




## Review

# The State of the Art Electricity Load and Price Forecasting for the Modern Wholesale Electricity Market

Vasileios Laitos<sup>1,\*</sup>, Georgios Vontzos<sup>1</sup>, Paschalis Paraschoudis<sup>2</sup>, Eleftherios Tsampasis<sup>3</sup>,  
Dimitrios Bargiotas<sup>1,\*</sup> and Lefteri H. Tsoukalas<sup>4</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, University of Thessaly, 383 34 Volos, Greece; gvontzos@uth.gr

<sup>2</sup> Department of Energy Systems, University of Thessaly, 415 00 Larisa, Greece; pparaschoudis@uth.gr

<sup>3</sup> Department of Digital Arts and Cinema, National and Kapodestrian University of Athens, 344 00 Psahna, Greece; etsampasis@uoa.gr

<sup>4</sup> Center for Intelligent Energy Systems (CiENS), School of Nuclear Engineering, Purdue University, West Lafayette, IN 47906, USA; tsoukala@purdue.edu

\* Correspondence: vlaitos@uth.gr (V.L.); bargiotas@uth.gr (D.B.)

**Abstract:** In a modern and dynamic electricity market, ensuring reliable, sustainable and efficient electricity distribution is a pillar of primary importance for grid operation. The high penetration of renewable energy sources and the formation of competitive prices for utilities play a critical role in the wider economic development. Electricity load and price forecasting have been a key focus of researchers in the last decade due to the substantial economic implications for both producers, aggregators and end consumers. Many forecasting techniques and methods have emerged during this period. This paper conducts an extensive and analytical review of the prevailing load and electricity price forecasting methods in the context of the modern wholesale electricity market. The study is separated into seven main sections. The first section provides the key challenges and the main contributions of this study. The second section delves into the workings of the electricity market, providing a detailed analysis of the three markets that have evolved, their functions and the key factors influencing overall market dynamics. In the third section, the main methodologies of electricity load and price forecasting approaches are analyzed in detail. The fourth section offers a comprehensive review of the existing literature focusing on load forecasting, highlighting various methodologies, models and their applications in this field. This section emphasizes the advances that have been made in all categories of forecasting models and their practical application in different market scenarios. The fifth section focuses on electricity price forecasting studies, summarizing important research papers investigating various modeling approaches. The sixth section constitutes a fundamental discussion and comparison between the load- and price-focused studies that are analyzed. Finally, by examining both traditional and cutting-edge forecasting methods, this review identifies key trends, challenges and future directions in the field. Overall, this paper aims to provide an in-depth analysis leading to the understanding of the state-of-the-art models in load and price forecasting and to be an important resource for researchers and professionals in the energy industry. Based on the research conducted, there is an increasing trend in the use of artificial intelligence models in recent years, due to the flexibility and adaptability they offer for big datasets, compared to traditional models. The combination of models, such as ensemble methods, gives us very promising results.

**Keywords:** electricity load forecasting; electricity price forecasting; wholesale energy market; deep learning; machine learning; statistical models



**Citation:** Laitos, V.; Vontzos, G.; Paraschoudis, P.; Tsampasis, E.; Bargiotas, D.; Tsoukalas, L.H. The State of the Art Electricity Load and Price Forecasting for the Modern Wholesale Electricity Market. *Energies* **2024**, *17*, 5797. <https://doi.org/10.3390/en17225797>

Academic Editor: Sara Carcangiu

Received: 29 October 2024

Revised: 15 November 2024

Accepted: 18 November 2024

Published: 20 November 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

The modern energy market faces several issues and challenges due to facts such as dynamic market regulations, the new policies established for environmental protection

and the high penetration of renewable energy sources (RESs). RESs such as wind and solar energy, despite their environmental benefits, cause variability and unpredictability in the operation of the power system. This variability complicates the balance of supply and demand side, making it harder to maintain grid stability. Additionally, the transition to decentralized energy networks, where large and small-scale consumers generate their own electricity, creates additional difficulties in managing the system. As a result, energy providers and regulators must constantly adjust to these conditions to meet demand and incorporate these new, variable energy sources.

Accurate forecasting is a key pillar for the optimal operation of the energy market. Accurate prediction of electricity helps grid operators to manage periods of peak and non-peak demand, ensuring the power balance in the transmission and distribution network. Simultaneously, electricity price forecasting helps energy companies and large-scale consumers decide the best times to buy or sell energy. This capability is crucial in a market with high price variability, where poor predictions can lead to significant financial losses, creating high risks for all parts of the system. In addition to helping with operational decisions, accurate forecasts also support strategic planning, investment in infrastructure, and the integration of RESs. By providing a clearer picture of future conditions, forecasting helps stabilize the market and ensures a reliable energy supply. Also, Distributed Energy Resources (DERs), like solar, wind farms and battery storage, are transforming the energy landscape by decentralizing power generation and storage. While they enhance flexibility and resilience within energy systems, they also add layers of complexity to load and price forecasting due to their variability and the unpredictability of these resources. So, advanced forecasting models that can account for fluctuating inputs and decentralized energy flows are required, making forecasting approaches a more challenging task.

Electricity Load (ELF) and Price Forecasting (EPF) are crucial elements for the efficient operation of modern wholesale electricity markets. The integration of RESs, particularly wind power, presents new challenges and opportunities in these domains. This paper offers a comprehensive review of recent advancements in ELF and EPF methodologies. We explore state-of-the-art studies, emphasizing the importance of accurate forecasts for maintaining stability and optimizing market operations, highlighting the interplay and differences between load and price forecasting. The contribution and key areas of our research include:

- The state-of-the-art short-term forecasting approaches for electricity load and prices, underscoring the need for precise predictions of demand fluctuations and price volatility.
- An in-depth mathematical analysis of various forecasting models, including statistical, machine and deep learning, and ensemble approaches.
- The application of statistical, artificial intelligence and hybrid models in the time series analysis of diverse load and price data, crucial for enhancing forecast accuracy.
- An extended review of evolving trends in ELF and EPF, incorporating economic and engineering perspectives that drive these trends.
- Case studies from different countries, providing practical examples of electricity market operations and the management of renewable energy integration in forecasting models.
- The identification of trends and potential gaps in the two fields we are studying is important. More specifically, it appears that in the case of EPF, ensemble methods seem to gain more and greater value due to the stability and reliability they provide in predictions.
- The formulation and identification of significant proposals for future studies and research, capable of further developing the tools used in the wholesale electricity market.

By addressing these areas, this paper aims to deepen the understanding of ELF and EPF methodologies, providing the most remarkable studies that have been performed in recent years in these fields. Thus, the main goal is the analysis of these works and the

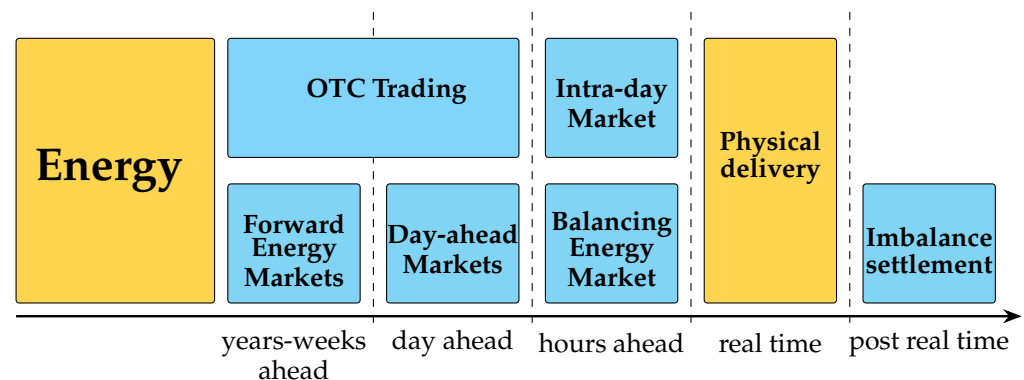
highlighting of the models and various metrics used for the different types and origins of datasets and not the emergence of the best models, as any comparison leading to safe conclusions should be carried out under the same conditions, such as the datasets and input variables of the models. Also, our study seeks to identify new avenues for research, promoting further innovation in electricity load and price forecasting for the modern wholesale market.

Regarding the structure of this paper, in Section 2, the electricity market working mechanism is analyzed in detail. Then, in Section 3, the most prevalent models used for load and price forecasting are presented in detail. Following that, in Section 4, the most comprehensive papers dealing with ELF are analyzed. Additionally, Section 5 presents the corresponding papers in the area of price forecasting. In Section 6, a discussion and comparison between load and price forecasting approaches is made. Finally, useful conclusions and future study proposals are discussed.

## 2. Electricity Market Working Mechanisms

The modern wholesale electricity market is a dynamic and complex system where electricity is bought and sold primarily through competitive bidding processes. It comprises various participants, including generators, retailers, and large consumers, who interact in day-ahead, intraday, and real-time markets to ensure an efficient and reliable supply of electricity. Market operators and power exchanges facilitate these transactions, ensuring transparency and fair pricing. The market incorporates advanced technologies and algorithms for demand forecasting, price setting, and grid management while also integrating RESs to promote sustainability. Regulatory frameworks vary by region but aim to balance competition, reliability, and environmental goals.

Bilateral contracts affect electricity markets by providing stability and predictability in pricing for both buyers and sellers, reducing market volatility. They facilitate long-term planning and investment in energy infrastructure and can help hedge against price fluctuations in the spot market [1]. The general idea about the electricity market working sequence is presented in Figure 1.



**Figure 1.** Energy market sequence [2].

### 2.1. Day-Ahead Market

The day-ahead market (DAM) involves buying and selling electricity one day before it is generated and delivered. It occurs either in the spot market of a power exchange or through direct agreements between two parties, typically power-trading companies, in over-the-counter (OTC) deals. The main functions of day-ahead markets are [3]:

- **Balancing Supply and Demand:** Ensuring that the anticipated electricity supply matches the expected demand for the following day.
- **Market Equilibrium:** Creating a fair and competitive environment by matching bids and offers through a transparent auction process.
- **Resource Optimization:** Facilitating the efficient use of generation resources by signaling which plants should be operational based on forecasted prices.

The day-ahead market aligns electricity supply and demand before production and delivery by gathering bids from producers and consumers to set hourly prices, ensuring market equilibrium. Market Pools or Power Exchanges publish these hourly prices for the following day. This information helps synchronize the anticipated electricity supply with expected customer usage. In the bidding process, participants indicate the volume of electricity they are able to purchase or supply each hour, along with the corresponding price levels in EUR/MWh (or USD/MWh) [4].

The regulations for day-ahead trading can differ greatly between national markets. These differences include lead times, trading periods, smallest tradable amounts of electricity, various pricing methods, gate closure times, and other specific details [3].

In the DAM, the bidding time period typically refers to the timeframe during which market participants submit their bids and offers for the next operating day. The general timelines and processes involved in the DAM, including the bid submission deadline and the delivery period, are the following:

- Bid Submission Deadline: 12:00 p.m. (noon) on day D-1.
- Market Clearing and Results Publication: Usually by early evening on day D-1.
- Operating Day (Day D): The delivery period typically starts at 12:00 a.m. (midnight) and covers the full 24 h of day D.

From the bid submission deadline to the start of the delivery period, there are typically approximately 12 h. Different cases of how the DAM operates in different markets follow: In the PJM Interconnection market, participants must submit their bids by 12:00 p.m. Eastern Time on the day before the delivery of the electricity, named as day D-1. The market results are typically published by the early evening of the same day. The delivery period for the transactions occurs from 12:00 a.m. to 11:59 p.m. on day D. For the NYISO (New York Independent System Operator), the bid submission deadline is set at 5:00 a.m. Eastern Time on day D-1. The market results are made available by 10:00 a.m. on the same day. Similar to PJM Interconnection, the delivery period lasts from 12:00 a.m. to 11:59 p.m. on day D.

Additionally, in the EPEX SPOT market, bids must be submitted by 12:00 p.m. (noon) Central European Time (CET) on day D-1. The market results are published shortly after, by 12:42 p.m. CET on day D-1. The delivery period for these transactions covers 12:00 a.m. (midnight) to 11:59 p.m. on day D. Finally, for the Nord Pool market, the bid submission deadline is also 12:00 p.m. (noon) Central European Time on day D-1. The market results are similarly published by 12:42 p.m. CET on the same day. As with the other markets, the delivery period runs from 12:00 a.m. (midnight) to 11:59 p.m. on day D [2].

