstructed to predict the new data points at a later stage [77]. This method is useful for short time series, extracting seasonality factors, smoothing, and extracting important information of different amplitudes. Some of the limitations of SSA are that it cannot handle large datasets, cannot reconstruct the original signal when the spectrum spreads, and computational complexity exists.

## 7.2. Forecasting Layer

The forecasting layer is used to predict the results by taking the data from the data-decomposition layer. This layer consists of two models, namely the statistical model and the machine-learning model.

### 7.2.1. Statistical Model

The most widely used statistical model, also referred to as the time series model, is the Autoregressive Integrated Moving Average (ARIMA) model. Two statisticians, George Box and Gwilym Jenkins, first introduced the ARIMA model in 1970 [46]. The ARIMA model is composed of an Autoregressive model (AR), a Moving Average (MA), and an integration of both AR and MA. This model is effective for short time series. However, this model cannot handle the nonlinear and nonstationary behavior of data series. Therefore, Box and Jenkins include seasonality to ARIMA, which is known as SARIMA, by introducing a seasonal exponential smoothing factor. This model is adaptive and can handle seasonal and nonlinear data series. However, the limitations of this method are (a) the computation burden is high, (b) it requires past values of data, and (c) it needs a good understanding of statistics.

### 7.2.2. Machine-Learning Model

Statistical models are limited in number and can lead to unsatisfactory prediction results due to higher computational burden and nonlinear and fluctuated behavior of power system data series [22]. Machine-learning models provide a promising alternative in this context. When the learning algorithm is selected, these models can work by themselves without being programmed [78]. The widely used machine-learning methods for forecasting are Support Vector Machine (SVM), Least Square Support Vector Machine (LSSVM), Random Forest (RF), and Gradient Boosting. The deep-learning model is another branch of machine-learning. Some of the popular deep-learning methods include Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recursive Neural Network (RNN), Long Short-Term Memory (LSTM), Generalized Regression Neural Network (GRNN), Back-Propagation Neural Network (BPNN), Radial Basis Function Neural Network (RBFNN), ELM (Extreme Learning Machine) and ELMAN neural network. A comparative analysis of these deep-learning methods is discussed in Table 3.

**Table 3.** A comparative analysis of deep-learning methods.

Algorithms	Advantages	Disadvantages
ANN	<ul style="list-style-type: none"> <li>Performance is not affected if there is any loss of information</li> <li>Fault tolerance</li> <li>Parallel processing capability</li> </ul>	<ul style="list-style-type: none"> <li>Depends on hardware.</li> <li>Unknown network duration.</li> <li>Interpretation of the network is difficult.</li> </ul>
CNN	<ul style="list-style-type: none"> <li>Can handle large datasets</li> <li>Extracts features automatically</li> <li>Unsupervised learning</li> </ul>	<ul style="list-style-type: none"> <li>High computational burden.</li> <li>Needs a high-speed processing unit.</li> </ul>
RNN	<ul style="list-style-type: none"> <li>Remembers each information</li> <li>Model size remains the same if the input size becomes bigger</li> </ul>	<ul style="list-style-type: none"> <li>Exploding and vanishing gradient problem.</li> <li>Difficult to train an RNN.</li> <li>Cannot process long data sequences.</li> </ul>

Table 3. Cont.

Algorithms	Advantages	Disadvantages
LSTM	<ul style="list-style-type: none"> <li>• Long-term information can be stored</li> <li>• Data can be trained in backward and forward directions</li> <li>• Exploding and vanishing gradient problem is solved</li> </ul>	<ul style="list-style-type: none"> <li>• Computational burden is high</li> <li>• Higher data storage is required during the training stage</li> <li>• Not suitable for parallel processing.</li> </ul>
GRNN	<ul style="list-style-type: none"> <li>• Training process is faster</li> <li>• Has only one free parameter</li> <li>• Guaranteed global optima</li> </ul>	<ul style="list-style-type: none"> <li>• Computational burden is high</li> <li>• Sensitive to noise.</li> </ul>
BPNN	<ul style="list-style-type: none"> <li>• Versatile algorithm</li> <li>• Computationally efficient</li> </ul>	<ul style="list-style-type: none"> <li>• Can stuck into local minima</li> <li>• Exploding and vanishing gradient problem</li> <li>• Requires parameter tuning.</li> </ul>
RBFNN	<ul style="list-style-type: none"> <li>• Can process high-dimensional data</li> <li>• Can train and test data quickly</li> <li>• Tolerate to noise</li> </ul>	<ul style="list-style-type: none"> <li>• Computational complexity increases</li> <li>• A nonlinear system cannot be modeled strongly.</li> </ul>
ELM	<ul style="list-style-type: none"> <li>• Generalization capability increases</li> <li>• Human intervention is minimal</li> <li>• Learning process is speedy</li> </ul>	<ul style="list-style-type: none"> <li>• Interpretation is difficult</li> <li>• Lacks control over the hidden layer.</li> <li>• Sensitive to noise.</li> </ul>
ELMAN	<ul style="list-style-type: none"> <li>• Can deal with nonlinear and complex data</li> <li>• Keeps the information of the previous data</li> </ul>	<ul style="list-style-type: none"> <li>• Cannot guarantee an optimal solution</li> <li>• Convergence rate is slow</li> <li>• Inflexibility</li> </ul>

### Hyperparameters of Machine-Learning Models

Some of the key hyperparameters of neural network architectures shown in Table 4 are discussed below:

**Learning rate ( $\eta$ ):** The learning rate helps to determine the strength of the model where the information from the previous iteration is replaced by new information [79]. If the learning rate is 0, then the model will not learn anything, and if it is 1, then the model progresses according to the new information.

**Number of hidden layers ( $d$ ):** The number of hidden layers influences the final output directly by determining the overall structure of the model. With the increase in the number of layers, the complexity of the model increases [80].

**Number of neurons ( $\omega$ ):** The number of neurons in a neural network refers to the count of individual processing units (or nodes) in a given layer of the network [81]. Each neuron receives input, processes it through an activation function, and produces an output that can be sent to the next layer.

**Activation function:** Activation function introduces nonlinear properties to the output of neurons. A neural network model behaves as a simple linear regression model without the activation function. The most commonly used activation functions are sigmoid, hyperbolic tangent rectified linear units (ReLU) [82], and Swish [83].



**Epochs:** During the training process, epochs are the number of complete passes through the entire training dataset [81].

**Batch size:** During the training process, the dataset is often too large to pass through all at once. That is why they are divided into smaller groups called batches. The batch size refers to the number of samples included in each of these groups [51].

**Kernel size:** A kernel is a small matrix used to perform convolution on input data. It slides over the input data to extract features by performing mathematical operations. Kernel Size refers to the dimensions of the kernel used in convolutional layers of a convolutional neural network (CNN). It specifies how many rows and columns the kernel contains [84].

**Number of filters:** The number of filters refers to the quantity of convolutional kernels in a particular convolutional layer [85].

**Stride:** Stride refers to the number of pixels by which a convolutional kernel moves across the input data during the convolution operation. The stride affects the dimensions of the output feature map [86].

**Pooling size:** Pooling is a down-sampling operation that reduces the spatial dimensions of the feature maps while retaining important information. Pooling size refers to the dimensions of the pooling window used in pooling layers of a convolutional neural network (CNN) [87].

**Dropout rate:** Dropout rate refers to the proportion of neurons that are randomly set to zero during training in a neural network, effectively dropping out these neurons. The dropout technique is designed to improve the generalization of a neural network by preventing it from becoming overly reliant on any single neuron or a small group of neurons [88].

**Spread parameter:** The spread parameter controls the shape of the Gaussian activation function by measuring the similarity between input patterns. It determines how quickly the influence of a given training sample diminishes as the distance from that sample increases [89].

**Momentum:** Momentum is designed to accelerate the convergence of the training process and help navigate through the loss landscape. It helps to carry forward the previous updates to the weights, smoothing out the optimization trajectory [90].

**Regularization parameter:** The regularization parameter controls the complexity of the model and prevents overfitting. Regularization techniques add a penalty to the loss function based on the magnitude of the model parameters [91].

**Table 4.** Hyperparameters of deep-learning methods.

Algorithms	Hyperparameters
ANN	<ul style="list-style-type: none"> <li>• Number of hidden layers (d)</li> <li>• Number of neurons (<math>\omega</math>)</li> <li>• Activation function</li> <li>• Learning rate</li> <li>• Epochs</li> <li>• Batch size</li> </ul>
CNN	<ul style="list-style-type: none"> <li>• Number of filters</li> <li>• Kernel size</li> <li>• Stride</li> <li>• Pooling size</li> <li>• Learning rate</li> <li>• Dropout rate</li> </ul>

Table 4. Cont.

Algorithms	Hyperparameters
RNN	<ul style="list-style-type: none"> <li>• Number of hidden layers (d)</li> <li>• Learning rate</li> <li>• Epochs</li> <li>• Batch size</li> <li>• Dropout rate</li> </ul>
LSTM	<ul style="list-style-type: none"> <li>• Number of LSTM units</li> <li>• Learning rate</li> <li>• Epochs</li> <li>• Batch size</li> <li>• Dropout rate</li> <li>• Sequence length</li> </ul>
GRNN	<ul style="list-style-type: none"> <li>• Spread parameter (<math>\sigma</math>)</li> <li>• Number of neurons (<math>\omega</math>)</li> </ul>
BPNN	<ul style="list-style-type: none"> <li>• Number of hidden layers (d)</li> <li>• Number of neurons (<math>\omega</math>)</li> <li>• Activation function</li> <li>• Learning rate</li> <li>• Epochs</li> <li>• Momentum</li> </ul>
RBFNN	<ul style="list-style-type: none"> <li>• Number of neurons (<math>\omega</math>)</li> <li>• Spread parameter(<math>\sigma</math>)</li> <li>• Learning rate</li> <li>• Epochs</li> </ul>
ELM	<ul style="list-style-type: none"> <li>• Number of neurons (<math>\omega</math>)</li> <li>• Activation function</li> <li>• Regularization parameter</li> </ul>
ELMAN	<ul style="list-style-type: none"> <li>• Number of hidden layers (d)</li> <li>• Activation function</li> <li>• Learning rate</li> <li>• Epochs</li> <li>• Momentum</li> <li>• Regularization parameter</li> </ul>

### Interdependencies Among Hyperparameters

The interdependencies among hyperparameters make it critical to use systematic tuning methods to find optimal values in combination. A breakdown of how some key parameters can interact across different models is given below:

#### 1. Learning rate and epochs

Learning rate is proportional to the convergence rate which means a higher learning rate can lead to faster convergence. However, having a high learning rate may overshoot the optimal solution. If the learning rate is too high, it might require fewer epochs as the model fails to converge. Conversely, a lower learning rate may necessitate more epochs

to reach convergence. There is a need to balance between learning rate with the number of epochs.

## 2. Batch size and learning rate

Smaller batch sizes can lead to more noisy gradient estimates, which can help escape local minima but may require a lower learning rate to stabilize training. Larger batch sizes provide a more accurate estimate of the gradient but might benefit from a higher learning rate. Tuning both together is essential for achieving optimal convergence speed and stability.

## 3. Number of neurons and layers

If the number of neurons or layers increases, it also increases the network's capacity to learn complex patterns. However, there is a risk of overfitting if regularization techniques or training data are not sufficient. Therefore, the architecture should match the complexity of the data; otherwise, the model may either underfit (few neurons/layers) or overfit (many neurons/layers).

## 4. Dropout rate and network complexity

The dropout rate helps mitigate overfitting in complex networks by randomly disabling a fraction of neurons during training. A high dropout rate in a simple network may hinder learning. The effectiveness of dropout is highly dependent on the network's architecture and the amount of training data available.

## 5. Spread parameter and number of neurons

A larger spread makes the network more general but can reduce its capacity to capture complex patterns, especially if there are fewer neurons. The choice of the spread parameter must align with the number of neurons to ensure the model captures relevant patterns without losing specificity.

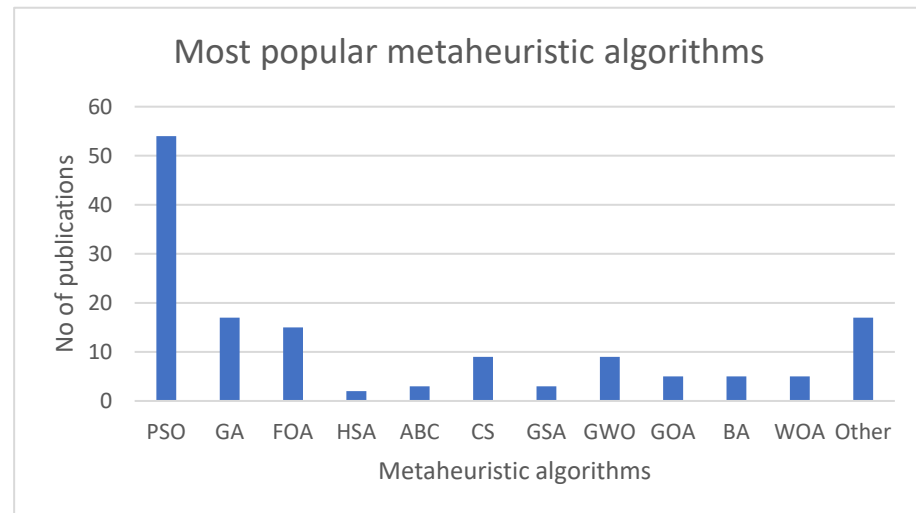
## 6. Activation function and learning rate

Different activation functions can respond differently to learning rates. For example, ReLU can cause dying neurons if the learning rate is too high, while sigmoid or hyperbolic tangent functions might cause vanishing gradients with a lower learning rate. Choosing the right activation function relative to the learning rate is important for maintaining gradient flow and ensuring effective learning.

### 7.3. Optimization Layer

The main objective of the optimization layer is to optimize the parameters of forecasting methods by different metaheuristic algorithms to improve the accuracy and performance of the prediction model.

The article surveys 165 articles, which gives an idea of which methods are extensively used. Among all the metaheuristic algorithms from the existing literature, the most popular ones for load forecasting are shown in Figure 5. These algorithms are Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Fruit-fly Optimization Algorithm (FOA), Harmony Search Algorithm (HAS), Artificial Bee Colony (ABC), Cuckoo Search (CS), Gravitational Search Algorithm (GSA), Gray Wolf Optimization (GWO), Grasshopper Optimization Algorithm (GOA), Bat Algorithm (BA) and Whale Optimization Algorithm (WOA). The chart reveals that the PSO has been used extensively in the literature. The next most used algorithm is GA. Artificial Bee Colony also gains popularity. Another two algorithms used are GWO and BA. The next section provides a genderized discussion of these algorithms.



**Figure 5.** A number of metaheuristic algorithms used in the literature.

**Genetic Algorithm (GA):** Genetic Algorithm uses biological concepts to find the approximate solution. This is based on Darwin’s survival of the fittest theory [92,93]. It finds the solution through biologically inspired processes such as selection, crossover, and mutation. A workflow of this algorithm is shown in Figure 6a. Initial population individuals can be altered and mutative. These can be crossed over in pairs to create a better one. These individuals can be mutated to replace the previous one for better fitness function in the next iteration.

**Grasshopper Optimization Algorithm (GOA):** Grasshopper algorithm mimics the foraging and swarming behavior of grasshoppers [94,95]. Their lifecycle consists of three stages—egg, nymph, and adult—going through a process called metamorphosis. A young agent takes a smaller step slowly, whereas an adult agent takes the bigger steps, and at this stage, they destroy the crops. This behavior is modeled mathematically to form this optimization technique. Figure 6b shows the workflow of this algorithm.

**Whale Optimization Algorithm (WOA):** The algorithm mimics the hunting behavior of whales. Their hunting behavior is known as the bubble-net feeding method, where whales hunt their prey close to the surface. During this process, they create distinctive bubbles along a circular path. There are three steps in this algorithm such as: sieging the prey, attacking the prey, and searching for prey. A detailed procedure of WOA can be found here [96]. A generalized workflow is shown in Figure 6c.

**Artificial Bee Colony (ABC):** This algorithm uses the social searching behavior of finding food [97]. Three types of bees performing in the colony, such as working bees, watchdogs, and scouts, are modeled. Working bees execute the exploitation procedure by nourishing food sources, which are converted to candidate solutions. Watchdogs identify the most promising food sources based on the feedback given by working bees. If there is no change in the food source, scout bees begin the exploration procedure. Figure 6d shows a generalized workflow of the ABC algorithm.

**Bat Algorithm (BA):** Bats use certain sound systems to locate prey, barriers, and nests while moving in the darkness. This phenomenon is the inspiration of this algorithm, where the best current position is determined by changing the speed and position of each bat [98]. Bats fly at random velocity toward a particular position. During the hunting stage, bats vary their frequency, loudness, and pulse emission rate. Search is intensified by random walk. When it reaches the stopping criteria, then it is considered that the best solutions are found. A generalized workflow is shown in Figure 6e.

**Cuckoo Search Algorithm (CS):** The Cuckoo Search Algorithm imitates the parasitic behavior of cuckoos to lay eggs in host birds’ nests [99]. The goal is to increase reproductivity while preventing the host birds not to finding the eggs. First, each cuckoo lays one egg and dumps it in a nest randomly. In the next generation, only high-quality eggs in the best

nests are selected. However, there is also a probability of the host bird finding the eggs and throwing away or building a new nest. Figure 6f shows a generalized workflow of the CS algorithm.

Particle Swarm Optimization (PSO): Particle Swarm imitates the flying behavior of a bird's flock [100]. First, the objective function is calculated at a specific point. The flying path is selected based on the current position and the previous best position. This step is repeated several times until the stopping criteria are met. A generalized workflow is shown in Figure 6g.