## 2.2. Intraday Market

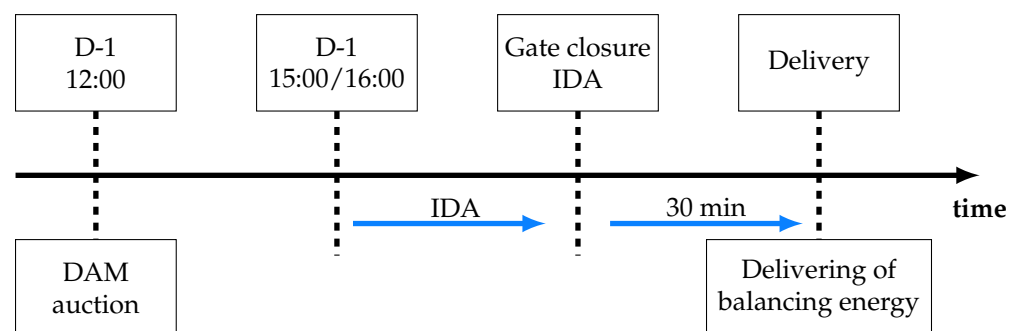
In the intraday market (IDA), the energy is traded on the same day it is delivered and is executed in power exchange markets or through direct agreements between buyers and sellers, like the DAM. The main functions of the IDA are managing imbalances by allowing market participants to handle deviations from their scheduled energy deliveries in the DAM. This helps in reducing penalties and associated costs. The IDA offers opportunities for market participants to optimize their positions and mitigate important financial risks, helping fulfill the balancing of group commitments and minimizing potential imbalance costs. As power assets become more flexible, it allows for rapid power generation adjustments, optimizing profitability and maintaining system stability [5].

The increasing penetration of intermittent RESs has heightened interest in IDAs, as market participants face greater challenges in maintaining balance after the DAM closes. The IDA allows for adjustments to be made closer to the actual delivery time, accommodating changes in demand forecasts and generation availability. Here are the general timelines [6]:

- Continuous Trading: IDAs often operate on a continuous trading basis, allowing participants to trade up until a few hours before delivery.

- Gate Closure Times: These vary by market but are typically set a few hours before the delivery period starts. For example, gate closure might occur 1 to 2 h before the delivery period.
- Delivery Period: The IDA covers the hours remaining in the operating day (day D).

Cases of how the IDA operates in different markets are as follows: In the EPEX SPOT (European Power Exchange) market, continuous trading is available up to 5 min before the delivery period in certain regions. The gate typically closes between 30 min and 1 h before delivery in most regions. The delivery period covers the remaining hours of the operating day. For Nord Pool, the trading is available up to 1 h before the delivery period. The gate closes 1 h prior to the start of the delivery period. As with EPEX SPOT, the delivery period includes the remaining hours of the operating day [2]. DAM and IDA operation hours are included in Figure 2.



**Figure 2.** Day-Ahead and Intraday Market.

### 2.3. Balancing Market

The balancing market (BM), which is presented in Figure 3, is a critical component of modern electricity markets, ensuring immediate equilibrium between electricity supply and demand in real-time operations. In contrast to the DAM, which plans energy transactions for the next day, the BM responds swiftly to unforeseen fluctuations in supply and demand. Participants such as generators and consumers submit bids to adjust electricity production or consumption at short notice. These bids are managed by the system operator, typically an Independent System Operator (ISO) or Regional Transmission Organization (RTO), which monitors grid conditions continuously. When discrepancies arise between predicted and actual grid conditions, the system operator activates these bids to maintain grid stability, ensuring a consistent match between electricity supply and demand [7].

The BM involves the activation of reserves to handle unforeseen imbalances. The general timelines are:

- Real-Time Operation: The balancing market operates in real time, with continuous adjustments being made to balance supply and demand.
- Activation of Reserves: Reserve power is activated as needed to address imbalances. This can occur at any time during the operating day.
- Settlement: Imbalances are settled after the operating day, based on the actual deviations from scheduled supply and demand.

The pricing in the BM reflects the real-time cost of adjusting electricity supply or demand, often resulting in higher prices compared to the DAM due to the immediate nature of the required actions. The costs in the BM are usually covered by imbalance charges on market participants whose differences from their planned positions require balancing actions. This mechanism encourages market participants to keep closely to their planned commitments, which improves the performance and sustainability of the power system. Additionally, this market mechanism plays a vital role in integrating RESs, which are often variable and less predictable, providing flexible platforms for managing the increased penetration of RESs into the grid [2].

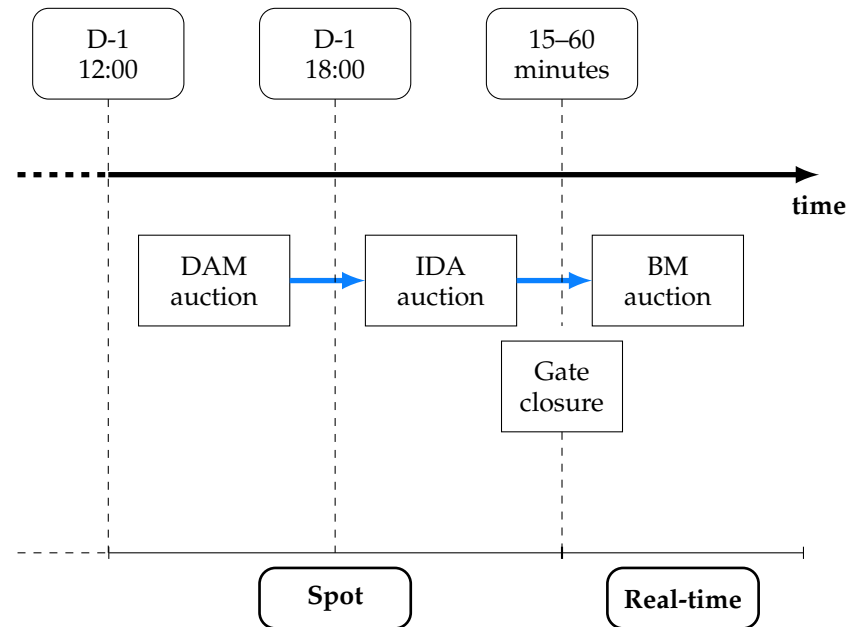


Figure 3. Balancing market [8].

### 3. Forecasting Methodologies for Load and Price

The methodologies used for ELF and EPF have been enhanced and have progressed to a high level in the last five to ten years. Despite the fact that, in the early years, we saw statistical models being used, with advances in technology, there is an increasing trend toward the use of machine learning (ML), deep learning (DL), and ensemble methods, which appear to dominate over the former due to the flexibility and high performance they offer. Starting from traditional methods, there will be an extensive mathematical presentation of the most fundamental models, both statistical and those involving machine, deep and ensemble learning methods. This section provides a detailed overview of the most common models used for these purposes. Initially, the statistical models are presented, followed by the machine and deep learning, and finally, the ensemble methods are analyzed. For a better understanding of the categories and the variety of the models, Figure 4 graphically presents the aforementioned forecasting models.

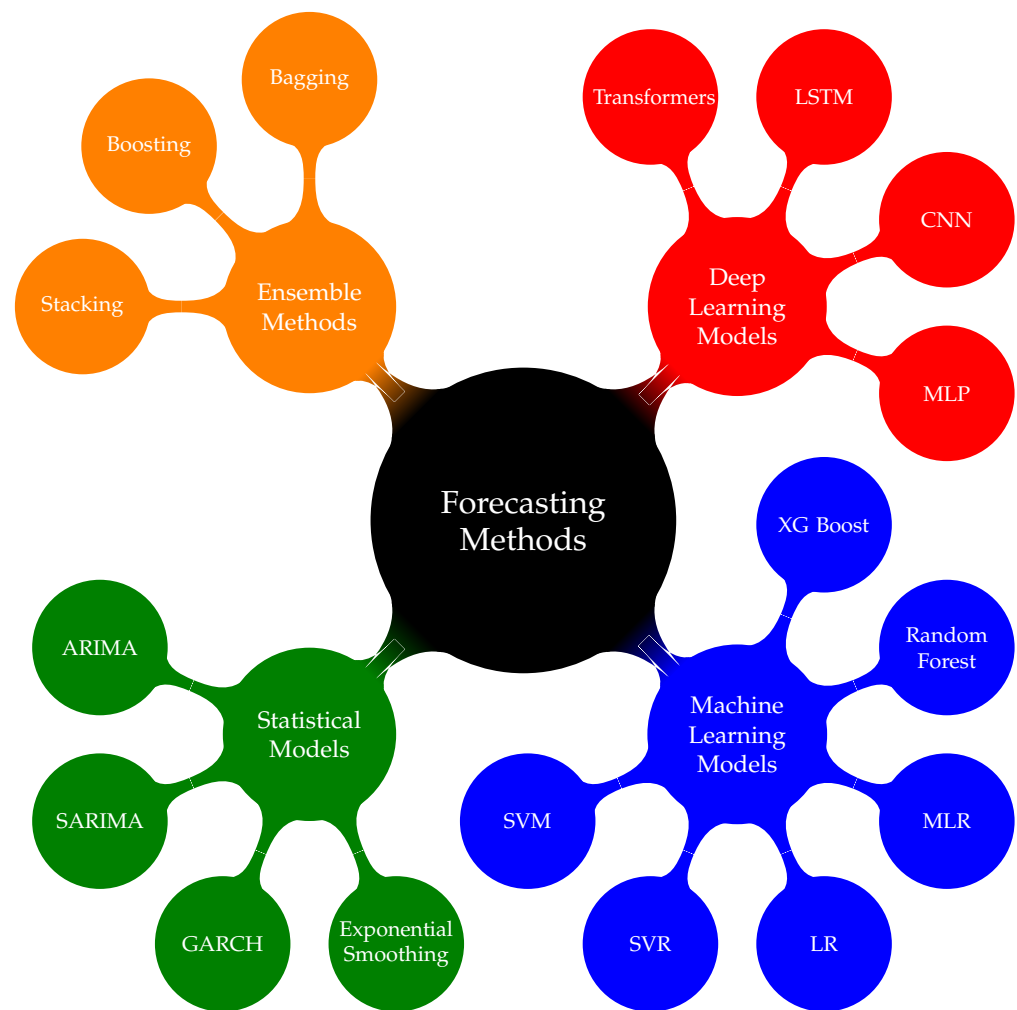
#### 3.1. Statistical Models

Statistical models are essential in ELF and EPF, providing robust frameworks for understanding and predicting trends in time series data. Among the most prevalent models are AutoRegressive Integrated Moving Average (ARIMA) models, which combine autoregression, differencing, and moving averages to model temporal dependencies and achieve stationarity. Seasonal ARIMA (SARIMA) models manage seasonal changes effectively by using seasonal differencing and seasonal autoregressive and moving average parts. Exponential smoothing methods, like Holt–Winters, are popular for being simple and effective in dealing with seasonality and trends in time series. GARCH models are used to capture and predict volatility patterns, mainly in financial data. State space models and Kalman filters offer a flexible way to model dynamic systems, especially for time-varying processes. These methods are important in the energy sector and are presented in detail below.

##### 3.1.1. AutoRegressive Integrated Moving Average

AutoRegressive Integrated Moving Average (ARIMA) analyzes time series data and predicts future trends based on historical values. This model is characterized by three main parameters:  $p$ ,  $d$ , and  $q$ . The first parameter  $p$  is the number of lag observations of the model, which constitutes the autoregressive part (AR). The second parameter  $d$  is the number of times the raw observations are differenced to achieve stationarity, which is the

integrated part (I). The third parameter  $q$  refers to the size of the moving average window, which is the moving average part of the model (MA) [9].



**Figure 4.** Forecasting methods overview.

The basic equations of this model can be expressed as follows:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \quad (1)$$

where  $y_t$  represents the value at time point  $t$ ,  $c$  is a constant,  $\phi_i$  are the coefficients for the AR part,  $\theta_j$  are the coefficients for the MA part, and  $\epsilon_t$  is the error term at time  $t$ . The autoregressive (AR) part of the model means that the output is directly influenced by its past values and can be written as:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t \quad (2)$$

The integrated (I) part of the model is used for differencing the raw observations, in order to make the data more stationary. It can be expressed as:

$$y'_t = y_t - y_{t-1} \quad (3)$$



Finally, the MA part represents the dependency between an observation and a residual error. It can be written as:

$$y_t = c + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad (4)$$

### 3.1.2. Seasonal AutoRegressive Integrated Moving Average

Seasonal AutoRegressive Integrated Moving Average (SARIMA) is an extension of the ARIMA model because of its capability to handle seasonality in time series. The model integrates both non-seasonal and seasonal components. Three parameters  $p, d, q$  represent the seasonal part and four parameters  $P, D, Q, s$  are for the non-seasonal part.  $p$  is the number of non-seasonal AR terms,  $d$  is the number of non-seasonal differences, and  $q$  is the number of non-seasonal MA terms. For the seasonal components,  $P$  indicates the number of seasonal autoregressive terms,  $D$  represents the number of seasonal differences, and  $Q$  represents the number of seasonal moving average terms. Finally, the parameter  $s$  is the length of the seasonal cycle [10].

The main formulation of SARIMA can be expressed with the following equation [10]:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{k=1}^P \Phi_k Y_{t-ks} + \sum_{l=1}^Q \Theta_l \epsilon_{t-ls} + \epsilon_t \quad (5)$$

where  $\phi_i$  are the non-seasonal AR factors,  $\theta_i$  are the factors of the MA part,  $\Phi_k$  are the factors for the seasonal AR part,  $\Theta_i$  are the MA seasonal factors,  $s$  is the seasonal period, and  $\epsilon_t$  is the error term at time  $t$ .

#### Non-Seasonal and Seasonal Parts

The non-seasonal part is similar to the ARIMA model and can be written as the Equation (1). The seasonal part operates for the seasonal fluctuations and can be written as:

$$Y_t = \sum_{k=1}^P \Phi_k Y_{t-ks} + \sum_{l=1}^Q \Theta_l \epsilon_{t-ls} \quad (6)$$

where  $Y_t$  represents the seasonal component of  $y_t$ .

#### Differencing

The seasonal and non-seasonal differencing parts can be expressed as follows: The non-seasonal differencing can be expressed as Equation (3), while the seasonal differencing can be written as:

$$Y'_t = Y_t - Y_{t-s} \quad (7)$$

### 3.1.3. Exponential Smoothing

The exponential smoothing (ES) method is suitable and effective in dealing with univariate datasets. It gives more importance to recent data values by applying weights that reduced in exponential way from the recent to the older data, so recent observations influence the forecast more than older ones.

There are three main exponential smoothing methods:

- Single Exponential Smoothing (SES): This method is suitable for time series without trend or seasonality.
- Double Exponential Smoothing (DES): Proper for data with a trend.
- Triple Exponential Smoothing (TES): Or Holt–Winters, which is suitable for data with both trend and seasonality patterns.



### Single Exponential Smoothing (SES)

The main equation for SES is:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t \quad (8)$$

where  $\hat{y}_{t+1}$  is the prediction for the time  $t + 1$ ,  $\alpha$  is the smoothing parameter with  $0 < \alpha < 1$ , and  $y_t$  is the actual and  $\hat{y}_t$  is the predicted value at time  $t$ .

### Double Exponential Smoothing (DES)

DES introduces a second equation to account for trends:

$$\hat{y}_{t+1} = l_t + b_t \quad (9)$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (10)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (11)$$

where  $l_t$  is the level component at time  $t$ ,  $b_t$  is the trend component at time  $t$ , and  $\alpha$  and  $\beta$  are smoothing parameters, with  $0 < \alpha, \beta < 1$ .

### Triple Exponential Smoothing (TES)—Holt–Winters

TES, or the Holt–Winters method, adds a seasonal component, as follows:

$$\hat{y}_{t+m} = l_t + b_t m + s_{t+m-s} \quad (12)$$

$$l_t = \alpha \frac{y_t}{s_{t-s}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (13)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (14)$$

$$s_t = \gamma \frac{y_t}{l_t} + (1 - \gamma)s_{t-s} \quad (15)$$

where  $l_t$  is the level component at time  $t$ ;  $b_t$  represents the trend component;  $s_t$  is the seasonal component;  $s$  is the length of the seasonal cycle;  $\alpha$ ,  $\beta$ , and  $\gamma$  represent the smoothing variables, with values between 0 and 1; and  $m$  is the number of periods ahead to be forecasted [11].

#### 3.1.4. Generalized Autoregressive Conditional Heteroskedasticity

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is a statistical model used to forecast the future fluctuations of time series. It is mainly useful in financial applications, where GARCH predicts the volatility of financial markets. This model utilizes lagged terms of both the error and its variance and can be expressed through the mean and variance equations, as follows [12]:

The mean equation:

$$y_t = \mu + \epsilon_t \quad (16)$$

where  $y_t$  represents the value at time  $t$ ,  $\mu$  is the mean, and  $\epsilon_t$  is the error term.

The variance equation:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (17)$$

where  $\sigma_t^2$  is the conditional variance at time  $t$ ,  $\alpha_0$  is a constant,  $\alpha_i$  are the coefficients for the lagged squared residuals ( $\epsilon_{t-i}^2$ ),  $\beta_j$  are the coefficients for the lagged variances ( $\sigma_{t-j}^2$ ),  $p$  is the order of the GARCH terms, and  $q$  is the order of the ARCH terms. Finally, the error term  $\epsilon_t$  is typically modeled as:

$$\epsilon_t = \sigma_t z_t \quad (18)$$

where  $z_t$  is a white noise process with zero mean and unit variance.

### 3.2. Machine Learning Models

Machine learning (ML) models are effective in ELF and EPF tasks and are capable of dealing with complex and nonlinear datasets that contain fluctuations and volatility. They are used for both regression and classification tasks and the most commonly models are presented below.

#### 3.2.1. Support Vector Machine

Support vector machine (SVM) is a supervised learning model, of which the basic working mechanism is to find the hyperplane that best separates the data into two different groups, maximizing the margin among them. Given a set of training data  $\{(x_i, y_i)\}_{i=1}^n$ , where  $x_i$  is the feature vector and  $y_i$  is the label, the algorithm aims to find the optimal hyperplane for the best data separation. There are two categories, the linear and non-linear approach [13].

##### Linear SVM

For linear SVM, the working function is given by:

$$f(x) = w \cdot x + b \quad (19)$$

where  $w$  is the weight vector, and  $b$  is the bias term.

The optimization problem of finding the optimal hyperplane can be formulated as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (20)$$

subject to the constraints:

$$y_i(w \cdot x_i + b) \geq 1 \quad \text{for all } i = 1, 2, \dots, n \quad (21)$$

##### Non-Linear SVM

For non-linearly separable data, SVM can be extended using the kernel trick. The function becomes as follows:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \quad (22)$$

where  $\alpha_i$  are the Lagrange multipliers, and  $K(x_i, x)$  is the kernel function, which computes the inner product between two feature vectors in a high-dimensional space.

#### 3.2.2. Support Vector Regression

Support vector regression (SVR) is an extension of the SVM model used for regression tasks. It aims to find a function that approximates the relationship between input features and continuous output values with minimal error while maintaining model simplicity.

The objective of the model is to find a function  $f(x) = w \cdot x + b$  that approximates the target values  $y_i$  within a margin of  $\epsilon$ , while keeping the model complexity as low as possible. The SVR optimization problem is expressed as:

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (23)$$

Subject to the constraints:

$$\begin{aligned} y_i - (w \cdot x_i + b) &\leq \epsilon + \xi_i \\ (w \cdot x_i + b) - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (24)$$

where  $w$  is the weight vector,  $b$  is the bias term,  $\xi_i$  and  $\xi_i^*$  are slack variables for points outside the margin,  $C$  is the regularization parameter, and  $n$  is the number of training samples [13].

### 3.2.3. Linear Regression

Linear regression (LR) is a machine learning method, which assumes a linear relationship between an input variable  $x$  and the target variable  $y$ . The model can be represented as [14]:

$$y = \beta_0 + \beta_1 x + \epsilon \quad (25)$$

where  $y$  is the target variable,  $\beta_0$  is the intercept,  $\beta_1$  is the slope (coefficient) of the input variable  $x$ , and  $\epsilon$  is the error term.

The objective is to find the values of  $\beta_0$  and  $\beta_1$  that minimize the sum of squared residuals (the difference between the observed and predicted values):

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \quad (26)$$

where,  $n$  is the number of observations.

### 3.2.4. Multilinear Regression

Multilinear regression (MLR) extends linear regression by modeling the relationship between a continuous target variable and multiple input features. It assumes a linear relationship between the target variable  $y$  and multiple input variables  $\mathbf{x} = (x_1, x_2, \dots, x_p)$ . Similarly with linear regression, this model can be represented as [14]:

$$y = \beta_0 + \dots + \beta_p x_p + \epsilon \quad (27)$$

where  $y$  is the target variable,  $\beta_0$  is the intercept,  $\beta_i$  are the coefficients of the input variables  $x_i$ , and  $\epsilon$  is the error term.

The objective is to find the values of  $\beta_0, \beta_1, \dots, \beta_p$  that minimize the sum of squared residuals:

$$\min_{\beta_0, \beta_1, \dots, \beta_p} \sum_{i=1}^n (y_i - (\beta_0 + \sum_{j=1}^p \beta_j x_{ij}))^2 \quad (28)$$

where  $n$  is the number of observations, and  $p$  is the number of input variables.

### 3.2.5. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a very effective and scalable algorithm used for supervised learning tasks. It belongs to the ensemble learning family and is based on the gradient boosting framework. XGBoost combines the outputs of several models, typically decision trees. It trains each tree sequentially, with each tree correcting the errors made by the previous trees.

The objective function in XGBoost includes two parts: the loss function, which measures the difference between predicted and true values, and a regularization term, which manages model complexity to reduce overfitting.

$$\text{Obj}(\Theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (29)$$

where  $\Theta$  represents the model parameters,  $L(y_i, \hat{y}_i)$  is the loss function measuring the difference between the true label  $y_i$  and the predicted label  $\hat{y}_i$ , and  $\Omega(f_k)$  is the regularization term for the  $k$ -th tree  $f_k$ .

Regarding the loss function, XGBoost typically uses the mean squared error (MSE):

$$L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (30)$$

where  $\hat{y}_i$  is the predicted probability of class 1.

Finally, in order for the models to avoid overfitting, the regularization term consists of two parts: the tree complexity term and the leaf weight regularization term. It can be written as:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|_2^2 \quad (31)$$

where  $T$  is the number of leaves in the tree  $f_k$ , and  $w$  are the weights associated with the leaf nodes [15].

### 3.2.6. Random Forest

Random forest (RF) is an ensemble learning algorithm which is a combination of multiple decision trees. This model generates a set of decision trees during training and produces the final output by taking the mode of the average predictions from each individual tree. Each decision tree is trained on a random subset of the training data, and at each split in a tree, a random subset of features is considered. This randomness helps in making the model robust and reducing variance.

Mathematically, the prediction  $\hat{y}$  of a random forest model is given by [16]:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x) \quad (32)$$

where  $T$  is the number of trees in the forest,  $f_t(x)$  is the prediction of the  $t$ -th tree for the input  $x$ . For classification, the mode of the class labels predicted by each tree is taken, while for regression, the average of all tree predictions is used.

### 3.3. Deep Learning Models

Similar to ML algorithms, the most commonly used DL models are the presented below.

#### 3.3.1. Multilayer Perceptron

The multilayer perceptron (MLP), which is presented in Figure 5, is a simple feed-forward neural network (FNN) model that consists of one input layer of nodes, hidden layers and one output layer, which creates the final prediction. Each node in each layer is connected to every node in the previous and next layer through neurons which correspond to synaptic weights, which are used for the transmission of information across the model. An activation function, such as the rectified linear unit (ReLU), is used in order to calculate the output of each node.

Given an input vector  $x$ , the output of a node  $j$  in the hidden layer  $h$  can be calculated as follows:

$$z_j^h = \sum_{i=1}^{n_{\text{input}}} w_{ij}^h x_i + b_j^h \quad (33)$$

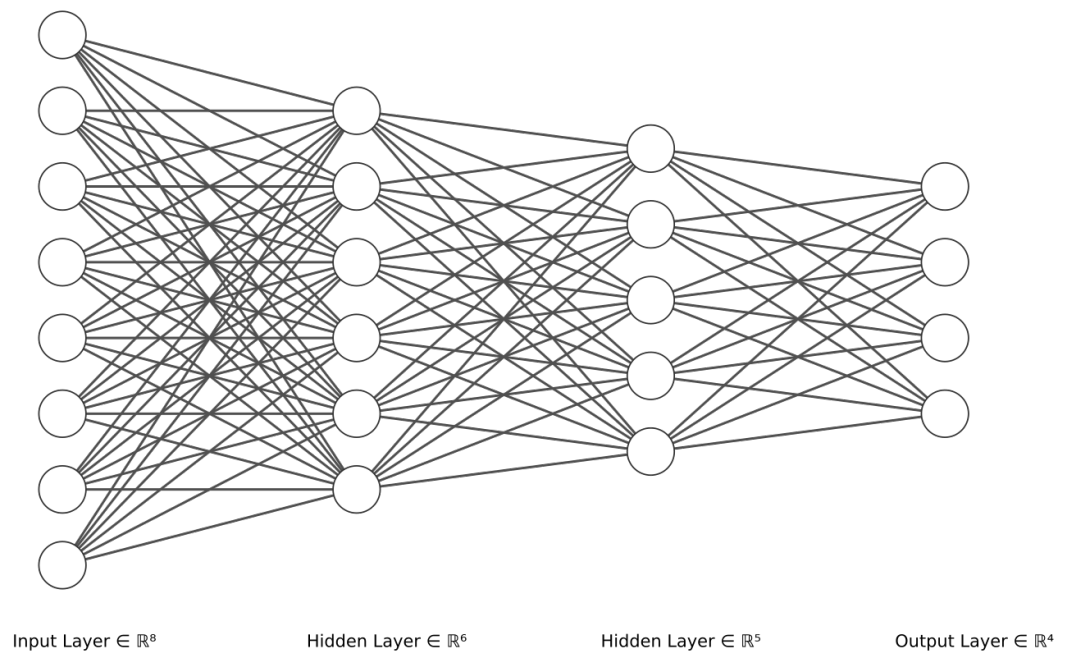
$$a_j^h = f(z_j^h) \quad (34)$$

where  $n_{\text{input}}$  is the number of nodes in the input layer,  $w_{ij}^h$  is the weight of the connection between the  $i$ -th node in the input layer and the  $j$ -th node in the hidden layer,  $b_j^h$  is the bias of the  $j$ -th node in the hidden layer, and  $f$  is the activation function.