Gravitational Search Algorithm (GSA): The laws of gravity and motion are the basis of this algorithm [101]. Each particle is considered to be an object, and its masses determine its performance. All objects attract each other due to the gravitational force. Lighter objects tend to move to the heavier objects due to this phenomenon. During the exploitation of the algorithm, heavier objects move more slowly than lighter objects. The global optima is found at the position where mass is heaviest. Figure 6h shows a generalized workflow of the GSA algorithm.

Gray Wolf Optimization (GWO): This algorithm imitates the predatory and hierarchical process of gray wolves [102]. Alpha, beta, delta, and omega are four types of gray wolves whose leadership hierarchies are simulated. The major steps that are considered in this algorithm are searching for prey, sieging the prey, and attacking the prey. Alpha is considered to be the fittest solution, followed by beta and gamma, and omega is considered to be the worst solution. Each type of wolf updates its position in the new generation based on the best solutions found from the first three groups in the previous generation. A generalized workflow is shown in Figure 6i.

Harmony Search Algorithm (HSA): This algorithm is inspired by the natural musical performance processes where a better state of harmony is searched by a musician [103]. Harmony search has selection and mutation stages whereas crossover is not explicitly used like GA. There are three stages in HAS: harmony memory usage, pitch adjusting, and randomization. Figure 6j shows a generalized workflow of the HSA algorithm.

Fruit-fly Optimization Algorithm (FOA): Fruit-fly Optimization Algorithm is based on the food-hunting behavior of a fruit-fly [104]. First, an individual fruit-fly searches for food in a random direction. They use smell sensation to search for food. Then, they use the vision to fly in that direction. A generalized workflow is shown in Figure 6k.

Table 5 shows a quantitative analysis of the most commonly used metaheuristic algorithms, including their advantages, disadvantages, process and developers, solving capability, and computation time.

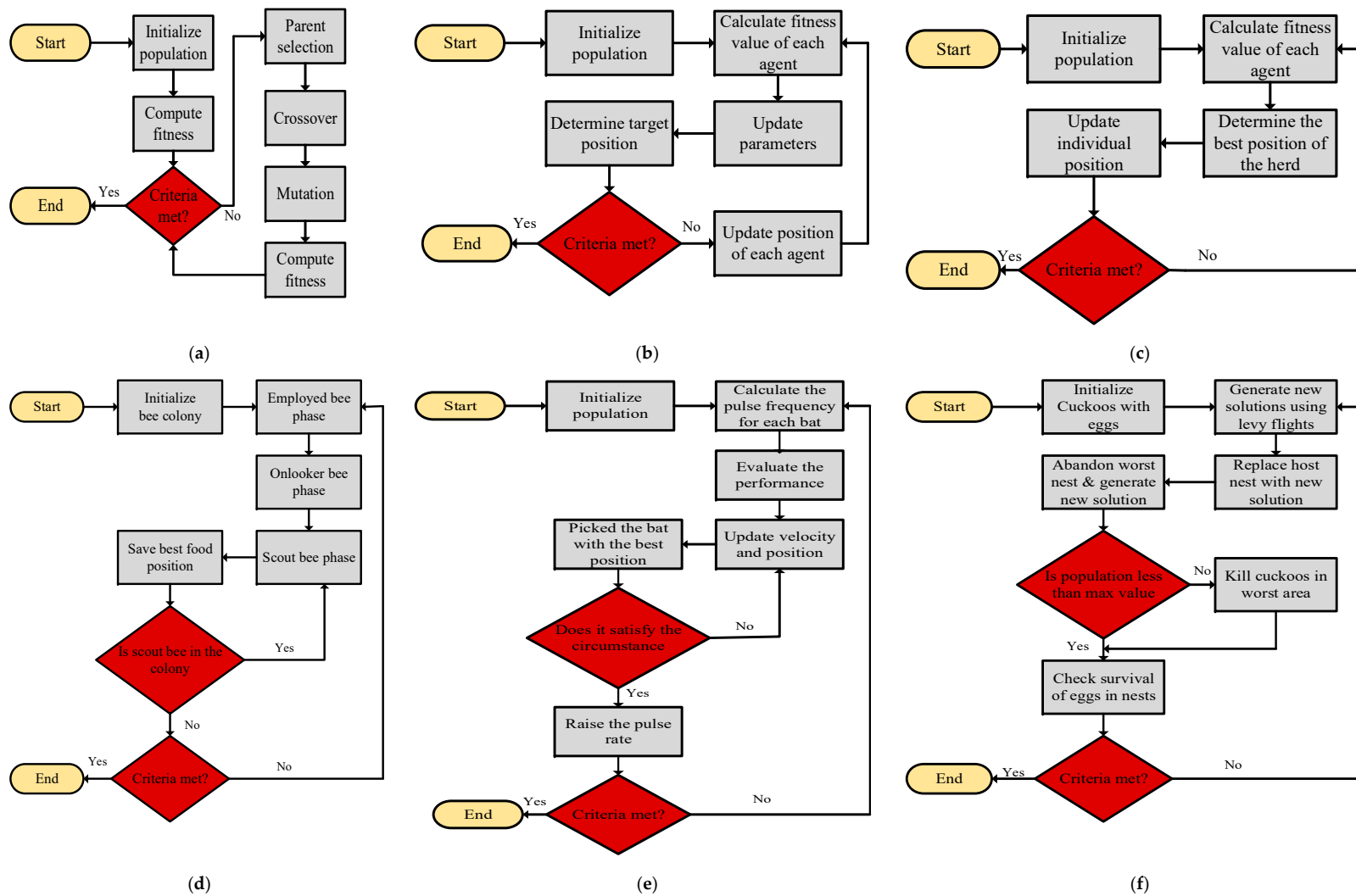


Figure 6. Cont.



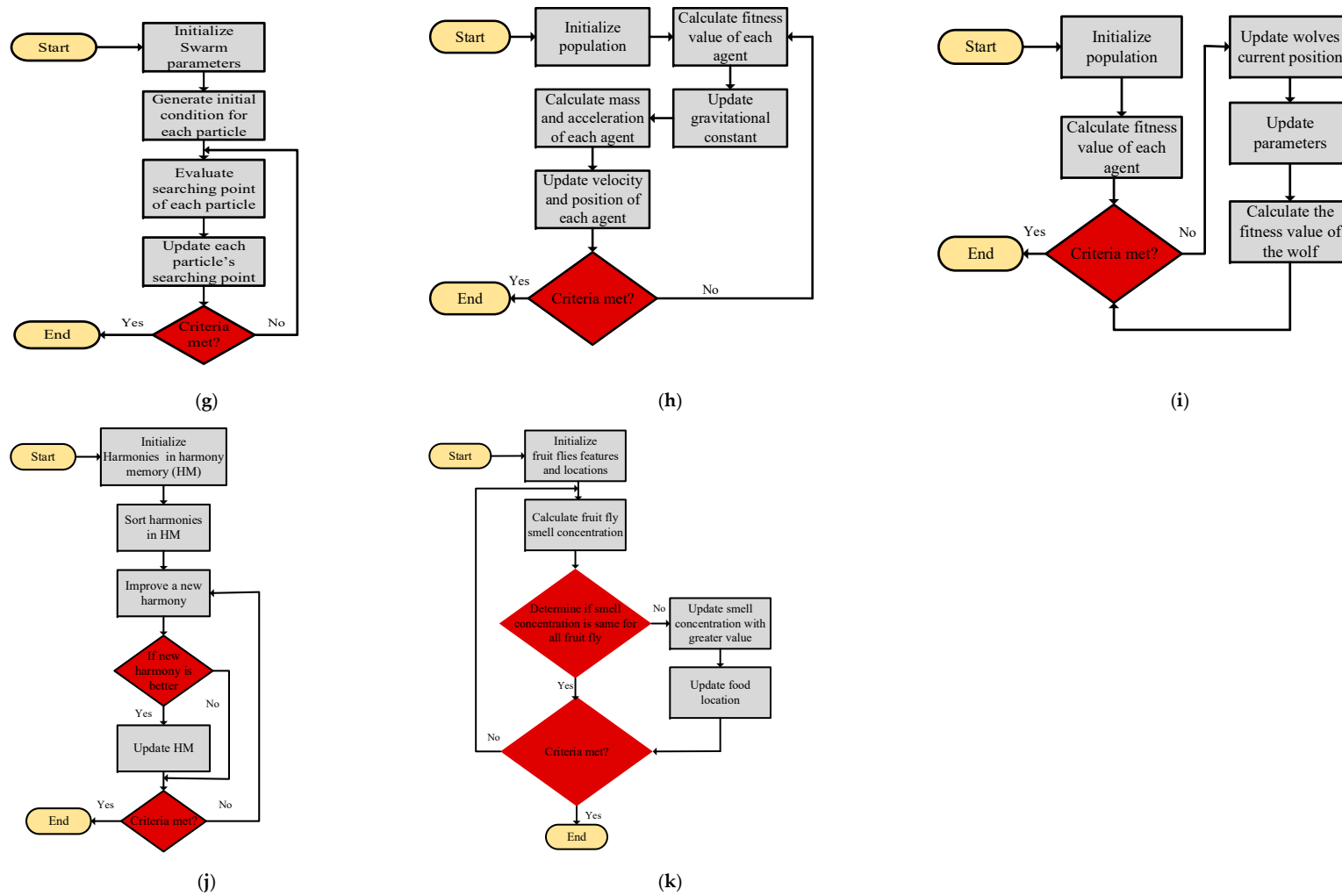


Figure 6. Workflows for (a) GA (b) GOA (c) WOA (d) ABC (e) BA (f) CS (g) PSO (h) GSA (i) GWO (j) HSA (k) FOA.

**Table 5.** A quantitative analysis of different metaheuristic algorithms.

Algorithms	Advantages	Disadvantages	Solving Capability of Simple Problem	Solving Capability of Complex Problems	Computational Time	Depends on Initial Solution
GA	<ul style="list-style-type: none"> <li>• Flexible</li> <li>• Parallel processing is possible</li> </ul>	<ul style="list-style-type: none"> <li>• Risk of premature convergence</li> <li>• Limited understanding of results</li> </ul>	Excellent	Poor	For larger datasets, requires huge computational time.	No
PSO	<ul style="list-style-type: none"> <li>• Tuning of fewer parameters</li> <li>• Simpler constraints</li> <li>• Can be applied to multi-objective optimization problem</li> </ul>	<ul style="list-style-type: none"> <li>• Premature convergence</li> <li>• Low-quality solution</li> </ul>	Excellent	Excellent	PSO requires much less time to converge than GA	No
FOA	<ul style="list-style-type: none"> <li>• Clear principle</li> <li>• Fewer parameters</li> </ul>	<ul style="list-style-type: none"> <li>• Falls into local optima</li> <li>• Strategic update is fixed</li> <li>• Fitness function is always positive</li> </ul>	Excellent	Poor	Simple computational process	Yes
ABC	<ul style="list-style-type: none"> <li>• Strong equation-searching ability</li> <li>• Simpler process</li> <li>• Avoids falling into local optima</li> </ul>	<ul style="list-style-type: none"> <li>• Insufficient population diversity</li> <li>• Weak developing capacity</li> </ul>	Excellent	Excellent	Slow global convergence	No
CS	<ul style="list-style-type: none"> <li>• Guaranteed global optima</li> <li>• Balanced mixing ability</li> </ul>	<ul style="list-style-type: none"> <li>• Requires tuning of parameters</li> </ul>	Excellent	Excellent	Slow convergence rate	No
GSA	<ul style="list-style-type: none"> <li>• Easier implementation</li> <li>• Adaptive learning capability</li> <li>• Offers high-precision results</li> </ul>	<ul style="list-style-type: none"> <li>• Gets stuck in local solution in last iteration</li> <li>• Parameters are complex</li> <li>• Becomes inactive after convergence</li> </ul>	Excellent	Poor	Less computation time	No
GWO	<ul style="list-style-type: none"> <li>• Requires fewer parameters</li> <li>• Simpler structure</li> <li>• Easier implementation</li> </ul>	<ul style="list-style-type: none"> <li>• Low solution accuracy</li> <li>• Falls into local optimum</li> </ul>	Excellent	Excellent	Slow convergence rate	No
GOA	<ul style="list-style-type: none"> <li>• Easier development</li> <li>• Higher accuracy</li> <li>• Offers high-precision results</li> </ul>	<ul style="list-style-type: none"> <li>• Falls into local optimum easily</li> <li>• Problems in exploiting search space</li> <li>• No theoretical convergence property</li> </ul>	Good	Excellent	Slow convergence rate	No
BA	<ul style="list-style-type: none"> <li>• Easier implementation</li> <li>• Shows efficiency in solving continuous and discrete problems.</li> </ul>	<ul style="list-style-type: none"> <li>• Premature convergence</li> <li>• Low precision</li> </ul>	Excellent	Excellent	Convergence becomes slower in later stages	No
WOA	<ul style="list-style-type: none"> <li>• Simpler structure</li> <li>• Higher accuracy</li> </ul>	<ul style="list-style-type: none"> <li>• Cannot handle variable inputs</li> <li>• Global optimum solution is not guaranteed</li> </ul>	Excellent	Excellent	Faster convergence rate	Yes

## 8. A Generalized Approach for Load-Forecasting Procedure

A generalized approach shown in Figure 7 is followed to develop a load-forecasting model, which is illustrated as follows:

Step 1: Collect historical load, weather, and event data from meters, data servers, etc.

Step 2: Prepare the load data.

Step 3: Analyze the load, weather, and event data.

Step 4: Prepare the model for the selected dataset.

Step 5: Choose an algorithm depending on time horizons and input parameters.

Step 6: Check whether the algorithm is appropriate for the given dataset or not.

Step 7: If not appropriate, then the hyperparameters are tuned using metaheuristic algorithms. In Step 7, the following steps are undertaken:

7.1. Parameters such as weights, threshold, bias, smoothing factor, and learning rate of forecasting methods need to be initialized.

7.2. Initial position and maximum number of iterations need to be set.

7.3. Read the load characteristics at a specific point.

7.4. Run the forecasting model and calculate the values at each specific point

7.5. Calculate the fitness function.

7.6. Check whether the stopping criteria are met. If yes, then go to Step 5.

7.7. If the stopping criteria are not met, then update the position and go to Step 7.4.

Step 8: If the algorithm is appropriate, then refine the model.

Step 9: Check whether there is any change in data or not.

Step 10: If there is any change in data, then go to Step 3.

Step 11: If there are no changes in data, then the model can be run for load forecasting.

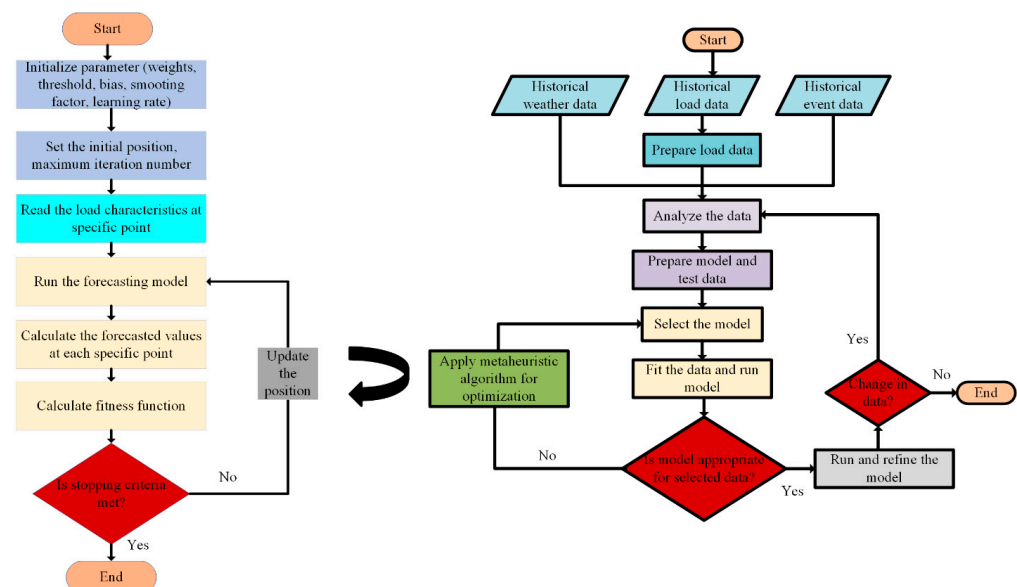


Figure 7. A generalized approach for applying metaheuristic algorithms in load forecasting.

## 9. An Overview of Short-Term Load Forecasting

A hybrid model consisting of Seasonal Support Vector Regression (SVR) and a Chaotic Gravitational Search Algorithm (CGSA) is proposed in [105]. To refine the searching space, the chaotic mapping function is applied to GSA. Electricity demand depends on seasonal factors, which are considered in this article by a seasonal mechanism in conjunction with SVR. The load-forecasting performance is improved from 2.587% to 3.199% by including seasonal indices.

Similarly, hyperparameters of SVR are optimized by the Particle Swarm Pattern Search optimization (PSwarm) algorithm [106]. The advantages of global optimization, such as Particle Swarm, and local minimization, such as Pattern Search, are combined to form this

hybrid algorithm called PSwarm. The hourly-based dataset used in this article is from a North American utility [107–109].

The feature selection and parameter optimization of SVR are carried out simultaneously by Comprehensive Learning Particle Swarm Optimization (CLPSO) in the framework of the Memetic Algorithm (MA) [110]. A similar approach has been adopted in [111] for short-term load forecasting. The article proposes a Seasonal SVR (SSVR) algorithm with Chaotic Simulated Annealing (CSA). Again, a cloud theory is employed with CSA to overcome the issue of the temperature annealing process. The model is tested on two datasets: one is from Northeastern China, and the other one is from New York Independent System Operator (NYISO), New York City.

Another article proposes SVR with a metaheuristic algorithm to improve the prediction accuracy [112]. Metaheuristic algorithm such as Tabu Search (TS) shows premature convergence and local optima trapping. To avoid these shortcomings, quantum computing mechanics is applied to TS. The forecasting index MAPE shows a smaller value of 1.32% for SVR-CQTS than 1.89% for SVR-QTS.