Similarly, the output of a node  $k$  in the output layer can be computed as:

$$z_k^{\text{output}} = \sum_{j=1}^{n_{\text{hidden}}} w_{jk}^{\text{output}} a_j^h + b_k^{\text{output}} \quad (35)$$

$$a_k^{\text{output}} = f(z_k^{\text{output}}) \quad (36)$$



**Figure 5.** Multilayer Perceptron.

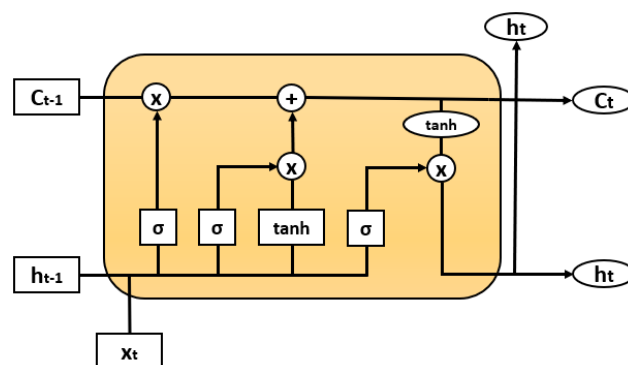
Regarding the training of the MLP model, it is carried out through the backpropagation method, in which the error between the actual and forecasting values is minimized by adjusting the weights and biases using gradient descent. The loss function used is based on each different task. Usually, mean squared error is used for time series forecasting and cross-entropy loss for classification tasks [17].

### 3.3.2. Long Short-Term Memory

Long short-term memory (LSTM) is a type of recurrent-based architecture, more suitable to manage sequential data, like energy consumption, and capture long-term dependencies between them from RNNs. LSTM networks, unlike regular RNNs, are capable of maintaining information for extended periods, which helps solve the vanishing gradient problem that often occurs in simple RNNs.

LSTM networks are composed of LSTM cells, whose main structure is presented in Figure 6 and has three main gates:

- Forget Gate: Determines what information to discard from the cell state.
- Input Gate: Selects which new information to store in the cell state.
- Output Gate: Chooses what part of the cell state proceeds to the output.



**Figure 6.** LSTM cell structure [18].

The equations for the LSTM cell are as follows:

Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (37)$$

Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (38)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (39)$$

Cell State Update

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (40)$$

Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (41)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (42)$$

where  $x_t$  is the input at time step  $t$ ;  $h_{t-1}$  represents the hidden state from the previous time step;  $C_{t-1}$  is the cell state from the previous time step;  $f_t$  is the forget gate activation;  $i_t$  is the input gate activation;  $\tilde{C}_t$  is the candidate cell state;  $C_t$  is the updated cell state;  $o_t$  is the output gate activation;  $h_t$  is the hidden state at time step  $t$ ;  $W_f, W_i, W_C, W_o$  are the weight matrices;  $b_f, b_i, b_C, b_o$  are the bias vectors;  $\sigma$  is the sigmoid activation function; and  $\tanh$  is the hyperbolic tangent activation function [18].

### 3.3.3. Convolutional Neural Network

A convolutional neural network (CNN) includes convolutional layers, pooling layers, and fully connected layers. More specifically, its structure is presented in Figure 7 and includes the following.

Convolutional Layer

In this layer, convolution operations are applied to the input data through a series of learnable filters (also known as kernels). Each filter slides across the input, generating a feature map. The convolution process is defined as:

$$(f * x)(i, j) = \sum_m \sum_n x(i + m, j + n) \cdot f(m, n) \quad (43)$$

where  $x$  is the input,  $f$  is the filter, and  $*$  denotes convolution.

Pooling Layer

This layer reduces the spatial dimensions (height and width) of the feature maps, helping to decrease the number of parameters and computational load in the network. One common pooling operation is max pooling, which is defined as:

$$y(i, j) = \max_{m, n} x(i + m, j + n) \quad (44)$$

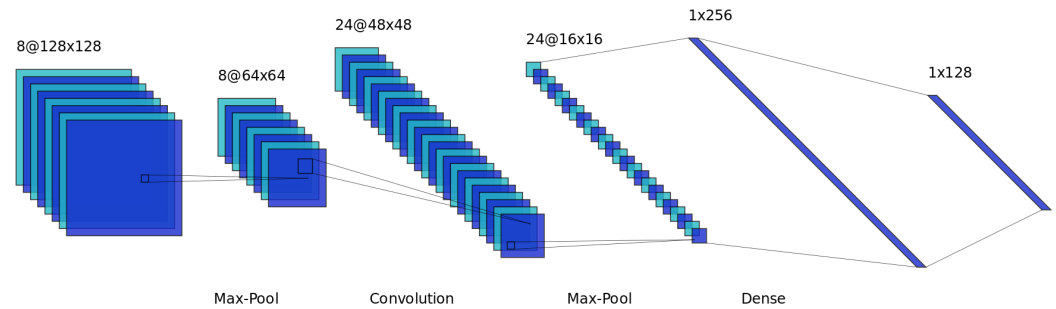
where  $x$  is the input and  $y$  is the pooled output.

Final Fully Connected Layer

The fully connected layer creates the final prediction. Each neuron of this layer is connected to every neuron in the previous layer. The output is computed as:

$$y = W \cdot x + b \quad (45)$$

where  $W$  is the weight matrix,  $x$  is the input vector, and  $b$  is the bias vector [18].



**Figure 7.** Convolutional neural network.

### 3.3.4. Transformer Models

Transformer models have revolutionized time series forecasting with their ability to handle complex dependencies through self-attention mechanisms. They use an encoder–decoder architecture to capture relationships in sequential data effectively.

The basic transformer architecture includes the following components [18]:

- Encoder: Processes input sequences.
- Decoder: Generates output sequences from encoded representations.

The core component is the self-attention mechanism, which computes the attention scores using the following equations:

Scaled Dot-Product Attention:

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

where  $Q$  is the query matrix,  $K$  is the key matrix,  $V$  is the value matrix, and  $d_k$  is the dimension of the key vectors.

Multi-Head Attention:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O$$

where each head is computed as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

and  $W_i^Q, W_i^K, W_i^V$  are parameter matrices for the  $i$ -th attention head, and  $W^O$  is the output matrix.

### 3.4. Ensemble Methods

The primary concept underlying ensemble methods is to capitalize on the advantages of different algorithms and combine their ultimate predictions, resulting in enhanced accuracy and robustness. These models can be applied to a variety of tasks, including statistical, as well as ML and DL models for classification, regression, and clustering tasks.

A key benefit of these algorithms is their capacity to decrease overfitting, which leads to bad forecasting performance and variance, which represents the level of fluctuation in predictions across different models within the ensemble, respectively. By combining the forecasts of several models, ensemble methods help minimize variance and produce more reliable predictions.

Regarding the disadvantages, they are typically more complicated to implement and interpret compared to individual models, and in most cases they require significantly more computational power. Their performance is highly dependent on the combination of models used, which makes identifying the best set of models a difficult task. Additionally, the ensemble's method quality is pivotal due to the fact that poorly performing models can negatively affect the ensemble final prediction, leading to issues such as overfitting or underfitting.



The most common techniques used to create ensemble models are bagging, boosting, and stacking, which are analyzed below [19].

#### 3.4.1. Bagging

Bagging consists of training each model individually and then combining their predictions using a weighted average. This approach can be used with any type of model and is effective at reducing variance and preventing overfitting. Initially, several iterations of the dataset are executed, each containing distinct subsets of the initial data. After, these subsets are used to train models separately, and finally their predictions are integrated to produce the final output. By merging predictions from different models in a weighted average process, bagging improves both the accuracy and reliability of the final predictions [18,19].

#### 3.4.2. Boosting

Boosting is a technique where each model is trained to correct the errors in predictive data analysis. Models are trained using labeled data to make predictions about unlabeled data. This method improves the accuracy of the final predictions but is more vulnerable to overfitting in comparison to bagging method. Boosting begins with a simple model and gradually adds more complex models, each focusing on the data points that were wrongly labeled by previous models. Each of the different forecasts is merged using a weighted combination based on their final accuracy [18,19].

#### 3.4.3. Stacking

Stacking, also called stacked generalization, involves combining the predictions from various base models—sometimes known as first-level models or base learners—to produce a conclusive prediction. This approach entails training multiple base models on an identical dataset and utilizing their output as input for a higher-level model, termed a meta-model or second-level model, to produce the final outcome. The fundamental concept is to improve predictive accuracy by combining the outputs of various base models, rather than depending on a single model [18,19].

### 3.5. Most Commonly Used Evaluation Metrics

In terms of related work, the prediction metrics that are most frequently employed include the following [20–22]:

Mean Absolute Error (MAE): Computes the mean value of the absolute differences between the predicted values and the actual values.

Mean Absolute Percentage Error (MAPE): Calculates the average of the absolute percentage deviations between the predicted and actual values.

Mean Squared Error (MSE): Measures the average squared difference between the predicted values and the actual values.

Root Mean Squared Error (RMSE): Calculates the square root of the mean of the squared differences between the forecasted and actual values.

R-squared ( $R^2$ ): This statistical measure evaluates the proportion of variance in the dependent variable that is explained by the independent variable(s) in a forecasting model, with values spanning from 0 to 1. A value of 1 indicates an excellent fit, whereas a value of 0 signifies no relationship between the variables.

The above metrics are defined with the following equations:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (46)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right| \quad (47)$$

$$MSE = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n} \quad (48)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - x_i)^2}{n}} \quad (49)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{x}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (50)$$

where  $x_i$ ,  $y_i$ ,  $\hat{x}_i$ , and  $\bar{y}$  are the forecasting values, the actual values, the mean of the forecasting values, and the mean of the actual values, respectively.

Apart from the above metrics, it is observed that some papers additionally use the following: Mean Arctangent Absolute Percentage Error (MAAPE), Mean Percentage Error (MPE), Error Sum of Squares (SSE), Sum of Squares (SSE), Mean of Mean Squared Error (FMSE), Symmetric Mean Absolute Percentage Error (sMAPE), Median Absolute Error (MedAE), Matthews Correlation Coefficient (MCC), Relative Root Mean Square (rRMSE), Normalized Root Mean Square Error (NRMSE), Coefficient of Variation (CV), and Accuracy, which is measured by the proportion of correct predictions out of the total number of predictions.

#### 4. Challenges and Solutions for Electricity Load Forecasting

In this section, important studies that have significantly contributed to the field of load forecasting are extensively analyzed. Initially, statistical models are presented, followed by papers focusing on ML and DL, and finally, emphasis is given to ensemble methods. Because most studies typically present models from both the ML and DL categories, it was chosen to present them in a common subsection for a more substantial presentation of the related works.

##### 4.1. Statistical Models

In the field of ELF, several methodologies and techniques have been investigated. These can be separated into two basic groups, the statistical and the modern models. Statistical models include models like AR, ARMA, ARIMA, and models with exogenous features, like ARMAX and ARIMAX, grey (GM) and ES. In this way, Sethi et al. in [23] compare the ARIMA method with other ML models for load prediction. Evaluation is performed using RMSE and Mean Bias Error (MBE), and as a case study, a commercial building (amusement park) in San Diego is used. Chou et al. in [24] conduct a comparative analysis using the SARIMA model. The study uses real-time consumption data collected from smart grid infrastructure that is connected to a three-floor building in the Xindian District, New Taipei City, Taiwan. Several metrics, including R-squared, Maximum Absolute Error (MaxAE), RMSE, MAE and MAPE are used for the evaluation process. Hadri et al. in [25] conduct a comprehensive research employing ARIMA, SARIMA, XGBoost, random forest (RF), and LSTM. For the evaluation period, Mean Percentage Error (MPE), MAPE, MAE, and RMSE are utilized. Simultaneously, Vrabecová et al. in [26] apply ARIMA models in order to create short-term forecasts on the public Irish CER dataset.

Some papers conduct comprehensive analysis of various statistical models used for ELF in different European countries. Following the above, Pelka et al. in [27] investigate the performance of different statistical models, including ARIMA, ETS, and the Prophet model. The target is to forecast monthly power demand, which estimates the correlation between past and future demand trends. As a case study, electricity time series data for 35 European countries in a monthly resolution are used. In addition, Elamin et al. in [28] introduce a forecasting approach, utilizing a SARIMA model, incorporating both exogenous and interaction variables, to predict hourly load electricity demand. Finally, Mado et al. in [29] smooth the training and testing datasets of the Java–Madura–Bali interconnection in Indonesia and propose a SARIMA model for short-term ELF, which considers double seasonal patterns in electricity consumption. They achieve MAPE equal to 1.56%.

#### 4.2. Machine and Deep Learning Models

The ELF sector has undergone high transition towards the use of ML and DL models, which offer superior accuracy and adaptability compared to traditional statistical models. ML models are capable of handling with complex non-linear relationships in electricity consumption data. Meanwhile, DL models, including RNNs, LSTMs, and CNNs, excel in handling large volumes of data and uncovering intricate temporal patterns. These advancements play a critical role for improving the performance and sustainability of modern electricity systems, enabling better demand management, grid stability, and integration of RES. According to the above, Abumohsen et al. in [30] utilize several DL models, such as RNN, GRU and LSTM, in order to predict electricity consumption based on current measurements data. Forecasting models are crucial for reducing costs, optimizing resources, and improving load distribution for electric companies. Among the models tested, the GRU exhibited superior performance, with MSE equal to 0.00215, MAE of 0.03266 and R-squared equal to 90.228%. Li et al. in [31] introduce a model named TS2ARCformer. This model utilizes multiple dimensions, demonstrating significant forecasting performance, in comparison with to benchmark models, like LSTM and GRU. The results show reductions of 43.2% and 37.8% in the aspect of MAPE.

Additionally, the model proposed in [32] utilizes neural network techniques enhanced with PSO to assess the Iranian power system. This approach, based on neural networks, showed lower prediction errors by effectively adapting to the underlying patterns of the consumption load. The evaluation was performed using the RMSE metric, which was equal to 0.028, 0.045, 0.035, and 0.040 for multiple lengths of training input data. In the studies [33,34], empirical modal decomposition (EMD) was used in order to break down the original electricity consumption time series data into various intrinsic mode functions (IMFs) with different frequencies and amplitudes.

Also, in another study, a hybrid prediction model is proposed. This model combines a temporal convolutional network (TCN) with variational mode decomposition (VMD). For better accuracy, an error correction process is followed. The weighted permutation entropy process is followed in order to determine the optimal decomposition number and penalty factor for VMD. The accuracy and performance of the proposed model are demonstrated using the Global Energy Competition 2014 dataset [35].

In the same way, Javed et al. in [36] research the short-term load forecasting problem, comparing eight state-of-the-art linear and non-linear models, which are ARX, KNN, AR-MAX, bagged trees, SVM, OE, NN-PSO, and ANN-LM with a single hidden dense layer and ANN-LM with two hidden dense layers, on real-time demand time series. Simultaneously, He et al., in [37], compare individual and decomposition-based forecasting methods. Regarding the the individual models, these include LSTM, SVR, MLP, LR, and random forest regressor. The decomposition-based methods include empirical mode decomposition-LSTM (EMD-LSTM) and ensemble EEMD-LSTM models.

In additional to the above, He et al. in [38] introduce a novel DNN architecture for short-term load forecasting. The model integrates various types parameters as input by employing appropriate NN elements designed to process each feature effectively. CNN components are used to extract features and patterns from the historical load time series, while RNN components are used to model the implicit dynamics. Additionally, dense layers are employed to transform other types of features, improving the generalization of the models in understanding complex trends and patterns. Shi et al. in [39] suggest a model to tackle the unpredictability in electrical load patterns by applying DL and ML techniques to forecast household energy use in Ireland. This research introduces an innovative pooling-based deep RNN (PDRNN), which combines various load curves of consumers into a comprehensive set of inputs. The performance of the PDRNN is compared with ARIMA, SVR, and RNN models. The results demonstrate that the PDRNN method outperforms the other models, with RMSE values (in kWh) of 0.4505 for PDRNN compared to 0.5593 for ARIMA, 0.518 for SVR, and 0.528 for RNN. Amarasinghe et al. in [40] suggested a framework to forecast short-term electricity consumption in residences

by utilizing smart meter data from various properties. Following this, Wand et al. in [41] propose an optimized SVM model using genetic algorithm (GA-SVM) in order to predict the hourly electricity demand of a commercial center located in Xin town, China.

Residential forecasting is an area that gathers a lot of scientific interest. Bessani et al. in [42] study the short-term residential forecasting task across multiple households employing Bayesian multivariate networks, which provide a structured, probabilistic approach to model complex dependencies between multiple variables, enabling efficient inference and decision-making even with incomplete data. This algorithm predicts the forthcoming household demand by analyzing historical consumption patterns, temperature, humidity, socioeconomic variables, and energy use. The assessment utilized actual data sourced from the Irish Intelligent Meter Project. According to the results, the proposed technique provides a robust single forecast model, which is capable of accommodating hundreds of households with various consumption patterns. The reported metrics include an MAE of 1.0085 kWh and a Mean Arctangent Absolute Percentage Error (MAAPE) of 0.5035, highlighting its effectiveness in accurate load forecasting.

Regarding RNN-based approaches, Caicedo-Vivas et al. in [43] analyze the issue of short-term load prediction by employing dataset from a Colombian grid operator. The proposed DL model utilizes a LSTM to predict load across a seven-day time horizon (168 h ahead). It combines historical data from a specific region in Colombia with calendar attributes like holidays and the month's index relevant to the forecasting week. Unlike traditional LSTM approaches, this model enables LSTM cells to process multiple load measurements concurrently, with supplementary information (holidays and current month) appended to the LSTM output. This methodology improves the model's ability to capture and utilize temporal and contextual factors for more accurate load forecasts. Also, Kong et al. in [44] conduct a comparative study between the LSTM model and the extreme learning machine (ELM), back-propagation NN (BPNN), and KNN. They demonstrated significant reduction in prediction errors by employing the LSTM structure. The aggregated average MAPE for their forecasts was 8.18%, while the average MAPE for individual forecasts was 44.39%. This study underscores the rich performance of LSTM models in short-term residential forecasting.

Following the above, Gao et al. in [33] investigate the impact of decomposition methods, which simplify complex data by breaking it down into underlying components, making it easier to analyze patterns, reduce dimensionality, and improve model interpretability. For this reason, the authors propose an EMD gated recurrent unit with feature selection (EMD-GRU-FS) for short-term electricity forecasting, using Pearson correlation to identify the relationship between subseries and the original series. Experimental results indicated that this method achieved average prediction accuracies of 96.9%, 95.31%, 95.72%, and 97.17% across four datasets. Additionally, Yuan et al. in [34] present an enhanced model combining integrated EMD and LSTM for short-term load forecasting, where the recurrent model is used for feature extraction and creating load predictions. For the final consumers, short-term electricity consumption predictions were obtained by aggregating multiple target prediction results. The results show that the proposed model performs achieving MAPE 2.6249% in winter and 2.3047% in summer, respectively.

Some studies integrate Internet of Things (IoT) solutions in order to enhance ELF and improve localized forecasting by gathering detailed, real-time data from devices like smart meters and local sensors. IoT-driven insights can optimize energy distribution, minimize supply mismatches, and support the integration of renewable sources, leading to a more reliable and efficient power grid. Based on these, Imani et al. in [45] introduce a clustering forecasting approach which divides the consumers' consumption based on two different features, collecting their data from an IoT connection from smart meters. As a case study, an Irish dataset is used and the MLP model is proposed. Additionally, Hadri et al. in [25] developed load forecasting techniques that can be integrated with occupancy predictions and context-aware control of building appliances. The researchers investigated the accuracy of multiple methods for predicting energy consumption. They deployed an

IoT and big data-based platform to collect real-time electricity values. The acquired data were used to build predictive models using XGBoost, RF, ARIMA, SARIMA and LSTM. The models' evaluation was performed using metrics such as RMSE, MAE, MAPE, R and Maximum Absolute Error (MaxAE). Finally, Raju et al. in [46] introduce an advanced and novel IoT method utilizing machine learning algorithms. LR, SVM, Gaussian Processes (GPs) and ensemble methods are used and MAE, MSE and RMSE metrics are measured in order to draw useful conclusions for a smart home in which an Arduino Uno system has been installed.

Other works focus on forecasting the energy usage of residential consumers. Accurate forecasting of residential consumers' load is critical because it enables efficient energy management, reduces operational costs, and improves grid stability. Based on these, in another study, the authors conduct a comparative study between ARIMA and SARIMA, as well as RF, XGBoost and LSTM models using MAPE, MAPE, Mean Percentage Error (MPE) and RMSE [25]. Li et al. in [47] developed a novel short-term forecasting model based on a modified LSTM model and an autoregressive model, which is used for feature selection. The aim is to create forecasts for household electricity consumption.

Other studies explore how advanced techniques can be built into deep learning models to make their conclusions more accurate and reliable. Online learning is also essential in this context because it enables models to update continually as new data arrive. This constant updating allows the system to adjust to real-time changes in the market, adapting its responses to reflect current conditions. This method, known as online or continual learning, supports accurate, timely insights that are crucial for businesses operating in fast-changing, data-driven settings. Following these, Fekri et al. in [48] introduced an online adaptive RNN, which is capable of continuous learning from recently acquired data and adaptation to various patterns.

Regarding CNN-based models, which have proven effective for time series forecasting by capturing local patterns and dependencies within sequential data, Andriopoulos et al. in [49] introduce an advanced CNN model that concentrates on mathematical analysis of statistical data and advanced preprocessing of time series data for short-term load forecasting. This approach converts the data into a 2-D image format to leverage CNN-based models, which excel in handling stationarity and temporal volatility. The proposed model is compared with LSTM, considered the state-of-the-art solution for load forecasting. Following CNN models, Alanazi et al. in [50] present a novel approach, combining a residual CNN model and a layered echo state network (ESN), in order to handle efficiently both spatial and temporal patterns. Residential datasets from Individual-Household-Electric-Power-Consumption (IHEPC) from Sceaux, Paris, and the Pennsylvania—New Jersey—Maryland (PJM) region are used. The developed model achieves MSE, MAE, and RMSE equal to 0.0713, 0.1978, and 0.2670 MW, respectively. Also, Jalali et al. in [51] study and propose a novel approach for short-term ELF using a deep neuroevolution algorithm in order to develop a CNN model in an automatic way, using an enhanced grey wolf optimizer (EGWO) as a modified advanced algorithm. Finally, Chitalia et al. in [52] present a strong forecasting framework in order to accommodate fluctuations in building operations across different types and locations of commercial buildings.

Another promising method used in load prediction studies is graph neural networks (GNNs), which capture dependencies within graph-structured data [53]. For instance, Vontzos et al. [54] used a GNN-RNN algorithm to forecast the consumption of a multi-story educational building that exhibits in a graph the structure of the building zones with multiply methodologies. Additionally, Hu et al.'s [55] study incorporates an urban building energy model that synthesizes the solar-based building interdependency and spatio-temporal graph convolutional network (ST-GCN) algorithm to predict hourly energy consumption of a university campus. In particular, we took a university campus located in the downtown area of Atlanta as an example to predict the hourly energy consumption. In terms of regional load forecasting, Wei et al. [56] propose a method which incorporates limited dynamic time warping (LDTW) and graph convolutional networks.



Finally, in many studies, weather data like temperature and humidity play a crucial role in ELF. In this way, Liao et al. in [57] propose the utilization of the XGBoost model, using weekly historical data from a power plant in hourly resolution, incorporating weather factors such as temperature and humidity. The work additionally addresses the computational challenges due to the complexity of the XGBoost hyperparameter optimization process.

#### 4.3. Ensemble Methods

Ensemble models have gained interest because they improve predictive performance by combining multiple algorithms and minimize individual models' disadvantages. These models improve forecasting accuracy and generalization, making them effective for complex datasets, reducing overfitting and variance. Advances in computational power and accessible tools have made implementing ensemble methods easier, boosting their popularity.

Advanced combinations of different models are followed in ensemble learning methods. In this way, Guo et al. in [58] propose a novel approach to forecasting wind speed using a modified EMD-based feed-forward neural network (FNN). The advantage of the proposed model is based on the fact that it is capable of addressing the nonlinear and non-stationary nature of wind speed data by decomposing the time series into multiple components and a residual series, using EMD techniques. The data used are from Zhangye, China. Alobaidi et al. in [59] propose the implementation of an ensemble learning framework to tackle the issue of forecasting the daily mean demand on an individual household level for the day-ahead market. The proposed model incorporates a regression network, making the final model perform with very high accuracy on a dataset from France. Du et al. in [60] introduce an advanced hybrid algorithm that effectively retrieves accurate wind speed series using singular spectrum analysis. This algorithm demonstrates superior adaptive search and optimization capabilities compared to other methods, offering faster performance, fewer parameters, and lower computational cost. As a case study, this work applies the model to wind speed data in 10 min resolution from three wind farms located in Shandong Province, eastern China.

Simultaneously and following fuzzy logic, Ma et al. in [61] combine a denoising technique with a dynamic fuzzy neural network to meet the challenges in wind speed forecasting tasks. The raw data are smoothed using singular spectrum analysis, optimized by brain storm optimization, to produce a cleaner dataset. A generalized dynamic fuzzy neural network is then used for forecasting. The model's simpler and more compact neural network structure allows for faster learning and accurate predictions. Experimental results on wind speed data at 10 min, 30 min, and 60 min intervals show that the model effectively approximates actual values and can serve as a reliable tool for smart grid planning. Also, Sadaei et al. in [62] combine fuzzy time series (FTS) and CNN models for short-term ELF. Multivariate consumption data, temperature and a fuzzified version of load time series in hourly resolution were converted into multi-channel images in order to be used as input in the proposed model.

At the same way, Rafi et al. in [63] combine CNN and LSTM networks. As a case study, the proposed method is applied to data from Bangladesh's electricity network in order to provide short-term forecasting. Finally, Wang et al. in [64] incorporate Q-learning with reinforcement learning. The aim for the final ensemble model is the creation of optimal ensemble weights, achieving predictions with the highest accuracy. As a case study, various datasets of 25 households in the Austin area of the United States are used. The evaluation metrics show that the ensemble model surpasses three single models in the testing period.

In summary, we observe that when combining predictions from strong models in ensemble algorithms, the final performance improves even further. Additionally, a significant advantage of ensemble methods over single models is that the potential poor performance of any individual model is minimized in the final outcome, leading to more reliable and stable results.

#### 4.4. Aggregated Studies for Electricity Load Forecasting

In order to present the studies analyzed above in a consolidated manner, Table 1 is provided.