To do microgrid load forecasting, the Least-Squares Support Vector Regression (LSSVR) model with a metaheuristic algorithm is proposed [113]. A Fruit-fly Optimization Algorithm (FOA) is used in this article which mimics the foraging behavior of fruit flies. To overcome the issues of premature convergence and becoming trapped in local optima faced by FOA, QCM is used to add quantum behavior, and a cat-mapping function is adopted to help a fruit-fly not to become trapped in local optima. This model shows a minimum error (MAPE = 1.01%) than the other existing methods.

A modified fruit-fly algorithm (MFOA) is used to optimize the parameters of SVR for STLF [114]. Using the modified version, the prediction error is decreased from SVR-FOA (MAPE = 1.8051%) to SVR-MFOA (MAPE = 1.6909%).

A hybrid model with a metaheuristic algorithm is proposed for short-term load forecasting of microgrids [115]. Empirical mode decomposition (EMD) is used to decompose the load data into Intrinsic mode function (IMF). Prediction of IMF components is carried out by two forecasting algorithms named Extended Kalman filter (EKF) and Extreme Learning Machine with Kernel (KELM). The parameters of this model are optimized by the Particle swarm optimization (PSO) algorithm.

The article in [116] follows the same approach as [115]. The only difference is that SVR is used rather than EKF and KELM to construct IMFs.

For short-term load forecasting, a hybrid model consisting of EMD, PSO, Genetic algorithm (GA), and SVR is proposed [117]. As discussed above, EMD is used for the decomposition of data series into lower and higher frequency components. The higher and lower frequency parameters of the SVR model are optimized by PSO and GA, respectively.

Another application of EMD is demonstrated in [118], which is combined with seasonal adjustment, PSO, and Least square support vector machine (LSSVM). After decomposing into smaller components, seasonal components are eliminated, and LSSVM is used to model the resultant data series, which is then optimized by PSO. The final prediction result is achieved by multiplying the seasonal indexes by the forecast results from PLSSVM.

The proposed model consisting of EMD, Gray rational analysis (GRA), MPSO, and LSSVM is used for short-term load forecasting and tested on the Jibei area of China [61]. The data series is decomposed into smaller subsequences by EMD and GRA. Then, these subsequences are forecasted by MPSO and LSSVM. The combined model shows better performance than BP, SVM, LSSVM, PSO-LSSVM, MSO-LSSVM, and EMD-MPSO-LSSVM.

As EMD is used to decompose the data, an end effect is produced during this process, which will vary the final result. To eliminate the end effect, an improved version of EMD (IEMD) is proposed to be hybridized with Autoregressive integrated moving average (ARIMA), Wavelet neural network (WNN), and FOA [119]. An extension version of EMD known as Ensemble empirical mode decomposition (EEMD) is used in [120] along with the ARIMA and Culture Particle Swarm Optimization (CPSO) algorithm.

Another hybrid model based on EEMD and subsection PSO (SS-PSO) is tested on the dataset from the Chongqing grid in China [121]. Each section of the optimization is divided into 12 subsections which will have only one minima in each subsection. A comparison is made among these 12 minimum values from 12 subsections, and the last optimization value is chosen. In the EMD model, IMFs only contain information at a specific time scale. However, if the input data have a change at a certain time scale, IMFs may contain these components, which is difficult to predict for EMD. Therefore, EEMD adds white noise to the input data which is distributed for different time scales.

Though EEMD can eliminate the added noise, it cannot effectively neutralize the noise. That is why a complete ensemble empirical mode decomposition (CEEMD) along with the Whale optimization algorithm (WHA) and LSSVM is used for load forecasting [122]. The WHA algorithm takes inspiration from the hunting behavior of humpback whales.

The article [123] proposes a Support vector machine (SVM) algorithm for load forecasting supported by PSO. In this model, significant temperature variations are considered.

A Complete Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is used with SVM and the Modified Gray Wolf Optimization Algorithm (MGWO) [124]. The daily peak load is decomposed into multiple smaller sequences, and then SVM and MGWO are used to predict these smaller sequences. The adaptive white noise smoothing factor is added in decomposition by CEEMDAN to utilize the characteristics of mean Gaussian white noise to eliminate the mode mixing. An improved CEEMDAN (ICEEMDAN) is used for data preprocessing in [125], along with the Elman neural network (ELM), whose parameters are optimized by the multi-objective dragonfly algorithm (MODA). The behavior of dragonflies is mimicked in DA.

A hybrid model combining SVM and Grasshopper optimization algorithm (GOA) is proposed for STLF [126]. The proposed model uses a similar day approach as it considers that the local climate conditions show better performance (MAPE = 1.5%) than other methods, such as GA-SVM (MAPE = 2.13%) and PSO-SVM (MAPE = 1.94%). The nonlinear behavior of load data makes it difficult to load forecasting.

A combined model of SVM and Singular spectrum analysis (SSA) and Cuckoo Search (CS) is proposed in [127]. To analyze the time series data, SSA is used to identify and extract trends or noise. The behavior of cuckoo birds is followed to develop the CS algorithm. The optimization of the parameters of SVR is completed by the CS algorithm. The model shows better performance than SVM, SSA-SVM, CS-SVM, SARIMA and BPNN.

The authors have proposed a hybrid method comprising SVM and the Manta Ray Foraging Optimization (MRFO) algorithm [128]. MRFO is based on the eating behavior of the manta rays living in the ocean. A comparison is made among the proposed MRFO algorithm (RMSE = 4.715) and the other existing metaheuristic algorithms Slime Mold Algorithm (SMA) (RMSE = 8.7450), Tug of war optimization (TWO) (RMSE = 9.159), Moth flame optimization (MFO) (RMSE = 9.075), Satin bowerbird optimization (SBO) (RMSE = 9.248) and FOA (RMSE = 9.740).

An SVR model with Differential empirical mode decomposition (DEMD) and Quantum Particle swarm optimization (QPSO) is proposed to do the load forecasting [129]. Quantum mechanics helps to resolve the issue of premature convergence of PSO. The data series is decomposed into IMFs, and SVR is used for high-frequency data forecasting chosen by QPSO. Autoregressive (AR) modeling is used to forecast the residuals as it shows monotonous behavior.

Another SVR model with chaotic quantum particle swarm optimization (SVRCQPSO) is proposed [130]. The chaotic phenomenon is applied by keeping the diversities of the components of PSO to avoid trapping onto a local optimum solution. If the Eastern region is considered, then SVRCQPSO shows better performance (MAPE= 1.5940%) than SVRCQPSO (MAPE = 1.9830%). A similar approach is followed in [131] with a difference of using GA rather than PSO. A comparison is made between SVRCQPSO and SVRCQGA for the Eastern region, and SVRCQGA shows better performance (MAPE = 1.5180%) than SVRCQPSO (MAPE = 1.5940%).

A learning algorithm named Extreme Learning Machine (ELM) is proposed with switching delayed particle swarm optimization (SDPSO) to obtain better forecasting results [132]. The proposed method shows 0.72% less MAPE than the state-of-the-art method Radial basis function neural network (RBFNN).

A combination of Wavelet Transform (WT) and Gray Model (GM), whose coefficients are optimized by PSO, is proposed in [64]. The input data series includes the temperature, humidity, wind speed, and day-ahead load data. Another application of WT in combination with LSSVM and FOA is proposed in [133]. To validate the proposed model, it is compared with WT-LSSVM (MAPE = 1.111%), FOA-LSSVM (MAPE = 1.353%), PSO-LSSVM (MAPE = 1.414%), and LSSVM (MAPE = 1.8457) and shows better performance (MAPE = 1.068%).

A hybrid model comprising of Bayesian neural network (BNN), Discrete wavelet transform (DWT), and GA is proposed for load prediction [134]. The input data are decomposed by DWT into different resolutions of components to extract the nonlinear information of load data. The Bayesian approach is to train NN by assigning a probability density function.

The optimization of weighting coefficients of different components is completed by GA. A hybrid model is proposed in [135] consisting Artificial Neural Network (ANN) optimized by an Artificial Bee Colony (ABC) metaheuristic algorithm. The behavior of honeybees while searching for food is the inspiration behind this algorithm. A comparison is made among ANN-ABC, ANN-GA, and ANN-PSO using an evaluation index MSE, which shows  $7.16 \times 10^{-4}$ ,  $3.95 \times 10^{-3}$ , and  $8.79 \times 10^{-4}$ , respectively.

The article proposes a combined model of Back-Propagation Neural Network (BPNN), Radical Basis Function Neural Network (RBFNN), Generalized Regression Neural Network (GRNN), and Genetic Algorithm Back-Propagation Neural Network (GABPNN) where the Cuckoo Search (CS) algorithm is used to optimize the coefficients of each model [136]. Cuckoos' nature of laying eggs and breeding gives the inspiration to develop CS.

A modified GRNN is proposed with a Multi-Objective Firefly Algorithm (MOFA), which considers seasonal patterns and data-preprocessing techniques [137]. The weighting coefficients and thresholds of GRNN are optimized by MOFA. Again, to deal with the nonlinearity of the load dataset, a hybrid model consisting of GRNN and FOA with decreasing step size is proposed [138].

A two-step parameter optimization based on Grid Traverse Algorithm (GTA) and PSO is used in combination with SVR for short-term load forecasting. To narrow the search space from global solution to local solution, GTA is used, and PSO selects the best solution from the local solutions that can be selected for SVR. Then, the SVR function determines the forecasted value [139].

The authors in [53] propose five ANN-based models such as BPNN, GABPNN, WNN, RBFNN, and GRNN. The original data are integrated into a dataset constructed by multiple seasonal patterns, and EMD is used to decompose this dataset into smaller IMFs. Then, ANN-based models are used to forecast the IMFs. Each model gives different forecasting values, which are then added to be optimized by a multi-objective flower pollination algorithm (MOFPA). A similar approach is followed to find the forecasting value combining BPNN, Cuckoo Search BPNN (CSBPNN), GRNN, FOAGRNN, and RBFNN with Non-dominated sorting genetic algorithm III (NSGA-III) [140].

Artificial intelligent-based Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are proposed for load forecasting [141]. The hyperparameters of both models require appropriate tuning to improve efficiency. Therefore, the Boosted Self-Adaptive Sine Cosine Algorithm (BSA-SCA) is used to tune the parameters.

## 10. Results

A prediction framework is proposed in this article, which contains a data-decomposition layer, forecasting layer and optimization layer. To validate the efficacy of this framework, a comparison has been made among different models shown in Table 6. Several observations can be made from this table:



**Remarks 1.** Forecasting models can predict results with the highest MAPE, such as SVR, LSSVM, LSTM, and RBFNN, showing 6.0183, 3.5215, 1.1829, and 3.6645, respectively, within their respective datasets.

**Remarks 2.** The forecasting layer, along with metaheuristic algorithms, shows superior performance than the forecasting layer with the data-decomposition layer. It also shows the importance of the hyperparameters tuning of the model using metaheuristic algorithms. For example, LSTM-WOA shows a MAPE of 0.7615 which is lower than 0.8 of VMD-LSTM [122].

**Remarks 3.** In the data-decomposition layer, CEEMD shows better performance than EEMD and EMD [122].

**Remarks 4.** A combination of the data-decomposition layer, forecasting layer, and optimization layer shows superior performance than any other combination, such as only the forecasting layer or forecasting and data-decomposition layer or forecasting and optimization layer. For example, VMD-LSTM-WOA has shown superior performance with MAPE of 0.6986 than LSTM-WOA (MAPE = 0.7615), VMD-LSTM (MAPE = 0.8) and LSTM (MAPE = 1.1829) [142].

**Table 6.** MAPE for different models proposed in the existing literature.

Reference	Model	MAPE (%)	Reference	Model	MAPE (%)
[116]	EMD-SVR-PSO	3.4323	[143]	NN-DE	0.7173
	EMD-SVR	3.9898		WT-CNN	1.0808
	SVRPSO	5.9826		VMD-GRNN-GSA	0.6722
	SVR	6.0183		EEMD-BPNN-FPA	1.0731
[122]	CEEMD-LSSVM-WOA	1.5602	[144]	WT-BPNN	1.7317
	EEMD-LSSVM-WOA	1.5724		EEMD-BPNN	1.2104
	EMD-LSSVM-WOA	3.1486		BPNN-CS	1.6463
	LSSVM-WOA	3.3885		EEMD-SVM-WOA	1.3249
	LSSVM	3.5215		RBFNN	3.6645
[142]	LSTM	1.1829	[145]	ARIMA	3.3396
	VMD-LSTM	0.8		BPNN	3.1382
	LSTM-WOA	0.7615			
	VMD-LSTM-WOA	0.6986			

Again, the accuracy and stability of forecasting models largely depend on the type of datasets. Therefore, to evaluate the performance, they must use the same datasets. After doing a careful investigation, it is found that articles [113,130,131] use the data from the Global Energy Forecasting Competition (GEFCOM), 2014. Figure 8 shows a comparative analysis of different forecasting models used in these three articles. These articles use different evaluation indices. However, MAPE is the most common among these three. The MAPE of LSSVR-CQFOA is 1.02, which is the lowest among these algorithms. Therefore, it shows superior performance than the other methods. The next highest-performing algorithms are SVRCQBA and SVRCQGA, whose MAPEs are 1.07 and 1.16, respectively. This also shows a similar interpretation of the results from the previous discussion: (1) properly tuning hyperparameters increases the prediction accuracy (2) choosing the right metaheuristic algorithm can increase the performance of the models.

A taxonomy is presented in Table 7, which includes the year of publication, data type, resolution of the dataset, evaluation indices, compared methods and contributions of the proposed methodology, and their advantages and drawbacks.

Table 7. A taxonomy of the state-of-the-art articles.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[141]	2023	LSTM-BSA-SCA	<ul style="list-style-type: none"> <li>LSTM introduces a memory state cell to store information effectively.</li> <li>GRU passes information through network.</li> <li>Exploitation and exploration are boosted by BSA.</li> </ul>	<ul style="list-style-type: none"> <li>Huge computation burden.</li> <li>Modest population size is used.</li> <li>Only a few iterations are conducted to find optima.</li> </ul>	Dataset from ENTSO-E [146]	Hourly	Solar, wind, weather, and load data have been considered.	LSTM-SCA, LSTM-ABC, LSTM-FA, LSTM-S	MSE, RMSE, MAE, R <sup>2</sup>
[147]	2023	MFGBM-HIGWO	<ul style="list-style-type: none"> <li>MFGBM shows higher prediction accuracy.</li> <li>Theoretical solution of complex model is possible.</li> <li>HIGWO can be applied to nonlinear high-dimensional model.</li> </ul>	<ul style="list-style-type: none"> <li>Interaction between variables is ignored.</li> </ul>	Dataset from Sichuan Province, China		The proposed method has a better nonlinear fitting, and the traditional GWO method is also improvised.	LSSVR, AdaBoost, RFS	MAPE, RMSE, MAE
[128]	2023	SVR-MRFO	<ul style="list-style-type: none"> <li>SVR can handle larger and nonlinear datasets.</li> <li>The proposed model shows 99.9% and 99.3% accuracy for training and testing datasets.</li> </ul>	<ul style="list-style-type: none"> <li>Accuracy is affected by initial parameters.</li> </ul>			The classical SVR model only predicts point values that can be overcome by MRFO, as it shows error controllability and fast convergence.	PSO-SVR, PSO-BP, EMD-SVR-AR, DEMD-SVR-AR, AFCM, ARIMA, SVRCGSA, SSVRCGSA	MSE, RMSE, MAE, MAPE, RAE, R <sup>2</sup> , NMSE
[148]	2023	FE-AGO-LWSVR	<ul style="list-style-type: none"> <li>Dimensionality reduction is solved by FE.</li> <li>AGO tunes the weighting coefficient and bias of LWSVR to provide stable, robust, and fast solutions.</li> </ul>	<ul style="list-style-type: none"> <li>Considered only historical data.</li> <li>Key factors affecting the forecasting are not considered.</li> </ul>	Dataset from NSW, Victoria, Australia and CAISO, USA	Hourly	The proposed method improves precision, stability, and convergence rate.	NARX, DNN, GTB, RF	MAPE, MAE
[149]	2023	ECWOA-SVR	<ul style="list-style-type: none"> <li>Initial values are generated by a chaotic mechanism, which improves the convergence rate.</li> <li>Population diversity is increased by elite opposition-based learning strategy. It helps to avoid falling into local optima.</li> </ul>	<ul style="list-style-type: none"> <li>Feature selection is not added.</li> <li>Online testing is not possible.</li> </ul>	Load data from Singapore and load and price data from GEFCOM, 2014.	Half-hourly and hourly	SVR is combined with WOA to balance the exploration and exploitation of the algorithm.	SVR, WOA-SVR, PSO-SVR, BPNN	MSE, RMSE, MAE, MAPE, R <sup>2</sup>