**Table 1.** Aggregated table with resent publications on load forecasting.

Ref.	Year	Models	Metrics	Dataset Origin
[23]	2020	ARIMA, MLR, RPART, CTREE, BAGGING, RF	MAPE, R2	Electricity load data of a residential area
[24]	2018	ANN, SVM, RL, CR Tree, Voting, Bagging, SARIMA-MetaFA-LSSVR, SARIMA-PDO-LSSVR	R2, RMSE, MAPE, MaxAE	Real-time data collected from a smart grid installed in an experimental building
[25]	2019	ARIMA, SARIMA, XGBoost, Random Forest (RF), Long Short-Term Memory (LSTM)	sMAPE, MAPE, MAE, RMSE	N/A
[26]	2018	Online SVR, MLP, ARIMA	MAPE	Public Irish CER dataset
[27]	2023	ARIMA, ETS, Prophet	MAPE	ENTSOE
[28]	2018	SARIMAX	RMSE, MAE, MAPE	Tokyo Electric Power Company (TEPCO)
[65]	2020	ARIMA, ARMA, SARIMAX	MSE, MAPE, MAE	Various datasets
[66]	2016	ANN, LR, MLR, Fuzzy Regression, ARIMA, ARMA, GARCH	Various Metrics	US utility
[29]	2022	DSARIMA	MAPE	Java-Madura-Bali interconnection, Indonesia
[30]	2023	LSTM, GRU, RNN	R-squared, MSE, MAE	Tubas Electricity Company, Palestine
[67]	2017	DNN, PSO	MAE, MSE	North Cyprus based energy company, Cyprus Turkish Electricity Authority (Kib-Tek)
[33]	2019	GRU, SVR, RF, EMD-GRU, EMD-SVR, EMD-RF	MAPE, RMSE	U.S. Public Utilities Data Set, Region in Northwest China
[34]	2022	RNN, EMD, LSTM, EMD-LSTM	MAPE, RMSE	Region in Guangdong Province, China
[37]	2019	LSTM, SVR, MLPR, LR, RFR, EMD-LSTM, EEMD-LSTM	RMSE, MAE, MAPE (%), R2	Hubei Province, China
[35]	2022	TCN, VMD, LSTM, GRU, VMD-LSTM, VMD-GRU, VMD-TCN	RMSE, MAPE, R2	Global Energy Competition 2014 dataset
[38]	2017	LR, SVR, DNN, CNN	MAPE, MAE	Hourly loads of a North China city
[39]	2018	ARIMA, SVR, RNN, DRNN, PDRNN	RMSE, MAE, NRMSE	920 smart metered customers from Ireland
[40]	2017	CNN, LSTM, SVM, ANN	RMSE	Benchmark dataset of electricity consumption for a single residential customer
[42]	2020	Bayesian networks	NRMSE, MAE, MedAE, MAAPE	Real data from the Irish smart meter project
[44]	2019	LSTM, BPNN, KNN, ELM	MAPE	Real residential smart meter data



Table 1. Cont.

Ref.	Year	Models	Metrics	Dataset Origin
[43]	2023	LSTM, XGBOOST	MAPE	Historical load values from a region in Colombia
[36]	2021	OE, ARX, ARMAX, SVM, ANN-PSO, KNN, ANN-LM	RMSE, MAPE, MAE	Real-time electrical load and meteorological data of the city of Lahore in Pakistan
[49]	2021	CNN, LSTM, MLP, ANN	MAE, RMSE	Three open-source, publicly available datasets
[57]	2019	XGBOOST, LSTM	MSE, RMSE, MAPE	State Grid Corporation
[47]	2021	Persistence, SW-ARIMA, SW-ETS, SW-SVR, ConvLSTM, CLSAF	Accuracy	Manhattan, New York, residential building
[48]	2021	MLP, KNN, RNN	MSE, MAE	Five residential consumers provided by London Hydro
[51]	2021	CNN, EGWO, LSTM, SVR, MLP	RMSE, MAE, MAPE	Australian Energy Market Operator
[52]	2020	LSTM, BiLSTM	RMSE, MAPE, CV	Bangkok, Thailand; Hyderabad, India; Virginia, USA; New York, USA; and Massachusetts, USA
[31]	2023	LSTM, GRU, TS2ARCformer, TCN	MAPE, MSE, MAE	National College Student Mathematical Contest in Electrical Engineering
[58]	2012	EMD-FNN	MSE, MAE, MAPE	Zhangye, China
[59]	2018	IANN, Bagging Ensemble Model	RMSE, MAPE, rRMSE	Individual household energy consumption
[60]	2016	XNN, GRNN, BPNN, PSO-BP, APSOSA-BP	AE, MAE, MAPE, RMSE	Wind farms in Shandong Province, eastern China
[61]	2017	FNN, SSA-GDFNN, BSO-SSA-GDFNN	FE, MAPE, MAE, RMSE	Historical short-term wind speed data collected from Shandong Peninsula of China
[62]	2019	A hybrid model of CNN and FTS	MAPE	N/A
[63]	2021	CNN, LSTM, CNN-LSTM, XGBOOST, RBFN	MAE, RMSE, MAPE	Bangladesh power system
[64]	2023	HI-TCN, HI-ELM, HI-GRU, HI-BPNN	MAE, MAPE, RMSE	Pecan Street datasets contain the load power data of 25 households in the Austin area of the United States
[50]	2023	CNN, CNN-LSTM, CNN-GRU, ESN	MSE, MAE, RMSE, MAPE, NRMSE	Individual-Household-Electric-Power-Consumption (IHEPC) dataset from residential houses in Sceaux, Paris, and Pennsylvania–New Jersey–Maryland (PJM)
[41]	2023	SVM, GA-SVM	MAPE, CV-RMSE, R2	Data of a large public building in Xi town, China
[45]	2018	MLP	MAPE	Irish social science data archive (ISSDA)

Table 1. Cont.

Ref.	Year	Models	Metrics	Dataset Origin
[46]	2020	LR, SVM, Fine Tree (FT) regression, Ensemble Bagged Regressor, Gaussian Process Regressor, Ensemble Boosted Regressor	RMSE, MSE, MAE	Domestic data collected from IoT system
[54]	2024	GCN-LSTM, MLP, CNN, LSTM	MAE, MSE, CV-RMSE, $R^2$	Academic building in Bangkok, Thailand
[55]	2022	Average, LR, MLP, XGBoost, GRU, ST-GCN	RMSE, MAPE	University campus in Atlanta, Georgia
[56]	2024	SVR, LSTM, LSTM-Att, TCN, Informer, GCN-LSTM	MAPE, RMSE, $R^2$	Substation data in area of Shanghai

## 5. Challenges and Solutions for Electricity Price Forecasting

Similar to ELF methods, in this section, studies that have significantly contributed to the field of price forecasting are extensively analyzed. Initially, statistical models are presented, followed by papers focusing on ML and DL, and finally, emphasis is given to ensemble methods. Similar to ELF, for a more substantial presentation of the related works, the related studies utilize ML and DL models are presented together.

### 5.1. Statistical Models

Similar to ELF approaches, the statistical models used to predict electricity prices rely on similar algorithms. Initially, traditional time series forecasting methods, like simple regression techniques, moving averages and smoothing methods, were employed. Then, more complex and advanced models like dynamic regression models, Box–Jenkins, and hybrid statistical models [68] to detect trends and seasonality in electricity price time series data were investigated. Various works, including [69–72], have explored ELF using the above methods. These foundational models paved the way for the creation of more powerful forecasting techniques. Skopal et al. in [69] propose a hybrid combination of ARIMA and a MLP model. The statistical model that identified linear trends, while the neural network handled the modeling of the nonlinear residuals. This approach led in lower forecast metrics, measured by the MAPE and mean absolute deviation (MAD) in comparison to the single models. Similarly, other studies, including [70,71], propose multiple autoregressive models, like SARIMA and GARCH, as well as more advanced hybrid algorithms to predict electricity prices, evaluating their performance through statistical metrics like  $R^2$  and MAPE. Their results highlight the effectiveness of these hybrid models. Additionally, a hybrid model combining ARIMA and multivariate linear regression and Holt–Winters is developed in [72]. For the evaluation process, datasets from the Iberian electricity market are used and MAPE is used. The results show that the combined model achieves satisfactory performance compared to other benchmark models. Marcjazz et al. in [73] focus on the selection of optimal calibration windows for day-ahead EPF. The research utilizes six years of data from three major power markets and involves fitting four autoregressive expert models to both raw and transformed price time series. Also, Sanchez et al. in [74] investigate an ARIMA model using multiple input data for the Pennsylvania–New Jersey–Maryland (PJM) and Spanish electricity market, in order to generate forecasts 24 h ahead.

Other studies [75–77] have focused on comparing different time series analysis methods for EPF. Abunofal et al. in [75] conduct a comparative analysis between different autoregressive models, like ARIMA, SARIMA and SARIMAX, which utilize exogenous features as input. The aim is to create forecasts for the German DAM. Between the models, the SARIMAX achieved the highest forecast accuracy. Also, Banitalebi et al. in [76] compare DES and TES for EPF from variability, utilizing elastic net regularization. The forecasted re-

sults show that TES provides exceptional performance, and regularization helps minimize the MSE. These outcomes aid in making well-informed decisions for power generation planning within the market. Additionally, various autoregressive statistical models and their derivatives have produced accurate forecasting results, as discussed in [77]. The study provides detailed computational procedures, in detail results, and evaluation metrics, like MAPE, highlighting promising outcomes and areas for further consideration.

Khairuddin et al. in [78] explore the Imbalance Cost Pass-Through (ICPT) within Malaysia's incentive-based regulation (IBR), emphasizing its role in enabling power producers to adjust tariffs in response to changing fuel prices, thereby strengthening economic resilience in electricity generation. The Energy Commission's biannual reviews of generation costs lead to either rebates or surcharges for customers. The study proposes forecasting ICPT prices to aid stakeholders in Peninsular Malaysia, utilizing three models and evaluating performance via the MAPE. Findings indicate that ARIMA is the most accurate forecasting method in comparison to MA and LSSVM, with minimal price differences observed. The forecasted ICPT prices align closely with actual tariffs, suggesting a potential price drop in the upcoming announcement. This research could positively impact Malaysia's energy sector sustainability.

GARCH approaches are advantageous in ELF tasks because they effectively capture price volatility and temporal dependencies, allowing for more accurate predictions of sudden price spikes and fluctuations in highly volatile electricity markets. Based on these, Entezari et al. in [79] examine methods to quantify risk and forecast the bias of electricity prices to revert to their historical average in Portugal's energy market. The data were used are from January 2016 to December 2021. They apply both symmetric and asymmetric GARCH models to mitigate data fluctuation peaks, with the asymmetric model capturing the distinct effects of positive and negative shocks. The MSGARCH model, estimated with two distinct states, successfully captures the shifts between low and high volatility periods, as well as the prolonged impact of shocks during high volatility. The advantage of this method is that it helps power producers predict price movements and adjust their operations with high accuracy accordingly, leading to improved market stability and efficiency. Finally, the authors in [80] research the financial impact and sustainability of a RESs. More specifically, they use a PV plant with nominal power equal to 5 MWs, which is located in Montenegro. This PV plant has direct market access instead of a long-term Power Purchase Agreement (PPA). Using ARIMA for price forecasting, it finds that this riskier strategy is more profitable and offers a shorter payback period. The study concludes that direct market access is a better option to enhance project returns and support energy transformation in developing countries.

## 5.2. Machine and Deep Learning Models

The integration of ML and DL models into wholesale EPF problems has been gaining increasing application in recent years. Gonzalez et al. in [81] examine the effectiveness of two advanced and hybrid models for forecasting next-day spot electricity prices on the APX-UK power exchange. The first model combines a fundamental supply stack model with an econometric model incorporating price-driver data. The second model extends this approach by including logistic smooth transition regression (LSTR) to handle regime changes during structural shifts. The evaluation results in the test set from both hybrid models, particularly the hybrid-LSTR, perform better than those from non-hybrid SARMA, SARMAX, and LSTR models. The models' prediction intervals (PIs) are assessed by comparing nominal to actual coverage, though no formal tests are conducted. The LSTR model achieves the best results, followed by the hybrid-ARX and SARMAX models. However, the hybrid-LSTR model tends to produce too many exceeding prices due to overly narrow PIs.

At the same time, open-source platforms provide great flexibility for such problems. Medina et al. in [82] investigate the application of advanced GPT tools, such as OpenAI's ChatGPT, to improve electricity price prediction in Spain's complex energy market. By

analyzing energy news and expert reports, they generate supplementary variables that enhance predictive models. Their research includes two methodologies: employing in-context example prompts and fine-tuning GPT models. The results show that insights derived from GPT models closely match post-publication price changes, indicating that these methods can significantly boost prediction accuracy and aid market stakeholders and companies in making informed decisions amidst electricity price volatility. Simultaneously, ref. [67] investigates the problem of short-term ELF using PSO methodologies in order to optimize back propagation feedforward networks. The results comparison is based on MAE and MSE, achieving 8.69% and 1.62%, respectively.