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[150]	2023	PSO-VMD-TCN-Attention	<ul style="list-style-type: none"> <li>Modal number and bandwidth constraint re optimized by PSO.</li> <li>Feature extraction is easier.</li> <li>TCN can handle longer sequences.</li> </ul>	<ul style="list-style-type: none"> <li>The number of iterations is small.</li> <li>Hard to implement.</li> </ul>	Dataset from Panama case published on Kaggle.	Hourly	Manual adjustments of parameters required by VMD are overcome using PSO.	PSO-VMD-LSTM, PSO-VMD-GRU, LSTM, GRU, TCN	MSE, RMSE, MAE, MAPE
[151]	2023	CS-GWO-DA-BiGRU	<ul style="list-style-type: none"> <li>Sensitivity to key features is improved by DA.</li> <li>Population diversity and global search ability are enhanced by CS-GWO.</li> </ul>	<ul style="list-style-type: none"> <li>Leads to unstable operation when exposed to new data series.</li> </ul>			Combination of feature and temporal attention mechanisms is used to form DA.	DA-BiGRU, PSO-DA-BiGRU, WOA-DA-BiGRU, CSO-DA-BiGRU	RMSE, MAE, SMAPE, R <sup>2</sup>
[152]	2023	WHODL-STLFS	<ul style="list-style-type: none"> <li>Computational complexity is reduced by WHO.</li> <li>Prediction accuracy is improved by parameter optimization by AAO.</li> </ul>	<ul style="list-style-type: none"> <li>Stability and convergence rates are ignored.</li> </ul>	Dataset from FE and Dayton grid	Hourly	A three-stage process is proposed combining WHODL-STLFS, ALSTM and AAO.	FCRBM, AFC-ANN, Bi-level, MI-ANN, LSTM	MAPE
[153]	2023	Bi-LSTM + Dropout-LF-PSO	<ul style="list-style-type: none"> <li>Convergence rate of LF-PSO is better than PSO.</li> <li>LF-PSO explores new search space to improve the tuning of hyperparameters.</li> </ul>	<ul style="list-style-type: none"> <li>Computational complexity increases with the increase of larger datasets.</li> <li>Depends on the initial values.</li> </ul>	Dataset from Smart Grid Smart City, Australia.	Half-hourly	The proposed method has outperformed the other state-of-the-art methods in terms of forecasting accuracy.	LSTM, GRU, SVR, ARIMA	MAE, RMSE, MAPE
[154]	2023	PSVMD-SSA-CGA	<ul style="list-style-type: none"> <li>Randomness and subjectivity of the parameters of VMD are avoided by SSA.</li> <li>Weighting assignment is solved by CGA.</li> </ul>	<ul style="list-style-type: none"> <li>Computational time is large.</li> </ul>	Load data from Quanzhou, Fujian, China	Half-hourly	PSVMD is used to break down the load data into several quantities and CGA is used for forecasting.	CGA, GA-VMD-CGA, SSA-VMD-CGA	MAE, RMSE, MAPE, R <sup>2</sup>
[155]	2023	LCR-AdaBoost-FA-ELM	<ul style="list-style-type: none"> <li>Implementation is easier.</li> <li>Weighting coefficient and biases of ELM are optimized by FA to reduce the prediction error.</li> </ul>	<ul style="list-style-type: none"> <li>Diversified datasets are not used.</li> </ul>	Load data of furniture factory.	Hourly	A day-ahead load forecasting is proposed and FA combined with ELM can reduce prediction error.	SVR, ELM, FA-SVR, FA-ELM, AdaBoost-FA-SVR	MAE, RMSE, MAPE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[156]	2023	EWT-SSA-GRNN	<ul style="list-style-type: none"> <li>Forecasting errors can be avoided by smoothing the load sequence by EWT.</li> <li>SSA performs better in uncertain environments.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is increased during the data reconstruction stage.</li> </ul>	Dataset from a city in southern Australia	Half-hourly	Problems associated with load forecasting, such as volatility uncertainty, can be solved by this method.	EWT-GRNN, GRNN, LSTM, SVR, CNN-RNN, RNN, VMD-GRNN, VDM-SSA-GFNN	MAE, RMSE, MAPE, MSE, R <sup>2</sup>
[157]	2023	BES-VMD-CNN-Bi-LSTM-EC	<ul style="list-style-type: none"> <li>The nonlinear nature of complex load data is addressed by Bi-LSTM.</li> <li>Stability and security are enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Diversified load types are not considered.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from GEFCOM 2012	Hourly	Prediction accuracy is increased by improving error correction, which considers short-term factors.	RF, SVM, LSTM, GRU, Bi-LSTM, CNN-LSTM, CNN-GRU, CNN-Bi-LSTM, O-VMD-CNN-LSTM, O-VMD-CNN-GRU, O-VMD-CNN-Bi-LSTM, BES-VMD-CNN-Bi-LSTM	MAPE, RMSE
[158]	2023	IPSO-DBiLSTM-VMD-attention mechanism	<ul style="list-style-type: none"> <li>The model can work with extremely volatile and nonlinear load sequence.</li> <li>Prediction accuracy and robustness are enhanced by estimating the parameters of DBi-LSTM using PSO.</li> </ul>	<ul style="list-style-type: none"> <li>Meteorological factors are not considered.</li> </ul>	Dataset from Ninth Electrical Attribute Modeling Competition		The data are decomposed into different quantities by VMD, DBiLSTM is used for price representation of the data, and IPSO helps to avoid local optima and premature convergence.	GS-VMD-DBiLSTM-Attention, PSO-VMD-DBiLSTM-Attention, LSTM, Bi-LSTM, Bi-LSTM-Attention, DBiLSTM, TCN	MAE, RMSE, MAPE, R <sup>2</sup>
[159]	2022	ISOA-SVM	<ul style="list-style-type: none"> <li>To overcome the problem of random feature selection of SVM, ISOA is used.</li> <li>Prediction accuracy and convergence rate are improved by ISOA.</li> </ul>	<ul style="list-style-type: none"> <li>Stability and generalization ability need to be improved.</li> <li>The model must be more universal.</li> </ul>	Load data from a power plant in eastern Slovakia	Half-hourly	Parameters of SVM are optimized by ISOA to improve the optimization performance and convergence rate.	SOA-SVM, SVM, BP	MAE, RMSE, MAPE, R <sup>2</sup>
[160]	2022	FC-FWA-LSSVM	<ul style="list-style-type: none"> <li>Can mitigate the effects of uncorrelated factors.</li> <li>Global search capability is increased.</li> </ul>	<ul style="list-style-type: none"> <li>Overfitting of data is still prevalent.</li> </ul>	Residential data from China	Half-hourly	Fuzzy cluster analysis is used for feature extraction which can reduce the data redundancy and prediction error.	BPNN, LSSVM, FWA-LSSVM	RE, RMSE, MAPE, AAE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[161]	2022	MAFOA-GRNN	<ul style="list-style-type: none"> <li>Can handle structured and unstructured data.</li> <li>Highly adaptable.</li> </ul>	<ul style="list-style-type: none"> <li>Adjustments of input parameters are needed for different datasets.</li> </ul>	Dataset from Wuhan, China	Hourly	Several weather factors are considered here.	PSO-GRNN, FOA-GRNN, DSFOA-GRNN, BP, SVM, GRNN	NRMSE, MAE, MAPE
[19]	2022	MEMD-PSO-SVR	<ul style="list-style-type: none"> <li>Peak load forecasting is possible.</li> <li>MEMD can effectively extract the important features from nonlinear data series.</li> </ul>	<ul style="list-style-type: none"> <li>Only temperature is considered to be the input variable.</li> <li>PSO has a low convergence rate.</li> </ul>	Dataset from NSW, Victoria, Australia	Half-hourly	To reduce the loss of an overestimated or underestimated power system, multi-dimensional input variables are considered.	SVR, BPNN, EEMD-SVR, EEMD-PSO-SVR, MEMD-SVR, MEMD-PSO-BPNN	RMSE, MAPE, R <sup>2</sup> , DA
[162]	2022	CEEMDAN-IGWO-GRU	<ul style="list-style-type: none"> <li>CEEMDAN can effectively suppress the load fluctuation interference.</li> <li>Diversity of load-forecasting conditions is considered.</li> <li>Search performance is improved in IGWO than in GWO.</li> </ul>	<ul style="list-style-type: none"> <li>Depends on the initial population.</li> </ul>	Dataset from Singapore's utility grid.	Half-hourly	CEEMDAN is used to suppress the load fluctuation, and GRU, which is optimized by IGWO, is used for the prediction of each component.	EEMD-GRU-MLR, PSO-VSM, GRU, CEEMDAN-GRU, IGWO-GRU, CIG, BP, ELM, DBN, SAE	MAE, MAPE, RMSE
[163]	2022	CEEMD-SSA-GRU	<ul style="list-style-type: none"> <li>Noise interference is eliminated by CEEMD.</li> <li>Adaptability is enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Some unnecessary IMFs are created.</li> <li>The computational burden is increased.</li> </ul>	Load data of an industrial user's factory		Problem with modal aliasing in historical data is solved, and the relationship between the time series characteristics of load data is explored.	GRU, SSA-GRU, EMD-SSA-GRU	MAE, RMSE, MAPE
[164]	2022	IPSO-Elman	<ul style="list-style-type: none"> <li>To screen the meteorological factors, Pearson coefficients are used.</li> <li>Reliability and prediction accuracy is increased.</li> </ul>	<ul style="list-style-type: none"> <li>Weighting coefficients of climate factors are not considered.</li> </ul>	Dataset from two regions	15 min	Various climate factors that affect load forecasting are considered here.		MAPE, RMSE, MAE
[165]	2022	VMD-CISSA-LSSVM	<ul style="list-style-type: none"> <li>The model shows stable operation, search accuracy, and convergence rate.</li> <li>It follows the trend of load data.</li> </ul>	<ul style="list-style-type: none"> <li>Temperature and holiday effects are not considered.</li> </ul>	Dataset from Shandong, China	Half-hourly	The proposed metaheuristic algorithm has avoided uneven initial population distribution and trapping into local minima.	Elman, ELM, LSSVM, GWO-ELM, PSO-Elman, SSA-LSSVM, CISSA-LSSVM, FA-CSSSA-ELM	MSE, MAE, MAPE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[142]	2022	VMD-IWOA-LSTM	<ul style="list-style-type: none"> <li>Load interference is eliminated.</li> <li>Shows strong practicability.</li> </ul>	<ul style="list-style-type: none"> <li>Complex relationships of load characteristics cannot be extracted.</li> </ul>	Dataset from a power grid company	Half-hourly	The searching area of IWOA is enhanced using a nonlinear attenuation factor and random difference variation.	LSTM, VMD-LSTM, WOA-LSTM, VMD-WOA-LSTM	MAPE, MAE, RMSE
[166]	2022	MSC-PSO-SVR	<ul style="list-style-type: none"> <li>Prediction performance is enhanced as MSC can choose any parameter size.</li> <li>Can handle small and nonlinear datasets.</li> </ul>	<ul style="list-style-type: none"> <li>Cannot handle multi-dimensional optimization problems.</li> </ul>	Dataset from a county of Jiangxi and Germany.		The proposed method can adapt to the candidate size.	BPNN, LSTM, RNN, RF, XGBoost	MAE, MAPE, RMSE, STD
[167]	2022	VMD-mRMR-tsPSO-LSSVR	<ul style="list-style-type: none"> <li>Multi-step prediction uncertainty is reduced.</li> <li>Different data series can be adopted.</li> <li>Prediction accuracy and stability are enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Computational time is high.</li> </ul>	Dataset from California	Hourly	A hybrid algorithm is proposed to enhance the diversity and can perform in extremely noisy environments with the help of PSO.	SVR, ANN, PSO-LSSVR, EMD-LSSVR, VMD-LSSVR, VMD-PSO-LSSVR	MAE, MAPE, MSE, RMSE, R <sup>2</sup>
[168]	2021	FE-SVR-mFFO	<ul style="list-style-type: none"> <li>Three performances are considered: convergence rate, prediction accuracy, and stability.</li> </ul>	<ul style="list-style-type: none"> <li>The model must be verified in more diversified datasets.</li> </ul>	Dataset from AEMO	Half-hourly	mFFO is used to select and tune hyperparameters of SVR, which will improve the convergence rate and prediction accuracy	EMD-SVRPSO, FS-TSFE-PSO, VMD-FFT-IOSVR, DCP-SVM-WO	MAPE, MSE, RMSE, R, WI
[169]	2021	NN-PSO	<ul style="list-style-type: none"> <li>Does not depend on the initial solution.</li> <li>Can detect the nonlinear relationships among datasets.</li> <li>Requires less statistical training.</li> </ul>	<ul style="list-style-type: none"> <li>Meteorological factors such as temperature and humidity are not considered.</li> <li>Number of iterations is not mentioned.</li> <li>It tends to overfit data.</li> </ul>	Dataset from Iran's power grid		PSO is used to tune the parameters for NN, and NN uses the back-propagation method for load forecasting.		MAPE, MAE, MSE
[170]	2021	VMD-GWO-SVR	<ul style="list-style-type: none"> <li>Each component can be predicted separately.</li> <li>Can follow the trend of load fluctuation.</li> </ul>	<ul style="list-style-type: none"> <li>Time series is being delayed.</li> <li>Faces problems while dealing with high-frequency components.</li> </ul>	Dataset from the power grid of Oslo and surrounding regions.	Hourly	The proposed method separates the important information from load data and, therefore, predicts the trend in load change.	SVR, VMD-SVR, GWO-SVR	MAE, MAPE, MSE, R <sup>2</sup>

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[171]	2021	HWOA-ELM	<ul style="list-style-type: none"> <li>Can do load prediction in a cloud environment.</li> <li>Random parameters influence is addressed.</li> </ul>	<ul style="list-style-type: none"> <li>Computation time is high.</li> <li>Generalization capability is low.</li> </ul>	Datasets from real-world measurements		Trapping into local optima and poor convergence rate is solved by the proposed method.	WOA-ELM, BSA-ELM, CSO-ELM, GWO-ELM	RMSE, MAPE, R <sup>2</sup>
[172]	2021	IGA-LS-SVM	<ul style="list-style-type: none"> <li>Uses equality constraints.</li> <li>Solves the load-forecasting problem by solving a set of linear equations.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> <li>Relative errors between variables need to be reduced to achieve higher accuracy.</li> </ul>	Dataset from Yunnan province.		Temperature, meteorological factors, holidays, and other factors affecting load forecasting are considered.	BP, LS-SVM	RMSE
[173]	2021	SVR-LR-RF-PSO	<ul style="list-style-type: none"> <li>Have excellent fitting ability.</li> <li>Shows robustness.</li> </ul>	<ul style="list-style-type: none"> <li>A comparison is being made only with single models.</li> <li>Number of iterations is quite small.</li> </ul>	Dataset from NSW, Australia	Half-hourly	A weighting factor is applied to three individual models.	SVR, LR RF	MAE, MAPE, RMSE, R <sup>2</sup>
[174]	2021	CNN-CHIO	<ul style="list-style-type: none"> <li>It reduces the overfitting of classifiers.</li> <li>CHIO uses its classifiers to increase the performance of CNN classifiers.</li> </ul>	<ul style="list-style-type: none"> <li>Only the residential dataset is used.</li> <li>Computational burden is high.</li> </ul>	Dataset from ISO-NE		Classifiers are used to extract the features from the dataset, and CHIO is used to tune the parameters.	SVM, RF, LR, LDA	MAPE, RMSE, MSE, MAE
[175]	2021	EOBL-CSSA-LSSVM	<ul style="list-style-type: none"> <li>Can eliminate the noise effect.</li> <li>Can improve the deficiencies of the machine-learning model.</li> </ul>	<ul style="list-style-type: none"> <li>Weather and holiday factors are not considered.</li> </ul>	Dataset from South-eastern grid, Australia	Half-hourly	VMD can reduce the noise effect, whereas metaheuristic algorithms can improve prediction accuracy.	Elman, PSO-ESN, SA-LSSVM, CAWOA-ELM, FA-CSSA-ELM	RMSE, MAPE, MAE, MSE
[176]	2021	SMN-PSO	<ul style="list-style-type: none"> <li>Adaptive to online learning.</li> <li>Has a low computational burden.</li> </ul>	<ul style="list-style-type: none"> <li>Number of iterations is not discussed.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from AEMO, Australia	Half-hourly	Different PSO variants are considered here.	EMD-DBN	RMSE, MAPE, MAE, MSE
[143]	2020	VMD-NNGSA-GRNNGSA	<ul style="list-style-type: none"> <li>Features can be selected effectively by combining NN and GSA from different signals.</li> <li>Gets higher prediction accuracy while considering different seasons.</li> </ul>	<ul style="list-style-type: none"> <li>Moderate computational burden.</li> </ul>	Dataset from PJM and Spanish electricity market.		A combinational forecasting algorithm is proposed, which can select the best inputs and outperform other state-of-the-art methods.	RBF, GRNN, NN-DE, WT-CNN	RMSE, MAE, MAPE, TIC



Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[177]	2020	MFNN-GOA	<ul style="list-style-type: none"> <li>Input variables are selected properly, which results in a short prediction time.</li> <li>Reliability of the system is enhanced.</li> <li>Loads can be forecasted at different times.</li> </ul>	<ul style="list-style-type: none"> <li>Iterations number is not provided.</li> </ul>	Dataset from the Youth power station, Salhiya	Hourly	Temperature and other factors affecting the load data are considered.	MFNN, MFNN-GA, MFNN-GWO	RMSE, MAE, MAPE
[178]	2020	PSO-ENN	<ul style="list-style-type: none"> <li>The key parameter is the learning rate, which is optimized by PSO.</li> <li>Simple structure.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> <li>Poor generalization capability.</li> <li>Depends on initial conditions.</li> </ul>	Dataset from Eastern Slovakia	Half-hourly	The learning rate of ENN can be found dynamically by PSO	ENN, GRNN, BPNN	RMSE, MAPE
[179]	2020	ICS-FARIMA	<ul style="list-style-type: none"> <li>Global optimization capability is enhanced.</li> <li>Nonlinear data series is properly preprocessed.</li> </ul>	<ul style="list-style-type: none"> <li>Slow convergence rate.</li> <li>Sensitive to scaling of data.</li> </ul>	Dataset from EIRGIRD, Ireland		ICS is used for parameter optimization of the forecasting algorithm.	RBF, RNN, FARIMA	MAPE, MAE
[180]	2020	ELM-RNN-SVM-MOPSO	<ul style="list-style-type: none"> <li>It increases the accuracy and stability.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> </ul>	Datasets from NSW, Queensland, and Victoria, Australia	Half-hourly	A multi-step forecasting algorithm is proposed where MOPSO is used for optimizing the weighting coefficients.	ELM, RNN, SVM	MAPE, MAE, RMSE
[181]	2020	QPSO-mFBM	<ul style="list-style-type: none"> <li>Stationary load series can be generated from nonstationary series by mFBM.</li> <li>QPSO shows superior performance than PSO.</li> </ul>	<ul style="list-style-type: none"> <li>Self-similarity exists.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from Eastern Slovakia	Half-hourly	QPSO avoids trapping into local optima by searching for a global solution.	FBM, PSO-mFBM, RNN	Max, mean, median, std. deviation
[182]	2020	GBRBM-GA	<ul style="list-style-type: none"> <li>Robustness.</li> <li>Error constraints are trained in a precise direction.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> </ul>	Dataset from Tianjin power station, China	15 min	Demand-side management is considered	PM, ARIMA, ANN, SVR	MAPE, RMSE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[47]	2020	SVM-IPSO	<ul style="list-style-type: none"> <li>Only relevant information is extracted.</li> <li>Prediction accuracy is enhanced by extracting holiday information.</li> <li>Second-order oscillation is used to increase the accuracy of SVM.</li> </ul>	<ul style="list-style-type: none"> <li>Features are chosen artificially.</li> <li>If the price is not considered, the results show similar to real value.</li> </ul>	Dataset from the Singapore power market	Half-hourly	Real-time electricity price is considered.	mRMR-GA-LSTM, BPNN, mRMR-BPNN	MAPE, MAE, RMSE, IA
[183]	2020	LSSVM-ELM-GRNN-WOA	<ul style="list-style-type: none"> <li>Linear equations are used for optimization problems.</li> <li>It shows fitting capability to nonlinear data series.</li> </ul>	<ul style="list-style-type: none"> <li>Meteorological factors are not considered.</li> </ul>	Dataset from NSW, Australia	Half-hourly	WOA is proposed to optimize the weighting coefficients of the combined model.	ARIMA, BP, GRNN	AE, MAE, RMSE, NMSE, MAPE
[184]	2020	ICEEMDAN-GWO-MKELM	<ul style="list-style-type: none"> <li>Can extract important information efficiently.</li> <li>Forecasting ability is improved by optimizing the weight and parameters of each kernel.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> </ul>	Dataset for NSW, TAS, Queensland, Victoria, SA from AEMO	Half-hourly	GWO is used to optimize the parameters of the kernel for ELM.	ICEEMDAN-ANN, ICEEMDAN-DBN, ICEEMDAN-ELM, ICEEMDAN-KELM, ICEEMDAN-RF, ICEEMDAN-SVR	MAE, MAPE, RMSE
[185]	2020	EMD-IGOA-PCA-ARIMA-IFPA-NN-WT	<ul style="list-style-type: none"> <li>The important features which are related to load demand are selected properly.</li> <li>IFPA is used to optimize the weights of NN to avoid overtraining.</li> </ul>	<ul style="list-style-type: none"> <li>Cannot handle large datasets.</li> <li>Features and factors are not selected concurrently.</li> </ul>	Dataset from Iran's electricity market		IGOA is used to select the best features and IFPA is used for optimization of weighting coefficients.	ARIMA, ADEBPNN, IEMDAW, TSOGA	MAPE, MAE, RMSE
[186]	2019	CF-SA-FFOA-SVM	<ul style="list-style-type: none"> <li>The prediction result is improved using the maximum temperature as the input variable.</li> <li>The prediction accuracy is also enhanced by grouping the raw data.</li> </ul>	<ul style="list-style-type: none"> <li>Number of iterations is low.</li> </ul>	Gas data from PetroChina Kunlun Gas Ltd.		The proposed algorithm considers the influence of temperature types.	PSO-SVM, BPNN, GM, ARIMA	MAPE, RMSE, MSE
[187]	2019	FNN-SCG-IBA-DWT	<ul style="list-style-type: none"> <li>Data are selected effectively.</li> <li>NN learning accuracy is improved.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> </ul>	Dataset from Portuguese National Electricity Transmission Grid and New England, USA	15 min and hourly	IBA is used for parameter selection over two optimization layers.	Elman-NN, RBF-NN, SVM, MRMRMS-RBF, MRMRMS-MLP, MRMRMS-WNN	MAPE, RMSE, MAE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[188]	2019	LMD-GSA-PSO-WNN, LMD-GSA-PSO-SVM, LMD-GSA-PSO-BPNN	<ul style="list-style-type: none"> <li>Data-preprocessing technique is improved.</li> <li>Extracts important information effectively.</li> <li>Fast convergence rate.</li> </ul>	<ul style="list-style-type: none"> <li>It is difficult to manage the weighting coefficients of two different metaheuristic algorithms combined.</li> </ul>	Dataset from Queensland, Australia	Half-hourly	Approximation of actual values can be done by the proposed method, which can be applied to a smart grid.	LMD-GSA-BPNN, GSA-PSO-BPNN, LMD-PSO-BPNN, LMD-GSA-WNN, LMD-PSO-WNN	MAE, MAPE, RMSE, R <sup>2</sup> , DA
[189]	2019	ELM-GA and SVM-GS	<ul style="list-style-type: none"> <li>Performance and accuracy of the model are enhanced.</li> <li>Redundant information is eliminated effectively.</li> </ul>	<ul style="list-style-type: none"> <li>The performance of the classifiers is low.</li> </ul>	Dataset from ISO-NE		Deep-learning methods are used to optimize the parameters.	LG, LM, LDA, ELM, SVM	RMSE, MAPE
[190]	2019	AS-GCLSSVM	<ul style="list-style-type: none"> <li>Input features are selected optimally.</li> <li>Cross-validation.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> <li>Implementation is complex.</li> <li>Nonlinear relationships among datasets are not considered.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from NSW, Vitoria, Queensland.	Half-hourly	The parameters of LSSVM are optimized by GWO and CV.	RS-LSSVM, PS-LSSVM, AS-LSSVM, PS-GCLSSVM, AS-GCLSSVM, RF-ANN	MAPE, MAE, R <sup>2</sup>
[144]	2019	EEMD-CSFPA-BPNN	<ul style="list-style-type: none"> <li>Complexity is reduced bus mooting the load series.</li> <li>CSFPA changes the fixed frequency during the exploration and exploitation stage.</li> <li>The model is fully data-driven.</li> </ul>	<ul style="list-style-type: none"> <li>Meteorological factors are not considered.</li> <li>Low number of iterations.</li> </ul>	Dataset from AEMO and IESO	Half-hourly and hourly	CSFPA enhances the forecasting performance and helps to create initial population and switch probability.	Cuckoo-BPNN, EEMD-BPNN, WT-BPNN	MAE, RMSE, MAPE
[145]	2019	EEMD-WOA-SVM	<ul style="list-style-type: none"> <li>Uncertain and irregular nature of the data series can be reduced.</li> <li>Requires a few parameters for optimization.</li> </ul>	<ul style="list-style-type: none"> <li>Cannot predict the same accurate results for different datasets.</li> </ul>	Dataset from NSW and Queensland	Half-hourly	A hybrid model is proposed, which consists of data preprocessing, parameter optimization, and load forecasting.	BPNN, RBFNN, ARIMA, EMD-PSO-BPNN, EMD-CSO-WNN, EMD-WOA-SVM	MAE, MAPE, RMSE, WI, E <sub>NS</sub> , E <sub>LM</sub>

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[191]	2019	FA-SVM-SSSC	<ul style="list-style-type: none"> <li>A methodology is developed to access the season-specific meteorological variables.</li> <li>Seasonality effect is integrated into the forecasting process.</li> </ul>	<ul style="list-style-type: none"> <li>Should have considered the relationship between load data and meteorological variables.</li> </ul>	Dataset from SLDC, Assam	hourly	Seasonal variables are considered, and FA-SVM, which is season-specific, is proposed.		MAPE
[192]	2019	AMBA-WNN	<ul style="list-style-type: none"> <li>The convergence rate is higher.</li> <li>Adaptive.</li> </ul>	<ul style="list-style-type: none"> <li>Only theoretical and statistical representations are used.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from a city in China		AMBA overcomes the problems of slow convergence and trapping into local minima.	WNN, PSO-WNN, AMP-PSO-WNN	MAE, MAPE, RMSE
[193]	2019	H-EMD-SVRPSO	<ul style="list-style-type: none"> <li>Noise is reduced.</li> <li>Datasets can be filtered.</li> <li>Future tendencies can be forecasted.</li> </ul>	<ul style="list-style-type: none"> <li>When datasets exhibit mode mixing, EMD becomes ineffective.</li> </ul>	Dataset from NSW, Australia	Half-hourly	The data series can be filtered, and future tendencies can be forecasted by SVRPSO.	SVR, SVRPSO, PSO-BP, SVR-GA, EMD-SVR-AR, EMD-PSO-GA-SVR	MAE, MAPE, RMSE, R
[119]	2018	IEMD-ARIMA-WNN-FOA	<ul style="list-style-type: none"> <li>Linear and nonlinear characteristics can be extracted properly.</li> <li>Robust and stable.</li> </ul>	<ul style="list-style-type: none"> <li>Can not handle long data sequences.</li> </ul>	Dataset from AEMO and NYISO	Half-hourly	Fitting the nonlinear component into load data is done by WNN, which is optimized by FOA.	ENN, SVM, ELM, WTNNEA, WGMIPSO	MAPE, MAE, MPE, RMSE
[122]	2018	CEEMD-LSSVM-WOA	<ul style="list-style-type: none"> <li>Easier implementation.</li> <li>Multi-step forecasting at different time intervals is considered to give more future information.</li> </ul>	<ul style="list-style-type: none"> <li>Residual noise exists.</li> <li>Spurious artifact exists.</li> </ul>	Load data from NSW, Australia, Singapore	Half-hourly	Wind speed, electric load, and price are considered.	GRNN, BPNN, WOA-LSSVM, EMD-WOA-LSSVM	AE, MAE, MSE, MAPE, DA
[113]	2018	LSSVR-CQFOA	<ul style="list-style-type: none"> <li>Searching space becomes diverse.</li> <li>Appropriate parameters are selected effectively.</li> </ul>	<ul style="list-style-type: none"> <li>Seasonal factors are not considered.</li> <li>Computational burden is high.</li> <li>Results are reproduced.</li> </ul>	Dataset from IDAS 2014 [194] and GEFCOM 2014 [195]	Hourly	FOA can avoid local minima by implementing a chaotic global perturbation strategy	LSSVR-CQPSO, LSSVR-CQTS, LSSVR-CQGA, LSSVR-CQBA, LSSVR-FOA, LSSVR-QFOA	RMSE, MAE
[120]	2018	EEMD-ARIMA-CPSO	<ul style="list-style-type: none"> <li>Robust model.</li> <li>Convergence rate is high.</li> <li>Noise is reduced.</li> <li>Population diversity is improved.</li> </ul>	<ul style="list-style-type: none"> <li>Residual noise exists.</li> <li>The accuracy of the result is affected by adding white noise.</li> </ul>	Dataset from Shanxi, China	Half-hourly	Computational speed and prediction accuracy are improved by CPSO.	ARMA, ARIMA, EMD-ARIMA	RMSE, MAE, MAPE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[124]	2018	CEEMDAN-MGWO-SVM	<ul style="list-style-type: none"> <li>Robust model.</li> <li>Search space is enhanced.</li> <li>Residual noise can be effectively reduced.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> </ul>	Dataset from Hebei Province, China		The parameters of SVM are optimized by MGWO which will enhance the global search ability.	EEMD-MGWO-SVM, MGWO-SVM, GWO-SVM, SVM, BPNN	RE, MAPE, R <sup>22</sup>
[196]	2018	GA-FL and AC-FL	<ul style="list-style-type: none"> <li>Flexible working environment is provided.</li> <li>Forecasting conditions do not depend on training datasets.</li> </ul>	<ul style="list-style-type: none"> <li>During the low-temperature period, the optimization process becomes inactive.</li> </ul>	Dataset from National Load Dispatch Center	Hourly	GA-FL and AC-FL can deal with knowledge complexity.		MAPE
[126]	2018	GOA-SVM	<ul style="list-style-type: none"> <li>Can handle nonlinear datasets.</li> <li>Prediction accuracy is higher.</li> <li>Fewer parameters must be adjusted.</li> </ul>	<ul style="list-style-type: none"> <li>Season-specific factors are not considered.</li> <li>Computational burden is high.</li> <li>Cannot handle large datasets.</li> </ul>	Dataset from SLDC, Assam, India		Regional climate factors that impact the load data are considered here.	GA-SVM, PSO-SVM	MAPE
[197]	2017	ANN-GA, ANN-PSO, ANN-CSA, ANN-BA	<ul style="list-style-type: none"> <li>ANN-GA gives a better solution for fewer iterations.</li> <li>ANN-PSO performs better than GA as the position of particles is updated.</li> <li>ANN-CSA performs better than GA and PSO.</li> <li>ANN-BA shows the fastest convergence rate among these four.</li> </ul>	<ul style="list-style-type: none"> <li>Number of iterations is not large.</li> </ul>	Dataset from Xintai power plant, China	Hourly	ANN is trained by a back-propagation-based metaheuristic method.		Percentage of error
[131]	2017	SVR-CQGA	<ul style="list-style-type: none"> <li>Can work with nonlinear datasets.</li> <li>Population diversity is enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Can not handle long data series.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from Taiwan's regional electricity company [198] and GEFCOM 2014 [199].	hourly	Search space is enlarged by integrating cat function and quantum mechanics.	SVR-QGA, SVR-CQTS, SVR-QTS, SVRCQPSO, SVRQPSO	MAPE, MSE, RMSE, MAE
[117]	2017	EMD-PSO-GA-SVR	<ul style="list-style-type: none"> <li>Adaptive to different datasets.</li> <li>Volatility of SVR is reduced.</li> <li>Generalization capability is enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Is prone to mode mixing.</li> </ul>	Dataset from NYISO, USA, and NSW, Australia	hourly	The hybrid model shows a generalized capability in load forecasting while dealing with different types of data.	SVR, SVRPSO, SVR-GA, AFCM	MAPE, RMSE, MAE

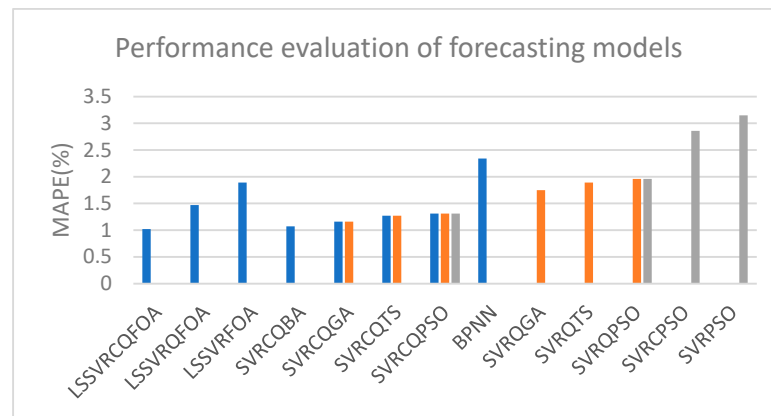
Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[127]	2017	CSA-SSA-SVM	<ul style="list-style-type: none"> <li>Nonlinear datasets are handled properly.</li> <li>Parameters are not selected artificially.</li> <li>Noise effects are eliminated.</li> </ul>	<ul style="list-style-type: none"> <li>Meteorological factors are not considered.</li> <li>Can not handle large data sequences.</li> </ul>	Dataset of NSW, Australia	Half-hourly, hourly	CS algorithm can train non-noisy datasets to construct an SVM model.	SVM, CS-SVM, SSA-SVM, SARIMA, BPNN	MAE, MSE, MAPE
[132]	2017	SDPSO-ELM	<ul style="list-style-type: none"> <li>Can avoid overtraining problems.</li> <li>Can avoid redundant nodes.</li> </ul>	<ul style="list-style-type: none"> <li>Computational burden is high.</li> <li>Meteorological factors are not considered.</li> </ul>	Dataset from Fujian province, China		The proposed method can avoid overtraining problems and unnecessary nodes.	RBFFNN	MAPE, MAE
[130]	2016	SVRCQPSO	<ul style="list-style-type: none"> <li>Search space is enhanced by applying quantum mechanics.</li> <li>Generalization capability is enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Prediction accuracy is not guaranteed for new datasets.</li> <li>Premature convergence is not fully avoided.</li> </ul>	Dataset from four regions of Taiwan [198], GEFCOM 2014 [199].	Hourly	A hybrid model is proposed with a chaotic mapping function and quantum metaheuristic algorithm.	ARIMA, BPNN, SVRPSO, SVRCPSO, SVRQPSO, SVRCGA	MAPE
[129]	2016	DEMD-QPSO-SVR-AR	<ul style="list-style-type: none"> <li>Instability impact is solved.</li> <li>Satisfactory parameter solutions are achieved.</li> </ul>	<ul style="list-style-type: none"> <li>Low iteration number is used.</li> <li>Computation burden is high.</li> </ul>	Dataset from NSW, Australia, and NYISO, USA	Half-hourly	To optimize the parameters of SVR, QPSO is used.	ARIMA, BP-ANN, GA-ANN, EMD-SVR-AR,	MAE, RMSE, MAPE
[116]	2016	EMD-SVRPSO	<ul style="list-style-type: none"> <li>Computational complexity does not depend on input variables.</li> <li>Generalization capability is enhanced.</li> </ul>	<ul style="list-style-type: none"> <li>Cannot handle large datasets.</li> <li>Prone to mode mixing.</li> <li>Sensitive to noise.</li> </ul>	Dataset from SGHEPC, China		A hybrid model is proposed for the residential dataset.	EMD-SVR, PSO-SVR, SVR	RMSE, MAE, MAPE, VAPE
[112]	2016	SVR-CQTS	<ul style="list-style-type: none"> <li>Search space is enhanced.</li> <li>Easy to implement.</li> <li>Population diversity is improved.</li> </ul>	<ul style="list-style-type: none"> <li>Cannot handle large datasets.</li> <li>Sensitive to noise.</li> </ul>	Dataset from Taiwan's regional electricity company [130,198].		To improve the forecasting accuracy, quantum mechanics is applied with Tabu Search to enhance the tabu memory.	SVR-QTS, SVRCQPSO, SVRQPSO, SVR-CPSO, SVRPSO	MAPE
[111]	2015	SSVR-CCSA	<ul style="list-style-type: none"> <li>Discrete nature of the temperature annealing process is avoided.</li> <li>Seasonal mechanism is used to align with the electric load.</li> </ul>	<ul style="list-style-type: none"> <li>Cannot handle large datasets.</li> <li>Sensitive to noise.</li> <li>Depends on the initial value.</li> </ul>	Dataset from Northeast China and NYISO, USA [200]	monthly	Premature convergence can be avoided by cat mapping function and cyclic effects can be adjusted through seasonal mechanism.	ARIMA, SSVR-SA, BPNN, SVR-SA, SVR-CSA	MAPE

Table 7. Cont.