Some works carry out extensive comparisons between ML and DL models. Conte et al. in [83] explore forecasting techniques to enhance sustainable electricity management in Brazil by predicting critical variables such as the price of differences settlement (PLD) and wind speed. They employ two approaches: the first combines deep neural networks (LSTM and MLP) optimized with a canonical genetic algorithm (GA), and the second uses machine committees. One committee includes MLP, decision tree, linear regression, and SVM, while the other consists of MLP, LSTM, SVM, and ARIMA. Their findings indicate that the hybrid GA + LSTM algorithm performs exceptionally well in PLD prediction, achieving a mean square error (MSE) of 4.68, and in wind speed forecasting, with an MSE of 1.26. These methodologies aim to enhance decision-making in Brazil's electricity market.

Transformer models are gaining more and more interest in EPF sector, mainly due to their attention mechanisms, which allow them to focus selectively on relevant time steps, enables parallelization and capture intricate temporal dependencies in sequential data. Laitos et al. in [84] utilize advanced DL algorithms, implementing four different novel models: a CNN incorporating an attention mechanism, a hybrid CNN-GRU with attention, a soft voting ensemble, and a stacking ensemble model were introduced. Additionally, they developed an optimized transformer model known as the multi-head attention model and the perceptron model is selected as benchmark. The results reveal substantial predictive accuracy, particularly with the hybrid CNN-GRU model, which achieved a MAPE of 6.333%. The soft voting ensemble and multi-head attention models also demonstrated strong performance, with MAPEs of 6.125% and 6.889%, respectively. These findings underscore the potential of transformer models and attention mechanisms in EPF, suggesting promising directions for future research.

Additionally, the influence of conventional energy forms plays a significant role in shaping the final hourly price. Jin et al. in [85] aim to enhance electricity market price forecasting by accounting for the discrete influences of fuel costs, large-scale generator costs, and demand. They introduce two innovative techniques: one focuses on feature generation and forecasting transformed variables based on fuel cost per unit and the maximum of the probability density function (PDF) derived from kernel density estimation (KDE). The second technique involves demand decomposition followed by feature selection, verified by gain or Shapley additive explanations (SHAP) values. In their case study in the Korean electricity market, they observed improvements across all indicators. Feature generation using fuel cost per unit improved forecasting accuracy by reducing monthly volatility and error, achieving a MAPE of 3.83%. The KDE-based method achieved an MAPE of 3.82%. Combining both techniques resulted in the highest accuracy with a MAPE of 3.49%, while using decomposed data alone had a MAPE of 4.41%.

Baskan et al. in [86] explore the use of ML and DL methods to predict day-ahead electricity prices over a 7-day horizon in the German spot market, focusing on the challenge of increasing price volatility. The study underscores the potential of ML models to deliver reliable predictions in volatile and unpredictable market conditions, presenting a compelling alternative to traditional models that rely heavily on human modeling and assumptions. Through both qualitative and quantitative analysis, the research evaluates model performance based on forecast horizon and robustness determined by hyperparameters. Three distinct test scenarios are manually selected for evaluation. Various models are trained, optimized, and compared using standard performance metrics. The results

show that DL models outperform tree-based and statistical models, demonstrating superior effectiveness in managing volatile energy prices.

Other studies address the complexity of data that includes numerous block bids and offers per hour. Pinhão et al. in [87] present an innovative methodology for EPF by predicting the underlying dynamics of supply and demand curves derived from market auctions. The proposed models integrate several seasonal effects and key market variables such as wind generation and load. The evaluation metrics show the superior performance compared to benchmark methods, emphasizing the importance of integrating market bids in EPF.

Some studies examine the integration of explainable AI into deep learning models in order to deliver more precise and dependable conclusions. Heistrene et al. in [88] integrate Shapley additive explanations (SHAP) with LSTM and CNN models in order to introduce an Error Compensation EPF method. Data from the Italian power market are used, and RMSE, MAE and MAPE metrics are used. Also, Heistrene et al. in [89] introduce an explainable AI deep learning method using a CNN model, in order to enhance the predictions of electricity prices from Italy's and ERCOT's electricity markets. The main findings of these studies are the important role of feature importance, which helps identify the impact of each input feature on model predictions and the models' interpretability, which helps researchers understand the functionality of each model.

Additionally, Duo et al. in [90] introduce a day-ahead forecasting method for China's Shanxi electricity market prices, using a hybrid extreme learning machine (ELM), which randomly initializes the weights of the hidden layer and only requires the calculation of the output weights, making it significantly faster compared to traditional neural networks. Initially, they preprocess trading data by eliminating weakly correlated features using the Spearman correlation coefficient. Next, they develop a day-screening model based on grey correlation theory to filter out irrelevant samples. In order to improve the accuracy, they optimize the parameters of the regularized ELM (RELM) using the marine predators algorithm (MPA). The proposed model demonstrates superior performance compared to other existing models, helping to develop strong quotation strategies.

Finally, Saglam et al. in [91] investigate the influence of exogenous independent variables on peak load forecasting by employing multiple combinations of multilinear regression (MLR) models. Data from 1980 to 2020 are used, incorporating factors such as import/export figures, population, and gross domestic product (GDP) to predict instantaneous electricity peak load. The performance of these techniques is evaluated through MAE, MSE, MAPE, RMSE, and R-squared. The results show that the ANN and GRO models achieve the highest forecasting accuracy.

### 5.3. Ensemble Methods

Ensemble models offer various advantages, as they leverage the capabilities of both statistical algorithms and machine and deep learning. Kushagra Bhatia et al. in [92] investigate the influence the impact of RESs on price forecasts, using them in the training process of the models. The study proposes a bootstrap aggregated-stack generalized architecture for very short-term EPF, aimed at helping market participants develop real-time strategies. Guo et al. in [93] introduce an advanced ensemble model, which is called BND-FOA-LSSVM, by combining three distinct algorithms. These are Beveridge–Nelson decomposition (BND), fruit fly optimization algorithm (FOA), and least square support vector machine (LSSVM). The results show that the proposed model achieves an MAPE of 3.48%, an RMSE of CNY 11.18/MWh, and an MAE of CNY 9.95/MWh. These metrics are significantly smaller compared to those of the ARIMA, LSSVM, BND-LSSVM, FOA-LSSVM, and EMD-FOA-LSSVM models.

Yang et al. in [94] research and create a data-driven ensemble DL model (GHTnet) to capture the temporal variation of real-time electricity price values. For this reason, an advanced CNN module based on inception architecture is developed, which is called GoogLeNet. The advantage of this model is the capability of capturing the high-frequency



features of data used, while the inclusion of time series summary statistics has been shown to enhance the forecasting of volatile price spikes. Datasets from 49 generators from the NYISO are used for the evaluation of the performance of the models. The final results show that the proposed model achieves significant improvements over that of state-of-the-art benchmark methods, improving the MAPE by 17.34% approximately.

Also, Jahangir et al. in [95] propose a novel method for EPF by using grey correlation analysis to select important parameters, denoising data with deep neural networks, implementing dimension reduction, and utilizing rough ANNs. Additionally, Khajeh et al. in [96] focus on using ANNs with an improved iterative training algorithm for EPF in such markets. This paper introduces a novel forecasting approach for short-term electricity price prediction, addressing the inherent complexity, nonlinearity, volatility, and time-dependent nature of electricity prices. The proposed strategy incorporates two key innovations: a two-stage feature selection algorithm and an iterative training algorithm, both designed to enhance the accuracy and reliability of the proposed model.

Following the above, Zhang et al. in [97] introduce a novel hybrid model based on wavelet transform (WT), ARIMA and least squares SVM (LSSVM). The hyperparameters of the model are optimized by PSO algorithm, making it capable of achieving high performance in EPF tasks. The proposed method is tested using data from New South Wales (NSW) in Australia's national electricity market. The result show that this method gives more accurate and effective results compared to other EPF methods.

Finally, Jaime et al. in [98] propose a hybrid EPF algorithm that utilizes a two-dimensional approach. In the vertical dimension, the final forecasts were generated by combining the outputs of XGBoost and LSTM models in a stacked array. In the horizontal dimension, the model leveraged available market data at the time of forecasting to create price predictions ranging from 1 h to 96 h ahead. Additionally, a new input feature, derived from the slope of thermal power generation output, significantly improved the accuracy of the model.

#### 5.4. Aggregated Studies for Electricity Price Forecasting

In order to present the studies analyzed above in a consolidated manner, Table 2 follows.

**Table 2.** Aggregated table with resent publications for price forecasting.

Ref.	Year	Models	Metrics	Dataset Origin
[69]	2015	DNN, ARIMA	MAPE, MAD, MFE	Czech electricity spot prices
[70]	2013	ARMA, GARCH	RMSE, MAPE	Electricity prices in the electricity market of New England
[71]	2021	SARIMA, GARCH	MAPE, RMSE	Indian Energy Exchange (IEX)
[72]	2019	ARIMA, Holt-Winters, MLR	MAPE	Iberian electricity market
[75]	2021	ARIMA, SARIMA, SARIMAX	MAE, RMSE	EEX (European Energy Exchange)
[76]	2021	Exponential Smoothing	MSE	Ontario, Canada
[77]	2020	ARIMA, ARIMAX	MAPE	Belgorod region in European part of Russian coincides
[78]	2024	ARIMA, MA, LSSVM	MAPE	Malaysia, the Energy Commission
[81]	2012	NAIVE, SARMA, SARMAX, LSTR	MAPE	APX power exchange for Great Britain
[82]	2024	Generative pre-trained transformers (GPT)	MCC	Spanish electricity market

Table 2. Cont.

Ref.	Year	Models	Metrics	Dataset Origin
[84]	2024	CNN Attention, Multi-Head Attention, CNN-GRU Attention, Stacking Ensemble, Soft Voting Ensemble, Perceptron	MSE, RMSE, MAPE	HEDNO (Hellenic Electricity Distribution Network Operator)
[79]	2024	GARCH, MSGARCH	Coefficient, Standard Error, t-Value, p-Value	Portugal's energy market
[83]	2024	MLP, LSTM, SVM, ARIMA, GA, PLD	MSE	National electric system of Brazil
[85]	2023	XGBoost and Decomposition Techniques	MAPE, RMSE, MDE	Korean power system
[86]	2023	LSTM, CNN-LSTM, ARIMA, Random Forests, Gradient Boosted Trees, Naive, SVM, KNN, Decision Trees	MAE, RMSE	German spot market
[87]	2022	XGBOOST, Market Curves, XGBOOST-Market Curves	MAE, WMAE	OMIE (Electricity Market Operator, Spain)
[91]	2024	ANN, SVR, MLR, PSO, DO, GRO	MAE, MSE, MAPE, RMSE	Turkish Statistical Institute (TSI), World Bank Open Database
[90]	2022	CRITIC-GRAMPA-RELM, MPA-RELM, GA-ELM, GA-SVM, ELM, SVM	SSE, MSE, RMSE, MAE	Shanxi spot market
[80]	2022	NNAR, ARIMA	ME, RMSE, MAE, MPE, MAPE	Hungarian Power Exchange
[92]	2021	LASO, ANN, RANDOM FOREST, RVM, XGBOOST-RVM	MAAPE, RMSE, MAE	ENTSO-E Transparency Platform
[94]	2022	ARIMA, Lasso, CNN, ANN, GRU, LSTNet, HFnet, Uber, W-GRU, W-HFnet, GHTnet	MAPE, MAE	New York Independent System Operator (NYISO)
[95]	2019	DR-R-SDAE, LSTM, R-DAE, SDAE, CNN, SVM	RMSE, MAE, MAPE	Ontario, Canada
[96]	2017	RNN, PCA with CNN, MLP with LM, CA with CNN, SVC-RFE with CNN	MAPE	Pennsylvania–New Jersey–Maryland (PJM) electricity markets
[97]	2012	ANN, ANN-ARIMA	MAPE, MAE, RMSE	Australian national electricity market
[98]	2023	DTR, LSTM, XGBOOST, Hybrid	MAE, RMSE	Several market players in the Alberta electricity market
[73]	2018	Several AR models	MAE	Nord Pool (NP; Northern Europe), PJM Interconnection (Northeastern United States), EPEX Germany and Austria (Central Europe)
[93]	2017	f LSSVM, BND-LSSVM, FOA-LSSVM	MAPE, MAE, RMSE	The daily electricity price in the northeast power grid of China
[74]	2016	several types of ARIMA	FMSE	Spanish electricity market
[88]	2024	XAI CNN, LSTM, XGBOOST	RMSE, MAE, MAPE	Italian electricity market, ERCOT Electricity Market
[89]	2023	CNN-based models	MAE	Italian electricity market, ERCOT Electricity Market



## 6. Discussion and Comparison Between Load and Price Forecasting

In the above analyses, qualitative techniques, models and approaches were presented in two very interesting sectors, ELF and EPF. The approach used by each researcher to solve a topic differs according to the data at his or her disposal and the ultimate goal of the problem. Based on what was presented, the comparison between ELF and EPF models does not reveal significant differences in their application and complexity. In many cases, however, we see that models focused on price forecasting require quite complex data preprocessing, a fact that is based on the often unpredictable fluctuations of the price, which is influenced by both static factors and unstable ones, such as political and social crises.