Reference	Year	Algorithm	Advantages	Limitations	Test Set	Input Resolution	Contribution	Compared Methods	Evaluation Indicators
[123]	2014	PSO-SVM	<ul style="list-style-type: none"> <li>Switching events are detected when required data are available.</li> <li>Significant changes in temperature are identified properly.</li> </ul>	<ul style="list-style-type: none"> <li>Tends to trap local optima.</li> <li>Low convergence rate.</li> </ul>	Dataset from Burbank Utility, USA	Hourly	Temperature sensitivity is considered here.	Classical method	MAPE
[115]	2014	EMD-EKF-KELM-PSO	<ul style="list-style-type: none"> <li>Can handle nonlinear and nonstationary datasets.</li> <li>Meteorological factors are considered.</li> </ul>	<ul style="list-style-type: none"> <li>Can not handle long datasets.</li> <li>Sensitivity to noise.</li> </ul>	Residential and commercial load data of Zhejiang Province, China	Hourly	A hybrid method with parameter optimization is proposed by designing offline optimization and online forecasting.	KELM	MAPE
[105]	2013	SSVR-CGSA	<ul style="list-style-type: none"> <li>Seasonal mechanism is added.</li> <li>Current best solution is refined.</li> </ul>	<ul style="list-style-type: none"> <li>Can not handle long datasets.</li> <li>Prone to noise effect.</li> </ul>	Dataset from Northeast China [201].	Monthly	The proposed methods can handle non-historical climate change datasets.	ARIMA, SVR-CGA	MAPE





**Figure 8.** Performance evaluation of forecasting models (blue—[113], orange—[131], grey—[130]).

## 11. Review Findings and Recommendations

The article considers the growing interest of researchers in the field of load forecasting combined with metaheuristic algorithms. After a careful review of the existing literature, the following findings are discussed here:

- The accuracy and reliability of a forecasting method depend on the input data series. Accessing datasets is a challenging task. Some of the studies have used historical datasets such as from AEMO, ENTSO-E, and NYISO, and some have used real datasets from the power grid. However, with the growth of penetration of renewable energy sources, the datasets are changing, and the developed model should be able to forecast these changes. To increase the robustness of STLF models, smart meter or social media data source provides an opportunity for new data sources. Machine-learning algorithms can be applied to identify the relevant data sources to develop forecasting models in the future.
- It has been found that the combined framework predicts results more accurately than individual ones. This is due to the hyperparameters tuning. Therefore, hyperparameters play a crucial role in increasing prediction accuracy. However, there is a lacking of exploring different metaheuristic algorithms for this parameter tuning. More algorithms should be explored, such as fruit-fly and flower pollination, differential evolution, etc., to check their efficacy in tuning hyperparameters correctly.
- The developed model should be universal as well as adaptive such as it should work on any datasets that are given as inputs. These models should adapt to any changes in the energy system conditions. However, it is difficult to assess the validity of a model based on a particular dataset. A comparison should be made with the other state-of-the-art methods on a particular dataset.
- Evaluation indices such as MAE, MAPE, MSE, and RMSE can be used to evaluate the prediction accuracy of the models.
- Meteorological factors such as temperature, wind speed, and humidity play an important part in the load dataset. Most of the literature ignores the weather information. Only a handful of them did weather forecasting. If weather information from the different meteorological bureaus can be found, then it would be useful to incorporate that information into weather data.
- Single predictive models are prone to premature convergence and trapping to local solutions. It is found in the literature that a hybrid or combined model can overcome these drawbacks. In addition to that, these hybrid models have shown improved performance efficiency and accuracy. Further research can focus on improving these hybrid models to incorporate more input features and other challenges faced by STLF.
- As mentioned earlier, most of the works are based on historical datasets. Very little work is carried out at the low-voltage distribution network level. In this case, the

forecasting model will have to deal with very volatile datasets. Again, data privacy could be an issue as well.

- Another problem with the dataset is the quality issue, which involves missing data, measurement problems, etc. Prediction accuracy is impacted by data quality issues. Further studies can assist in developing models that can deal with missing data or measurement errors.
- The load data series can be of different time scales such as daily, weekly, etc.; these also include seasonal patterns. Future research should focus on developing models that can train datasets on multiple time scales.
- Artificial intelligence-based methods have been widely used in the existing literature rather than statistical models because of their data processing and feature extraction capabilities. Further studies can investigate advanced-level deep-learning methods to handle a large number of datasets that contain different input features.
- Load data contains nonlinear and nonstationary data series. Statistical models cannot handle the nonlinear and nonstationary behavior of these datasets. As a result, they assume that the input data are linear and stationary. Future research can explore the development of more flexible statistical models that can handle the nonlinearity of data.
- Real-time load forecasting becomes a necessity to increase the accuracy and timeliness of a forecasting model. Online learning algorithms can create a path for real-time data acquisition when it becomes available and adjust the forecasting result accordingly. Further studies can focus on developing learning algorithms that can work online.
- Most of the forecasting models act like black boxes which means they are difficult to interpret. However, interpretability is an important factor that can help the power industry to have an insight into the elements that drive the prediction and make decisions about resource allocation. Therefore, future research can explore the algorithms that are easier to interpret and understand.
- Electricity demand depends on consumer behavior, environment, etc. which can influence the prediction results. Therefore, it is important to incorporate this knowledge information into the forecasting model to improve the accuracy and diversity of the results. Further research can be undertaken to investigate the sophisticated models to integrate this information as input features.
- The Particle Swarm Optimization algorithm has gained popularity among all the other metaheuristic algorithms. The recent advancement in other algorithms should be explored, such as differential algorithms and generalized normal distribution optimization algorithms to have a fast convergence rate.

## 12. Conclusions

To improve the accuracy of the forecasting techniques, researchers have worked on numerous single and hybrid predictive models. This article has presented a comprehensive survey of the existing state-of-the-art methods used for short-term load forecasting, which include metaheuristic algorithms. The hybrid models are found to be more efficient than the single models. The metaheuristic algorithms optimize the parameters of the single models to minimize the prediction error percentage. An analysis of different data-decomposition methods and deep-learning methods is summarized. The hyperparameters of the deep-learning models are identified and their interdependencies have been discussed. Furthermore, a quantitative analysis is presented on the most commonly used metaheuristic algorithms, which include their advantages, disadvantages, and solving capability of a problem. Every algorithm has its own advantages and limitations based on accuracy, speed, efficiency, etc. Their applications depend on the rate of speed, data size, and parameter setting. However, the Particle Swarm Optimization algorithm is found to be widely used in existing literature. Genetic algorithms are also seen to be used in some articles. A huge gap is still prevalent in the existing literature on the usage of recently used metaheuristic algorithms in other applications. Advanced algorithms, such as differential evolution and generalized normal distribution optimization algorithms, can be used for future work.

Again, the developed model should be universal to cope with any dataset. Most of the previous works are based on historical datasets. If researchers can use real-world data to check the validity of a forecasting model, it would be more valuable for the power industry.

**Author Contributions:** Methodology, U.M.; Formal analysis, U.M.; Writing—original draft, U.M.; Writing—review & editing, S.A. and P.W.; Supervision, S.A. and P.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Advanced Queensland Industry Research Fellowship program and grant number AQIRF105-2022RD5.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

SSA	Singular spectrum analysis	ENTSO-E	European Network of Transmission System Operators for Electricity
VSTLF	Very short-term load forecasting	ABC	Artificial bee colony
STLF	Short-term load forecasting	RFR	Random forest regression
MTLF	Medium-term load forecasting	GBRBM	Gauss–Bernoulli restricted Boltzmann’s machine
LTLF	Long-term load forecasting	PM	Persistence model
LSTM	Long short-term memory	FCRBM	Factored conditional restricted Boltzmann machine
AEMO	Australian energy market operator	GWDO	Genetic wind-driven optimization
MSE	Mean square error	MI-ANN	Mutual information-based artificial neural network
RMSE	Root mean square error	AFC-ANN	ANN-based accurate and fast converging
MAE	Mean absolute error	mRMR	Minimal redundancy maximal relevance
APE	Absolute percentage error	ISO-NE	Independent System Operator New England
MAPE	Mean absolute percentage error	GS	Grid Search
NRMSE	Normalized root mean square error	RNN	Ridgelet neural network
SVM	Support vector machines	ENN	Elman neural network
GRU	Gated recurrent units	MHNN	Modified hybrid neural network
CNN	Convolutional neural networks	BFA	Bacterial foraging algorithm
ELM	Ensemble learning machine	GSA	Gravitational search algorithm
Bi-LSTM	bidirectional LSTM	DA	Direction accuracy
MLP	Multilayer perceptron	STD	Standard deviation
SVR	Support vector regression	LSSVR	Least-squares support vector regression
MRFO	Manta ray foraging optimization	CEEMDAN	Complete ensemble empirical mode decomposition adaptive noise
LR	Linear Regression	SVRCQPSO	support vector regression with chaotic quantum particle swarm optimization
SCG	Scaled Conjugated Gradient	DEMD	Differential empirical mode decomposition
AE	Autoencoder	NYISO	New York Independent System Operator
R	Correlation coefficient	GEFCOM	Global Energy Forecasting Competition
GWO	Gray wolf optimization	SGHEPC	State Grid Handan Electric Power Company
ANN	Artificial neural network	WOA	Whale optimization algorithm
FPA	Flower Pollination	FTS	Fuzzy time series
RF	Random Forest	FOA	Fruit-fly optimization algorithm
MGF	Mean generating function	IDAS	Island data acquisition system
RSM	Response surface method	TS	Tabu search
MAD	Mean absolute deviation	ANYISO	American New York Independent System Operator
BPNN	Back-propagation neural network	CSA	Cuckoo search algorithm
EEMD	Ensemble empirical mode decomposition	SARIMA	Seasonal autoregressive integrated moving average
CV	Coefficient of variation	DBN	Deep belief network
VMD	Variational mode decomposition	LSSVM	Nonlinear least square support vector machine
WNN	Wavelet neural network	GRNN	Generalized regression neural network
GA	Genetic algorithm	PSO	Particle swarm optimization
SCA	Sine cosine algorithm		

## References

1. Lin, Y.; Luo, H.; Wang, D.; Guo, H.; Zhu, K. An ensemble model based on machine learning methods and data preprocessing for short-term electric load forecasting. *Energies* **2017**, *10*, 1186. [\[CrossRef\]](#)
2. Nalcaci, G.; Özmen, A.; Weber, G.W. Long-term load forecasting: Models based on MARS, ANN and LR methods. *Cent. Eur. J. Oper. Res.* **2019**, *27*, 1033–1049. [\[CrossRef\]](#)
3. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [\[CrossRef\]](#)
4. Badr, M.M.; Ibrahim, M.I.; Mahmoud, M.; Alasmay, W.; Fouda, M.M.; Almotairi, K.H.; Fadlullah, Z.M. Privacy-preserving federated-learning-based net-energy forecasting. In *SoutheastCon 2022*; IEEE: New York, NY, USA, 2022; pp. 133–139.
5. Xia, M.; Shao, H.; Ma, X.; de Silva, C.W. A stacked GRU-RNN-based approach for predicting renewable energy and electricity load for smart grid operation. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7050–7059. [\[CrossRef\]](#)
6. Sevlian, R.; Rajagopal, R. A scaling law for short term load forecasting on varying levels of aggregation. *Int. J. Electr. Power Energy Syst.* **2018**, *98*, 350–361. [\[CrossRef\]](#)
7. 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. [\[CrossRef\]](#)
8. Khuntia, S.R.; Rueda, J.L.; van Der Meijden, M.A. Forecasting the load of electrical power systems in mid-and long-term horizons: A review. *IET Gener. Transm. Distrib.* **2016**, *10*, 3971–3977. [\[CrossRef\]](#)
9. Koprinska, I.; Rana, M.; Agelidis, V.G. Correlation and instance based feature selection for electricity load forecasting. *Knowl. Based Syst.* **2015**, *82*, 29–40. [\[CrossRef\]](#)
10. Jiang, P.; Liu, F.; Song, Y. A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. *Energy* **2017**, *119*, 694–709. [\[CrossRef\]](#)
11. Groß, A.; Lenders, A.; Schwenker, F.; Braun, D.A.; Fischer, D. Comparison of short-term electrical load forecasting methods for different building types. *Energy Inform.* **2021**, *4*, 1–16. [\[CrossRef\]](#)
12. Kim, J.; Moon, J.; Hwang, E.; Kang, P. Recurrent inception convolution neural network for multi short-term load forecasting. *Energy Build.* **2019**, *194*, 328–341. [\[CrossRef\]](#)
13. Kathirgamanathan, A.; Patel, A.; Khwaja, A.S.; Venkatesh, B.; Anpalagan, A. Performance comparison of single and ensemble CNN, LSTM and traditional ANN models for short-term electricity load forecasting. *J. Eng.* **2022**, *2022*, 550–565. [\[CrossRef\]](#)
14. Liao, Z.; Pan, H.; Huang, X.; Mo, R.; Fan, X.; Chen, H.; Liu, L.; Li, Y. Short-term load forecasting with dense average network. *Expert Syst. Appl.* **2021**, *186*, 115748. [\[CrossRef\]](#)
15. Matrenin, P.; Safaraliev, M.; Dmitriev, S.; Kokin, S.; Ghulomzoda, A.; Mitrofanov, S. Medium-term load forecasting in isolated power systems based on ensemble machine learning models. *Energy Rep.* **2022**, *8*, 612–618. [\[CrossRef\]](#)
16. Hammad, M.A.; Jereb, B.; Rosi, B.; Dragan, D. Methods and models for electric load forecasting: A comprehensive review. *Logist. Sustain. Transp* **2020**, *11*, 51–76. [\[CrossRef\]](#)
17. Lindberg, K.; Seljom, P.; Madsen, H.; Fischer, D.; Korpås, M. Long-term electricity load forecasting: Current and future trends. *Util. Policy* **2019**, *58*, 102–119. [\[CrossRef\]](#)
18. Habbak, H.; Mahmoud, M.; Metwally, K.; Fouda, M.M.; Ibrahim, M.I. Load forecasting techniques and their applications in smart grids. *Energies* **2023**, *16*, 1480. [\[CrossRef\]](#)
19. Huang, Y.; Hasan, N.; Deng, C.; Bao, Y. Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting. *Energy* **2022**, *239*, 122245. [\[CrossRef\]](#)
20. Zhou, M.; Jin, M. Holographic ensemble forecasting method for short-term power load. *IEEE Trans. Smart Grid* **2017**, *10*, 425–434. [\[CrossRef\]](#)
21. Azeem, A.; Ismail, I.; Jameel, S.M.; Harindran, V.R. Electrical load forecasting models for different generation modalities: A review. *IEEE Access* **2021**, *9*, 142239–142263. [\[CrossRef\]](#)
22. Kuster, C.; Rezgui, Y.; Mourshed, M. Electrical load forecasting models: A critical systematic review. *Sustain. Cities Soc.* **2017**, *35*, 257–270. [\[CrossRef\]](#)
23. Hou, H.; Liu, C.; Wang, Q.; Wu, X.; Tang, J.; Shi, Y.; Xie, C. Review of load forecasting based on artificial intelligence methodologies, models, and challenges. *Electr. Power Syst. Res.* **2022**, *210*, 108067. [\[CrossRef\]](#)
24. Akhtaruzzaman, M.; Hasan, M.K.; Kabir, S.R.; Abdullah, S.N.H.S.; Sadeq, M.J.; Hossain, E. HSIC bottleneck based distributed deep learning model for load forecasting in smart grid with a comprehensive survey. *IEEE Access* **2020**, *8*, 222977–223008. [\[CrossRef\]](#)
25. Al Mamun, A.; Sohel, M.; Mohammad, N.; Sunny, M.S.H.; Dipta, D.R.; Hossain, E. A comprehensive review of the load forecasting techniques using single and hybrid predictive models. *IEEE Access* **2020**, *8*, 134911–134939. [\[CrossRef\]](#)
26. Hiba, C.; Tarek, K.M.; Belkacem, C. Deep Neural Network Architectures for Electrical Load Forecasting: A Review. *Facilities* **2022**, *2*, 3.
27. Kondaiah, V.; Saravanan, B.; Sanjeevikumar, P.; Khan, B. A review on short-term load forecasting models for micro-grid application. *J. Eng.* **2022**, *2022*, 665–689. [\[CrossRef\]](#)
28. Aslam, S.; Herodotou, H.; Mohsin, S.M.; Javaid, N.; Ashraf, N.; Aslam, S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew. Sustain. Energy Rev.* **2021**, *144*, 110992. [\[CrossRef\]](#)

29. Haben, S.; Arora, S.; Giasemidis, G.; Voss, M.; Greetham, D.V. Review of low voltage load forecasting: Methods, applications, and recommendations. *Appl. Energy* **2021**, *304*, 117798. [[CrossRef](#)]
30. Ma, P.; Cui, S.; Chen, M.; Zhou, S.; Wang, K. Review of family-level short-term load forecasting and its application in household energy management system. *Energies* **2023**, *16*, 5809. [[CrossRef](#)]
31. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
32. Tan, M.; Chen, B.; Pang, R.; Vasudevan, V.; Sandler, M.; Howard, A.; Le, Q.V. Mnasnet: Platform-aware neural architecture search for mobile. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2820–2828.
33. Fida, K.; Abbasi, U.; Adnan, M.; Iqbal, S.; Gasim Mohamed, S.E. A comprehensive survey on load forecasting hybrid models: Navigating the Futuristic demand response patterns through experts and intelligent systems. *Results Eng.* **2024**, *23*, 102773. [[CrossRef](#)]
34. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*.
35. Tranfield, D.; Denyer, D.; Smart, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br. J. Manag.* **2003**, *14*, 207–222. [[CrossRef](#)]
36. Khatoon, S.; Ibraheem; Singh, A.K.; Priti. Effects of various factors on electric load forecasting: An overview. In Proceedings of the 2014 6th IEEE Power India International Conference (PIICON), Delhi, India, 5–7 December 2014; pp. 1–5.
37. Black, J.D.; Henson, W.L.W. Hierarchical Load Hindcasting Using Reanalysis Weather. *IEEE Trans. Smart Grid* **2014**, *5*, 447–455. [[CrossRef](#)]
38. Patel, R.; Patel, M.R.; Patel, R.V. A review: Introduction and understanding of load forecasting. *J. Appl. Sci. Comput* **2019**, *4*, 1449–1457.
39. Zheng, G.; Chen, S.; Fan, S. A power load forecasting method based on matching coefficient of meteorological factors and similar load modification. In Proceedings of the 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics, Nanchang, China, 26–27 August 2012; pp. 216–219.
40. Nagasaka, K.; Al Mamun, M. Long-term peak demand prediction of 9 Japanese power utilities using radial basis function networks. In Proceedings of the IEEE Power Engineering Society General Meeting, Denver, CO, USA, 6–10 June 2004; pp. 315–322.
41. Fahad, M.U.; Arbab, N. Factor affecting short term load forecasting. *J. Clean Energy Technol.* **2014**, *2*, 305–309. [[CrossRef](#)]
42. He, F.; Zhou, J.; Feng, Z.-k.; Liu, G.; Yang, Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Appl. Energy* **2019**, *237*, 103–116. [[CrossRef](#)]
43. Hyde, O.; Hodnett, P. An adaptable automated procedure for short-term electricity load forecasting. *IEEE Trans. Power Syst.* **1997**, *12*, 84–94. [[CrossRef](#)]
44. Papalexopoulos, A.D.; Hesterberg, T.C. A regression-based approach to short-term system load forecasting. *IEEE Trans. Power Syst.* **1990**, *5*, 1535–1547. [[CrossRef](#)]
45. Al-Hamadi, H.; Soliman, S. Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model. *Electr. Power Syst. Res.* **2004**, *68*, 47–59. [[CrossRef](#)]
46. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
47. Dai, Y.; Zhao, P. A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. *Appl. Energy* **2020**, *279*, 115332. [[CrossRef](#)]
48. Khairalla, M. Meta-Heuristic Search Optimization and its application to Time Series Forecasting Model. *Intell. Syst. Appl.* **2022**, *16*, 200142. [[CrossRef](#)]
49. Ghaemi, Z.; Tran, T.T.D.; Smith, A.D. Comparing classical and metaheuristic methods to optimize multi-objective operation planning of district energy systems considering uncertainties. *Appl. Energy* **2022**, *321*, 119400. [[CrossRef](#)]
50. Silveira, C.L.B.; Tabares, A.; Faria, L.T.; Franco, J.F. Mathematical optimization versus Metaheuristic techniques: A performance comparison for reconfiguration of distribution systems. *Electr. Power Syst. Res.* **2021**, *196*, 107272. [[CrossRef](#)]
51. Yu, T.; Zhu, H. Hyper-parameter optimization: A review of algorithms and applications. *arXiv* **2020**, arXiv:2003.05689.
52. Xiao, L.; Shao, W.; Liang, T.; Wang, C. A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting. *Appl. Energy* **2016**, *167*, 135–153. [[CrossRef](#)]
53. Xiao, L.; Shao, W.; Yu, M.; Ma, J.; Jin, C. Research and application of a combined model based on multi-objective optimization for electrical load forecasting. *Energy* **2017**, *119*, 1057–1074. [[CrossRef](#)]
54. Liang, Y.; Niu, D.; Ye, M.; Hong, W.-C. Short-Term Load Forecasting Based on Wavelet Transform and Least Squares Support Vector Machine Optimized by Improved Cuckoo Search. *Energies* **2016**, *9*, 827. [[CrossRef](#)]
55. Yang, Y.; Che, J.; Li, Y.; Zhao, Y.; Zhu, S. An incremental electric load forecasting model based on support vector regression. *Energy* **2016**, *113*, 796–808. [[CrossRef](#)]



56. Fan, G.-F.; Peng, L.-L.; Hong, W.-C.; Sun, F. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing* **2016**, *173*, 958–970. [[CrossRef](#)]
57. Li, C.; Li, S.; Liu, Y. A least squares support vector machine model optimized by moth-flame optimization algorithm for annual power load forecasting. *Appl. Intell.* **2016**, *45*, 1166–1178. [[CrossRef](#)]
58. Kaur, A.; Nonnenmacher, L.; Coimbra, C.F.M. Net load forecasting for high renewable energy penetration grids. *Energy* **2016**, *114*, 1073–1084. [[CrossRef](#)]
59. Koukaras, P.; Bezas, N.; Gkaidatzis, P.; Ioannidis, D.; Tzovaras, D.; Tjortjis, C. Introducing a novel approach in one-step ahead energy load forecasting. *Sustain. Comput. Inform. Syst.* **2021**, *32*, 100616. [[CrossRef](#)]
60. Bianchi, F.M.; Santis, E.D.; Rizzi, A.; Sadeghian, A. Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. *IEEE Access* **2015**, *3*, 1931–1943. [[CrossRef](#)]
61. Niu, D.; Dai, S. A Short-Term Load Forecasting Model with a Modified Particle Swarm Optimization Algorithm and Least Squares Support Vector Machine Based on the Denoising Method of Empirical Mode Decomposition and Grey Relational Analysis. *Energies* **2017**, *10*, 408. [[CrossRef](#)]
62. Chen, Y.; Hu, X.; Zhang, L. A review of ultra-short-term forecasting of wind power based on data decomposition-forecasting technology combination model. *Energy Rep.* **2022**, *8*, 14200–14219. [[CrossRef](#)]
63. Lu, P.; Ye, L.; Zhao, Y.; Dai, B.; Pei, M.; Tang, Y. Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges. *Appl. Energy* **2021**, *301*, 117446. [[CrossRef](#)]
64. Bahrami, S.; Hooshmand, R.-A.; Parastegari, M. Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. *Energy* **2014**, *72*, 434–442. [[CrossRef](#)]
65. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
66. An, X.; Jiang, D.; Zhao, M.; Liu, C. Short-term prediction of wind power using EMD and chaotic theory. *Commun. Nonlinear Sci. Numer. Simul.* **2012**, *17*, 1036–1042. [[CrossRef](#)]
67. Wang, Y.; Guo, P.; Ma, N.; Liu, G. Robust Wavelet Transform Neural-Network-Based Short-Term Load Forecasting for Power Distribution Networks. *Sustainability* **2023**, *15*, 296. [[CrossRef](#)]
68. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
69. Yuan, J.; Wang, L.; Qiu, Y.; Wang, J.; Zhang, H.; Liao, Y. Short-term electric load forecasting based on improved Extreme Learning Machine Mode. *Energy Rep.* **2021**, *7*, 1563–1573. [[CrossRef](#)]
70. Yeh, J.-R.; Shieh, J.-S.; Huang, N.E. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2010**, *2*, 135–156. [[CrossRef](#)]
71. Wang, D.; Yue, C.; ElAmraoui, A. Multi-step-ahead electricity load forecasting using a novel hybrid architecture with decomposition-based error correction strategy. *Chaos Solitons Fractals* **2021**, *152*, 111453. [[CrossRef](#)]
72. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 4144–4147.
73. Hong, Y.; Wang, D.; Su, J.; Ren, M.; Xu, W.; Wei, Y.; Yang, Z. Short-Term Power Load Forecasting in Three Stages Based on CEEMDAN-TGA Model. *Sustainability* **2023**, *15*, 11123. [[CrossRef](#)]
74. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
75. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A.A. *Analysis of Time Series Structure: SSA and Related Techniques*; CRC press: Boca Raton, FL, USA, 2001.
76. Stratigakos, A.; Bachoumis, A.; Vita, V.; Zafiroopoulos, E. Short-Term Net Load Forecasting with Singular Spectrum Analysis and LSTM Neural Networks. *Energies* **2021**, *14*, 4107. [[CrossRef](#)]
77. Hassani, H.; Thomakos, D. A review on singular spectrum analysis for economic and financial time series. *Stat. Its Interface* **2010**, *3*, 377–397. [[CrossRef](#)]
78. Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A review of machine learning in building load prediction. *Appl. Energy* **2021**, *285*, 116452. [[CrossRef](#)]
79. Gonsalves, T.; Upadhyay, J. Chapter Eight—Integrated deep learning for self-driving robotic cars. In *Artificial Intelligence for Future Generation Robotics*; Shaw, R.N., Ghosh, A., Balas, V.E., Bianchini, M., Eds.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 93–118. [[CrossRef](#)]
80. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
81. Alvarez, J.M.; Salzmann, M. Learning the number of neurons in deep networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2270–2278.
82. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
83. Mercioni, M.A.; Holban, S. P-swish: Activation function with learnable parameters based on swish activation function in deep learning. In Proceedings of the 2020 International Symposium on Electronics and Telecommunications, Timisoara, Romania, 5–6 November 2020; IEEE: New York, NY, USA; pp. 1–4.

84. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaria, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)]
85. Shin, H.-C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)] [[PubMed](#)]
86. Guo, C.; Liu, Y.-l.; Jiao, X. Study on the influence of variable stride scale change on image recognition in CNN. *Multimed. Tools Appl.* **2019**, *78*, 30027–30037. [[CrossRef](#)]
87. Akhtar, N.; Ragavendran, U. Interpretation of intelligence in CNN-pooling processes: A methodological survey. *Neural Comput. Appl.* **2020**, *32*, 879–898. [[CrossRef](#)]
88. Poernomo, A.; Kang, D.-K. Biased dropout and crossmap dropout: Learning towards effective dropout regularization in convolutional neural network. *Neural Netw.* **2018**, *104*, 60–67. [[CrossRef](#)]
89. Zhang, Y.; Niu, J.; Na, S. A novel nonlinear function fitting model based on FOA and GRNN. *Math. Probl. Eng.* **2019**, *2019*, 2697317. [[CrossRef](#)]
90. Karmakar, S.; Shrivastava, G.; Kowar, M.K. Impact of learning rate and momentum factor in the performance of back-propagation neural network to identify internal dynamics of chaotic motion. *Kuwait J. Sci.* **2014**, *41*, 151–174.
91. Yildirim, H.; Özkale, M.R. The performance of ELM based ridge regression via the regularization parameters. *Expert Syst. Appl.* **2019**, *134*, 225–233. [[CrossRef](#)]
92. Mirjalili, S. Evolutionary algorithms and neural networks. *Stud. Comput. Intell.* **2019**, *780*, 43–53.
93. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; MIT press: Cambridge, MA, USA, 1992.
94. Saremi, S.; Mirjalili, S.; Lewis, A. Grasshopper optimisation algorithm: Theory and application. *Adv. Eng. Softw.* **2017**, *105*, 30–47. [[CrossRef](#)]
95. Nayak, P.C.; Prusty, R.C.; Panda, S. Grasshopper optimization algorithm optimized multistage controller for automatic generation control of a power system with FACTS devices. *Prot. Control Mod. Power Syst.* **2021**, *6*, 8. [[CrossRef](#)]
96. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [[CrossRef](#)]
97. Karaboga, D.; Basturk, B. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *J. Glob. Optim.* **2007**, *39*, 459–471. [[CrossRef](#)]
98. Yang, X.-S. A new metaheuristic bat-inspired algorithm. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 65–74.
99. Yang, X.-S.; Deb, S. Cuckoo search via Lévy flights. In *Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 9–11 December 2009*; pp. 210–214.
100. Kennedy, J.; Eberhart, R. Particle swarm optimization. In *Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995*; pp. 1942–1948.
101. Rashedi, E.; Nezamabadi-Pour, H.; Saryazdi, S. GSA: A gravitational search algorithm. *Inf. Sci.* **2009**, *179*, 2232–2248. [[CrossRef](#)]
102. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [[CrossRef](#)]
103. Geem, Z.W.; Kim, J.H.; Loganathan, G.V. A new heuristic optimization algorithm: Harmony search. *Simulation* **2001**, *76*, 60–68. [[CrossRef](#)]
104. Xing, B.; Gao, W.-J.; Xing, B.; Gao, W.-J. Fruit fly optimization algorithm. *Innov. Comput. Intell. A Rough Guide 134 Clever Algorithms* **2014**, *62*, 167–170.
105. Ju, F.-Y.; Hong, W.-C. Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting. *Appl. Math. Model.* **2013**, *37*, 9643–9651. [[CrossRef](#)]
106. Ceperic, E.; Ceperic, V.; Baric, A. A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines. *IEEE Trans. Power Syst.* **2013**, *28*, 4356–4364. [[CrossRef](#)]
107. Reis, A.R.; Da Silva, A.A. Feature extraction via multiresolution analysis for short-term load forecasting. *IEEE Trans. Power Syst.* **2005**, *20*, 189–198.
108. Amjady, N.; Keynia, F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy* **2009**, *34*, 46–57. [[CrossRef](#)]
109. Deihimi, A.; Showkati, H. Application of echo state networks in short-term electric load forecasting. *Energy* **2012**, *39*, 327–340. [[CrossRef](#)]
110. Hu, Z.; Bao, Y.; Xiong, T. Comprehensive learning particle swarm optimization based memetic algorithm for model selection in short-term load forecasting using support vector regression. *Appl. Soft Comput.* **2014**, *25*, 15–25. [[CrossRef](#)]
111. Geng, J.; Huang, M.-L.; Li, M.-W.; Hong, W.-C. Hybridization of seasonal chaotic cloud simulated annealing algorithm in a SVR-based load forecasting model. *Neurocomputing* **2015**, *151*, 1362–1373. [[CrossRef](#)]
112. Lee, C.-W.; Lin, B.-Y. Application of Hybrid Quantum Tabu Search with Support Vector Regression (SVR) for Load Forecasting. *Energies* **2016**, *9*, 873. [[CrossRef](#)]
113. Li, M.-W.; Geng, J.; Hong, W.-C.; Zhang, Y. Hybridizing chaotic and quantum mechanisms and fruit fly optimization algorithm with least squares support vector regression model in electric load forecasting. *Energies* **2018**, *11*, 2226. [[CrossRef](#)]



114. Kavousi-Fard, A.; Samet, H.; Marzbani, F. A new hybrid Modified Firefly Algorithm and Support Vector Regression model for accurate Short Term Load Forecasting. *Expert Syst. Appl.* **2014**, *41*, 6047–6056. [[CrossRef](#)]
115. Liu, N.; Tang, Q.; Zhang, J.; Fan, W.; Liu, J. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Appl. Energy* **2014**, *129*, 336–345. [[CrossRef](#)]
116. Wang, X.; Wang, Y. A hybrid model of EMD and PSO-SVR for short-term load forecasting in residential quarters. *Math. Probl. Eng.* **2016**, *2016*, 9895639. [[CrossRef](#)]
117. Fan, G.-F.; Peng, L.-L.; Zhao, X.; Hong, W.-C. Applications of Hybrid EMD with PSO and GA for an SVR-Based Load Forecasting Model. *Energies* **2017**, *10*, 1713. [[CrossRef](#)]
118. Chen, Y.; Yang, Y.; Liu, C.; Li, C.; Li, L. A hybrid application algorithm based on the support vector machine and artificial intelligence: An example of electric load forecasting. *Appl. Math. Model.* **2015**, *39*, 2617–2632. [[CrossRef](#)]
119. Zhang, J.; Wei, Y.-M.; Li, D.; Tan, Z.; Zhou, J. Short term electricity load forecasting using a hybrid model. *Energy* **2018**, *158*, 774–781. [[CrossRef](#)]
120. Li, W.-Q.; Chang, L. A combination model with variable weight optimization for short-term electrical load forecasting. *Energy* **2018**, *164*, 575–593. [[CrossRef](#)]
121. Liu, Z.; Sun, W.; Zeng, J. A new short-term load forecasting method of power system based on EEMD and SS-PSO. *Neural Comput. Appl.* **2014**, *24*, 973–983. [[CrossRef](#)]
122. Du, P.; Wang, J.; Yang, W.; Niu, T. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. *Renew. Energy* **2018**, *122*, 533–550. [[CrossRef](#)]
123. Selakov, A.; Cvijetinović, D.; Milović, L.; Mellon, S.; Bekut, D. Hybrid PSO–SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank. *Appl. Soft Comput.* **2014**, *16*, 80–88. [[CrossRef](#)]
124. Dai, S.; Niu, D.; Li, Y. Daily Peak Load Forecasting Based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Support Vector Machine Optimized by Modified Grey Wolf Optimization Algorithm. *Energies* **2018**, *11*, 163. [[CrossRef](#)]
125. Wang, J.; Yang, W.; Du, P.; Li, Y. Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. *Energy* **2018**, *148*, 59–78. [[CrossRef](#)]
126. Barman, M.; Choudhury, N.D.; Sutradhar, S. A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. *Energy* **2018**, *145*, 710–720. [[CrossRef](#)]
127. Zhang, X.; Wang, J.; Zhang, K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electr. Power Syst. Res.* **2017**, *146*, 270–285. [[CrossRef](#)]
128. Li, S.; Kong, X.; Yue, L.; Liu, C.; Khan, M.A.; Yang, Z.; Zhang, H. Short-term electrical load forecasting using hybrid model of manta ray foraging optimization and support vector regression. *J. Clean. Prod.* **2023**, *388*, 135856. [[CrossRef](#)]
129. Peng, L.-L.; Fan, G.-F.; Huang, M.-L.; Hong, W.-C. Hybridizing DEMD and Quantum PSO with SVR in Electric Load Forecasting. *Energies* **2016**, *9*, 221. [[CrossRef](#)]
130. Huang, M.-L. Hybridization of chaotic quantum particle swarm optimization with SVR in electric demand forecasting. *Energies* **2016**, *9*, 426. [[CrossRef](#)]
131. Lee, C.-W.; Lin, B.-Y. Applications of the chaotic quantum genetic algorithm with support vector regression in load forecasting. *Energies* **2017**, *10*, 1832. [[CrossRef](#)]
132. Zeng, N.; Zhang, H.; Liu, W.; Liang, J.; Alsaadi, F.E. A switching delayed PSO optimized extreme learning machine for short-term load forecasting. *Neurocomputing* **2017**, *240*, 175–182. [[CrossRef](#)]
133. Sun, W.; Ye, M. Short-Term Load Forecasting Based on Wavelet Transform and Least Squares Support Vector Machine Optimized by Fruit Fly Optimization Algorithm. *J. Electr. Comput. Eng.* **2015**, *2015*, 862185. [[CrossRef](#)]
134. Ghayekhloo, M.; Menhaj, M.; Ghofrani, M. A hybrid short-term load forecasting with a new data preprocessing framework. *Electr. Power Syst. Res.* **2015**, *119*, 138–148. [[CrossRef](#)]
135. Awan, S.M.; Aslam, M.; Khan, Z.A.; Saeed, H. An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting. *Neural Comput. Appl.* **2014**, *25*, 1967–1978. [[CrossRef](#)]
136. Xiao, L.; Wang, J.; Hou, R.; Wu, J. A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting. *Energy* **2015**, *82*, 524–549. [[CrossRef](#)]
137. Xiao, L.; Shao, W.; Wang, C.; Zhang, K.; Lu, H. Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. *Appl. Energy* **2016**, *180*, 213–233. [[CrossRef](#)]
138. Hu, R.; Wen, S.; Zeng, Z.; Huang, T. A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. *Neurocomputing* **2017**, *221*, 24–31. [[CrossRef](#)]
139. Jiang, H.; Zhang, Y.; Muljadi, E.; Zhang, J.J.; Gao, D.W. A Short-Term and High-Resolution Distribution System Load Forecasting Approach Using Support Vector Regression with Hybrid Parameters Optimization. *IEEE Trans. Smart Grid* **2018**, *9*, 3341–3350. [[CrossRef](#)]
140. Zhang, S.; Wang, J.; Guo, Z. Research on combined model based on multi-objective optimization and application in time series forecast. *Soft Comput.* **2019**, *23*, 11493–11521. [[CrossRef](#)]
141. Bacanin, N.; Jovanovic, L.; Zivkovic, M.; Kandasamy, V.; Antonijevic, M.; Deveci, M.; Strumberger, I. Multivariate energy forecasting via metaheuristic tuned long-short term memory and gated recurrent unit neural networks. *Inf. Sci.* **2023**, *642*, 119122. [[CrossRef](#)]

142. Zhuang, Z.; Zheng, X.; Chen, Z.; Jin, T. A reliable short-term power load forecasting method based on VMD-IWOA-LSTM algorithm. *IEEJ Trans. Electr. Electron. Eng.* **2022**, *17*, 1121–1132. [[CrossRef](#)]
143. Heydari, A.; Majidi Nezhad, M.; Pirshayan, E.; Astiaso Garcia, D.; Keynia, F.; De Santoli, L. Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. *Appl. Energy* **2020**, *277*, 115503. [[CrossRef](#)]
144. Pan, L.; Feng, X.; Sang, F.; Li, L.; Leng, M.; Chen, X. An improved back propagation neural network based on complexity decomposition technology and modified flower pollination optimization for short-term load forecasting. *Neural Comput. Appl.* **2019**, *31*, 2679–2697. [[CrossRef](#)]
145. Liu, T.; Jin, Y.; Gao, Y. A New Hybrid Approach for Short-Term Electric Load Forecasting Applying Support Vector Machine with Ensemble Empirical Mode Decomposition and Whale Optimization. *Energies* **2019**, *12*, 1520. [[CrossRef](#)]
146. Electrical Demand, Generation by Type, Prices and Weather in Spain. Available online: <https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather> (accessed on 21 January 2024).
147. Yin, C.; Mao, S. Fractional multivariate grey Bernoulli model combined with improved grey wolf algorithm: Application in short-term power load forecasting. *Energy* **2023**, *269*, 126844. [[CrossRef](#)]
148. Zulfiqar, M.; Kamran, M.; Rasheed, M.B.; Alquthami, T.; Milyani, A.H. A hybrid framework for short term load forecasting with a novel feature engineering and adaptive grasshopper optimization in smart grid. *Appl. Energy* **2023**, *338*, 120829. [[CrossRef](#)]
149. Lu, Y.; Wang, G. A load forecasting model based on support vector regression with whale optimization algorithm. *Multimed. Tools Appl.* **2023**, *82*, 9939–9959. [[CrossRef](#)]
150. Geng, G.; He, Y.; Zhang, J.; Qin, T.; Yang, B. Short-Term Power Load Forecasting Based on PSO-Optimized VMD-TCN-Attention Mechanism. *Energies* **2023**, *16*, 4616. [[CrossRef](#)]
151. Gong, R.; Li, X. A Short-Term Load Forecasting Model Based on Crisscross Grey Wolf Optimizer and Dual-Stage Attention Mechanism. *Energies* **2023**, *16*, 2878. [[CrossRef](#)]
152. Motwakel, A.; Alabdulkreem, E.; Gaddah, A.; Marzouk, R.; Salem, N.M.; Zamani, A.S.; Abdelmageed, A.A.; Eldesouki, M.I. Wild Horse Optimization with Deep Learning-Driven Short-Term Load Forecasting Scheme for Smart Grids. *Sustainability* **2023**, *15*, 1524. [[CrossRef](#)]
153. Kiruthiga, D.; Manikandan, V. Levy flight-particle swarm optimization-assisted BiLSTM + dropout deep learning model for short-term load forecasting. *Neural Comput. Appl.* **2023**, *35*, 2679–2700. [[CrossRef](#)]
154. Su, J.; Han, X.; Hong, Y. Short Term Power Load Forecasting Based on PSVMD-CGA Model. *Sustainability* **2023**, *15*, 2941. [[CrossRef](#)]
155. 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](#)]
156. Fan, G.F.; Li, Y.; Zhang, X.Y.; Yeh, Y.H.; Hong, W.C. Short-term Load Forecasting Based on a Generalized Regression Neural network optimized by an improved sparrow search algorithm using the empirical wavelet decomposition method. *Energy Sci. Eng.* **2023**, *11*, 2444–2468. [[CrossRef](#)]
157. Wang, N.; Li, Z. Short term power load forecasting based on BES-VMD and CNN-Bi-LSTM method with error correction. *Front. Energy Res.* **2023**, *10*, 1076529. [[CrossRef](#)]
158. Huang, Y.; Huang, Z.; Yu, J.; Dai, X.; Li, Y. Short-term load forecasting based on IPSO-DBiLSTM network with variational mode decomposition and attention mechanism. *Appl. Intell.* **2023**, *53*, 12701–12718. [[CrossRef](#)]
159. Zhang, S.; Zhang, N.; Zhang, Z.; Chen, Y. Electric Power Load Forecasting Method Based on a Support Vector Machine Optimized by the Improved Seagull Optimization Algorithm. *Energies* **2022**, *15*, 9197. [[CrossRef](#)]
160. Zhao, X.; Shen, B.; Lin, L.; Liu, D.; Yan, M.; Li, G. Residential Electricity Load Forecasting Based on Fuzzy Cluster Analysis and LSSVM with Optimization by the Fireworks Algorithm. *Sustainability* **2022**, *14*, 1312. [[CrossRef](#)]
161. Jiang, F.; Zhang, W.; Peng, Z. Multivariate Adaptive Step Fruit Fly Optimization Algorithm Optimized Generalized Regression Neural Network for Short-Term Power Load Forecasting. *Front. Environ. Sci.* **2022**, *10*, 873939. [[CrossRef](#)]
162. Chen, Z.; Jin, T.; Zheng, X.; Liu, Y.; Zhuang, Z.; Mohamed, M.A. An innovative method-based CEEMDAN-IGWO-GRU hybrid algorithm for short-term load forecasting. *Electr. Eng.* **2022**, *104*, 3137–3156. [[CrossRef](#)]
163. Li, C.; Guo, Q.; Shao, L.; Li, J.; Wu, H. Research on Short-Term Load Forecasting Based on Optimized GRU Neural Network. *Electronics* **2022**, *11*, 3834. [[CrossRef](#)]
164. Liu, J.; Yin, Y. Power Load Forecasting Considering Climate Factors Based on IPSO-Elman Method in China. *Energies* **2022**, *15*, 1236. [[CrossRef](#)]
165. Wang, G.; Wang, X.; Wang, Z.; Ma, C.; Song, Z. A VMD&ndash;CISSA&ndash;LSSVM Based Electricity Load Forecasting Model. *Mathematics* **2022**, *10*, 28.
166. Xian, H.; Che, J. Multi-space collaboration framework based optimal model selection for power load forecasting. *Appl. Energy* **2022**, *314*, 118937. [[CrossRef](#)]
167. Yuan, F.; Che, J. An ensemble multi-step M-RMLSSVR model based on VMD and two-group strategy for day-ahead short-term load forecasting. *Knowl. Based Syst.* **2022**, *252*, 109440. [[CrossRef](#)]
168. Hafeez, G.; Khan, I.; Jan, S.; Shah, I.A.; Khan, F.A.; Derhab, A. A novel hybrid load forecasting framework with intelligent feature engineering and optimization algorithm in smart grid. *Appl. Energy* **2021**, *299*, 117178. [[CrossRef](#)]

169. Shafiei Chafi, Z.; Afrakhte, H. Short-Term Load Forecasting Using Neural Network and Particle Swarm Optimization (PSO) Algorithm. *Math. Probl. Eng.* **2021**, *2021*, 5598267. [[CrossRef](#)]
170. Zhou, M.; Hu, T.; Bian, K.; Lai, W.; Hu, F.; Hamrani, O.; Zhu, Z. Short-Term Electric Load Forecasting Based on Variational Mode Decomposition and Grey Wolf Optimization. *Energies* **2021**, *14*, 4890. [[CrossRef](#)]
171. Peng, H.; Wen, W.-S.; Tseng, M.-L.; Li, L.-L. A cloud load forecasting model with nonlinear changes using whale optimization algorithm hybrid strategy. *Soft Comput.* **2021**, *25*, 10205–10220. [[CrossRef](#)]
172. Bao-De, L.; Xin-Yang, Z.; Mei, Z.; Hui, L.; Guang-Qian, L. Improved genetic algorithm-based research on optimization of least square support vector machines: An application of load forecasting. *Soft Comput.* **2021**, *25*, 11997–12005. [[CrossRef](#)]
173. Xian, H.; Che, J. A variable weight combined model based on time similarity and particle swarm optimization for short-term power load forecasting. *IAENG Int. J. Comput. Sci.* **2021**, *48*, 915–924.
174. Aslam, S.; Ayub, N.; Farooq, U.; Alvi, M.J.; Albogamy, F.R.; Rukh, G.; Haider, S.I.; Azar, A.T.; Bukhsh, R. Towards Electric Price and Load Forecasting Using CNN-Based Ensembler in Smart Grid. *Sustainability* **2021**, *13*, 12653. [[CrossRef](#)]
175. Wang, X.; Gao, X.; Wang, Z.; Ma, C.; Song, Z. A Combined Model Based on EOBL-CSSA-LSSVM for Power Load Forecasting. *Symmetry* **2021**, *13*, 1579. [[CrossRef](#)]
176. Wu, X.; Wang, Y.; Bai, Y.; Zhu, Z.; Xia, A. Online short-term load forecasting methods using hybrids of single multiplicative neuron model, particle swarm optimization variants and nonlinear filters. *Energy Rep.* **2021**, *7*, 683–692. [[CrossRef](#)]
177. Talaat, M.; Farahat, M.A.; Mansour, N.; Hatata, A.Y. Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach. *Energy* **2020**, *196*, 117087. [[CrossRef](#)]
178. Xie, K.; Yi, H.; Hu, G.; Li, L.; Fan, Z. Short-term power load forecasting based on Elman neural network with particle swarm optimization. *Neurocomputing* **2020**, *416*, 136–142. [[CrossRef](#)]
179. Wu, F.; Cattani, C.; Song, W.; Zio, E. Fractional ARIMA with an improved cuckoo search optimization for the efficient Short-term power load forecasting. *Alex. Eng. J.* **2020**, *59*, 3111–3118. [[CrossRef](#)]
180. Yang, Y.; Shang, Z.; Chen, Y.; Chen, Y. Multi-Objective Particle Swarm Optimization Algorithm for Multi-Step Electric Load Forecasting. *Energies* **2020**, *13*, 532. [[CrossRef](#)]
181. Song, W.; Cattani, C.; Chi, C.-H. Multifractional Brownian motion and quantum-behaved particle swarm optimization for short term power load forecasting: An integrated approach. *Energy* **2020**, *194*, 116847. [[CrossRef](#)]
182. Kong, X.; Li, C.; Zheng, F.; Wang, C. Improved Deep Belief Network for Short-Term Load Forecasting Considering Demand-Side Management. *IEEE Trans. Power Syst.* **2020**, *35*, 1531–1538. [[CrossRef](#)]
183. Shang, Z.; He, Z.; Song, Y.; Yang, Y.; Li, L.; Chen, Y. A Novel Combined Model for Short-Term Electric Load Forecasting Based on Whale Optimization Algorithm. *Neural Process. Lett.* **2020**, *52*, 1207–1232. [[CrossRef](#)]
184. Li, T.; Qian, Z.; He, T. Short-Term Load Forecasting with Improved CEEMDAN and GWO-Based Multiple Kernel ELM. *Complexity* **2020**, *2020*, 1209547. [[CrossRef](#)]
185. Salami, M.; Sobhani, F.M.; Ghazizadeh, M.S. A hybrid short-term load forecasting model developed by factor and feature selection algorithms using improved grasshopper optimization algorithm and principal component analysis. *Electr. Eng.* **2020**, *102*, 437–460. [[CrossRef](#)]
186. Lu, H.; Azimi, M.; Iseley, T. Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine. *Energy Rep.* **2019**, *5*, 666–677. [[CrossRef](#)]
187. Bento, P.M.R.; Pombo, J.A.N.; Calado, M.R.A.; Mariano, S.J.P.S. Optimization of neural network with wavelet transform and improved data selection using bat algorithm for short-term load forecasting. *Neurocomputing* **2019**, *358*, 53–71. [[CrossRef](#)]
188. Wang, R.; Wang, J.; Xu, Y. A novel combined model based on hybrid optimization algorithm for electrical load forecasting. *Appl. Soft Comput.* **2019**, *82*, 105548. [[CrossRef](#)]
189. Ahmad, W.; Ayub, N.; Ali, T.; Irfan, M.; Awais, M.; Shiraz, M.; Glowacz, A. Towards Short Term Electricity Load Forecasting Using Improved Support Vector Machine and Extreme Learning Machine. *Energies* **2020**, *13*, 2907. [[CrossRef](#)]
190. Yang, A.; Li, W.; Yang, X. Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines. *Knowl.-Based Syst.* **2019**, *163*, 159–173. [[CrossRef](#)]
191. Barman, M.; Dev Choudhury, N.B. Season specific approach for short-term load forecasting based on hybrid FA-SVM and similarity concept. *Energy* **2019**, *174*, 886–896. [[CrossRef](#)]
192. Zhang, B.; Liu, W.; Li, S.; Wang, W.; Zou, H.; Dou, Z. Short-term load forecasting based on wavelet neural network with adaptive mutation bat optimization algorithm. *IEEJ Trans. Electr. Electron. Eng.* **2019**, *14*, 376–382. [[CrossRef](#)]
193. Hong, W.-C.; Fan, G.-F. Hybrid Empirical Mode Decomposition with Support Vector Regression Model for Short Term Load Forecasting. *Energies* **2019**, *12*, 1093. [[CrossRef](#)]
194. Xiong, Y. Study on Short-Term Micro-Grid Load Forecasting Based on IGA-PSO RBF Neural Network. Master's Thesis, South China University of Technology, Guangzhou, China, 2016.
195. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [[CrossRef](#)]
196. Luy, M.; Ates, V.; Barisci, N.; Polat, H.; Cam, E. Short-Term Fuzzy Load Forecasting Model Using Genetic-Fuzzy and Ant Colony-Fuzzy Knowledge Base Optimization. *Appl. Sci.* **2018**, *8*, 864. [[CrossRef](#)]
197. Ray, P.; Arya, S.R.; Nandkeolyar, S. Electric load forecasts by metaheuristic based back propagation approach. *J. Green Eng.* **2017**, *7*, 61–82. [[CrossRef](#)]

198. Hong, W.-C. Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model. *Energy Convers. Manag.* **2009**, *50*, 105–117. [[CrossRef](#)]
199. 2014 Global Energy Forecasting Competition. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0169207016000133> (accessed on 6 February 2024).
200. Available online: [http://www.nyiso.com/public/markets\\_operations/index.jsp](http://www.nyiso.com/public/markets_operations/index.jsp) (accessed on 26 January 2024).
201. Wang, J.; Zhu, W.; Zhang, W.; Sun, D. A trend fixed on firstly and seasonal adjustment model combined with the  $\epsilon$ -SVR for short-term forecasting of electricity demand. *Energy Policy* **2009**, *37*, 4901–4909. [[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.