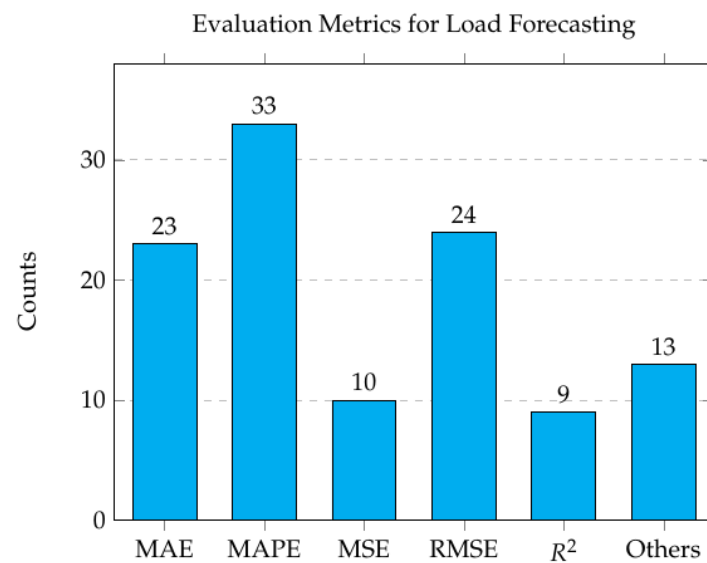
Load forecasting models primarily focus on predicting the demand for electricity, utilizing historical consumption data, weather conditions, and demographic factors. These models often employ statistical models and ML techniques to achieve high accuracy. On the other hand, price forecasting models aim to predict the wholesale electricity market prices, which are influenced by a broader range of variables, including supply-demand dynamics, fuel prices, regulatory changes, and the behavior of market participants. Price forecasting models frequently integrate advanced ML and DL algorithms to manage the complexity and volatility of price data.

Here, a fundamental difference that is observed between the methods in ELF and EPF may also emerge. Due to the fact that the value configuration depends on many diverse factors, such as the varying and unpredictable forms of RESs, it is observed that statistical models fall short compared to ML and DL because of their inability to handle a large number of inputs for their training. Ensemble models, which combine multiple forecasting methods, have shown considerable promise in both load and price forecasting by capitalizing on the strengths of different approaches. Despite their differences, both types of models are crucial for optimizing the electricity market and ensuring a reliable and cost-effective power supply.

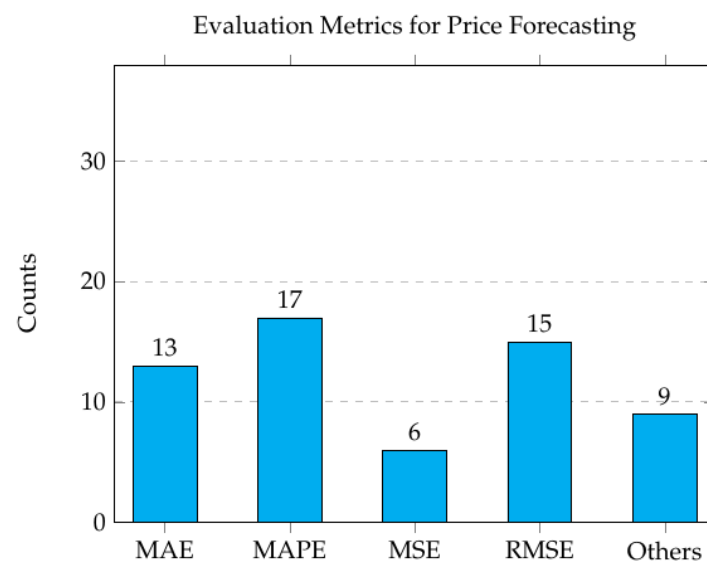
- Observing the ELF papers we have selected, we see that 30 are from 2019 onwards, and 13 are from before 2019. This observation indicates that significant progress has been made in the last five years in the development of models aimed at load forecasting.
- Regarding the EPF works, we see that 22 are from 2019 onwards and 7 before 2019. This observation indicates that significant progress has been made in the last five years in the development of models aimed at price forecasting.

Also, regarding Figures 8 and 9, which present the total number of related papers analyzed for each metric for both ELF and EPF, the following can be noticed:

- Regarding the studies related to ELF, it is observed that most of them use MAPE and MAE, followed by RMSE, with percentages 75%, 52% and 55%, respectively.
- Regarding the works on EPF, and similar to ELF, most of them employ MAPE, MAE and RMSE, with values of 57%, 43% and 50%, respectively.
- Regarding the EPF works, it is noted that none of the works use  $R^2$ , indicating the instability of this metric to provide reliable conclusions due to the greater fluctuations observed in the price time series compared to the load.



**Figure 8.** Total number of ELF studies per metric.



**Figure 9.** Total number of EPF studies per metric.

## 7. Conclusions and Future Study Proposals

In this comprehensive review, we have systematically examined the methodologies, challenges, and advancements in electricity load and price forecasting for the modern wholesale electricity market, two fields that are often interconnected in a cause–effect relationship. Our analysis covers a broad spectrum of forecasting techniques, from traditional statistical models to contemporary ML and advanced DL approaches. The dynamic and complex nature of electricity markets necessitates accurate and reliable forecasting to ensure operational efficiency, cost-effectiveness, and grid stability.

The main target of this review is to highlight the diverse range of methodologies employed in load and price forecasting, providing the reader with state-of-the-art methods for the two fields. For this purpose, an extensive study was conducted, based on which studies using different datasets and various forecasting approaches were selected. Traditional techniques like ARIMA and SARIMA remain foundational, while ML and DL models such as ANN, SVR, and ensemble methods have demonstrated significant improvements in predictive accuracy and adaptability. Accurate performance evaluation is crucial for forecasting models, and metrics such as MAE, MSE and MAPE are commonly used. A notable difference worth highlighting between ELF and EPF is that none of the EPF studies

use  $R^2$  as a metric. The main reason for this is that the time series of prices exhibit high volatility and fluctuations, including extreme and even negative values, which makes this metric less effective in providing clear conclusions. However, the selection of appropriate metrics is context-dependent, influenced by the specific needs of market participants and regulatory bodies.

Furthermore, based on what has been analyzed, it is observed that ensemble methods most often provide greater stability and effectiveness compared to single models, especially in cases where the outputs of each model are combined using a voting method to generate the final prediction. This phenomenon, which could potentially become a notable trend in the field, is more pronounced in price forecasting studies, a fact mainly due to the sharp fluctuations and variations in prices, especially during periods of crisis. Additionally, the literature does not include studies that connect behavior analysis with electricity consumption, two fields that show a relatively high correlation between them, mainly regarding the consumption of residential users and the load curve of islands.

The integration of diverse data sources, including historical load and price data, weather conditions, and economic indicators, enhances the robustness of forecasting models. Advanced data pre-processing techniques and the use of high-frequency data are pivotal in capturing market volatility and trends. The review identifies several challenges, including data scarcity, fluctuation of the data according to each different area, market volatility, and the non-linear nature of electricity prices. Innovative solutions, such as novel hybrid models combining multiple forecasting techniques and real-time data analytics, show promise in addressing these issues. The future of load and price forecasting lies in the continued evolution of artificial intelligence and ML. The adoption of DL, real-time analytics, and the Internet of Things (IoT) offers new opportunities for enhancing forecasting accuracy and responsiveness. Furthermore, collaboration between academia, industry, and regulatory bodies is essential for the development of standardized frameworks and best practices.

In conclusion, effective load and price forecasting are integral to the success of modern wholesale electricity markets. The advancements in forecasting techniques hold great potential to improve market operation and sustainability, reduce costs, and enhance the reliability of electricity supply. Ongoing research and innovation, supported by robust data infrastructures and collaborative efforts, will be key to addressing existing challenges and harnessing new opportunities in this dynamic field.

Regarding potential gaps in the fields that were analyzed, it is worth noting that, due to the trends observed in recent years on a global level concerning the use of standalone large-scale batteries in the grid, which aim at smoothing out electricity prices, there is not an extensive literature on this highly interesting issue.

So, in terms of future study proposals, firstly, managing with non-stationary and nonlinear load data is a challenge, as conventional STLF models typically assume that load data remains stationary and linear. Nonetheless, load data can exhibit non-stationarity and nonlinearity as a result of shifts in consumer behavior, the emergence of new technologies, and various other factors. Future investigations might focus on devising STLF models capable of managing non-stationary and nonlinear load data, potentially utilizing advanced ML models or more adaptable statistical frameworks.

Secondly, the incorporation of online learning is another avenue for improvement. Online learning, a branch of ML, adapts to real-time changes in the energy system. These algorithms continuously update their learning based on new data, allowing them to refine forecasts in real-time. Future studies could aim to develop STLF models that leverage online learning algorithms, thereby improving the accuracy and timeliness of forecasts.

Thirdly, the development of advanced models and the evaluation of their results using cross-validation techniques or not represents a notable direction for future research, which will provide us with meaningful comparative results and conclusions, as cross-validation ensures a reliable estimate of model performance on future, unseen data.

Fourth, the development of models for DERs, like standalone batteries, is becoming more crucial with the growing adoption of technologies such as rooftop solar panels and

integrated energy storage systems. These advancements pose new challenges for STLF models. Future research could focus on creating STLF models that integrate DERs, either by forecasting renewable energy generation or by assessing their impact on electricity demand.

Finally, the investigation of behavior analysis in the field of ELF is crucial, as discussed above. By understanding the patterns and habits of consumers, especially residential users, it is possible to refine load forecasting models. Human behavior significantly influences energy consumption, such as variations in demand based on time of day, weather conditions, or lifestyle choices. Integrating behavior analysis could improve the accuracy of forecasting models by capturing these dynamic factors and could be particularly useful in predicting consumption patterns during peak periods, contributing to more efficient energy management for flexibility and demand response programs.

**Author Contributions:** Conceptualization, V.L. and G.V.; methodology, V.L.; software, V.L. and G.V.; validation, V.L., G.V., P.P., E.T., D.B. and L.H.T.; formal analysis, V.L. and G.V.; investigation, V.L. and G.V.; resources, V.L. and G.V.; data curation, V.L. and G.V.; writing—original draft preparation, V.L. and G.V.; writing—review and editing, V.L., G.V., P.P., E.T., D.B. and L.H.T.; visualization, V.L. and G.V.; supervision, D.B. and L.H.T.; project administration, D.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** This study did not utilize any datasets.

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

## Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	Autoregressive Integrated Moving Average
SVM	Support Vector Machine
TL	Transfer Learning
SARIMA	Seasonal Autoregressive Integrated Moving Average
MLP	Multilayer Perceptron
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
ELF	Electricity Load Forecasting
EPF	Electricity Price Forecasting
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
ESN	Echo State Network
GRU	Gated Recurrent Unit
LR	Linear Regression
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
PJM	Pennsylvania–New Jersey–Maryland Interconnection
IHEPC	Individual-Household-Electric-Power-Consumption
NARX	Nonlinear Autoregressive Exogenous Model
DNN	Deep Neural Network
XGBoost	Extreme Gradient Boosting
RF	Random Forest
SVR	Support Vector Regression
KNN	K-Nearest Neighbors
BPNN	Backpropagation Neural Network
DAM	Day-Ahead Market
IDA	Intraday Market
BM	Balancing Market
ES	Exponential Smoothing

DES	Double Exponential Smoothing
TES	Triple Exponential Smoothing
EMD	Empirical Mode Decomposition
ELM	Extreme Learning Machine
NYISO	New York Independent System Operator
DER	Distributed Energy Resources

## References

- Dagoumas, A. Impact of Bilateral Contracts on Wholesale Electricity Markets: In a Case Where a Market Participant Has Dominant Position. *Appl. Sci.* **2019**, *9*, 382. [\[CrossRef\]](#)
- Makrygiorgou, J.J.; Karavas, C.S.; Dikaiaikos, C.; Moraitis, I.P. The Electricity Market in Greece: Current Status, Identified Challenges, and Arranged Reforms. *Sustainability* **2023**, *15*, 3767. [\[CrossRef\]](#)
- Loizidis, S.; Konstantinidis, G.; Theocharides, S.; Kyprianou, A.; Georghiou, G.E. Electricity Day-Ahead Market Conditions and Their Effect on the Different Supervised Algorithms for Market Price Forecasting. *Energies* **2023**, *16*, 4617. [\[CrossRef\]](#)
- Gea-Bermúdez, J.; Das, K.; Koduvere, H.; Koivisto, M.J. Day-Ahead Market Modelling of Large-Scale Highly-Renewable Multi-Energy Systems: Analysis of the North Sea Region towards 2050. *Energies* **2021**, *14*, 88. [\[CrossRef\]](#)
- Mousavi, F.; Nazari-Heris, M.; Mohammadi-Ivatloo, B.; Asadi, S. Chapter 1—Energy market fundamentals and overview. In *Energy Storage in Energy Markets*; Mohammadi-Ivatloo, B., Mohammadpour Shotorbani, A., Anvari-Moghaddam, A., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 1–21. [\[CrossRef\]](#)
- Kılıç, D.K.; Nielsen, P.; Thibbotuwawa, A. Intraday Electricity Price Forecasting via LSTM and Trading Strategy for the Power Market: A Case Study of the West Denmark DK1 Grid Region. *Energies* **2024**, *17*, 2909. [\[CrossRef\]](#)
- Roumkos, C.; Biskas, P.N.; Marneris, I.G. Integration of European Electricity Balancing Markets. *Energies* **2022**, *15*, 2240. [\[CrossRef\]](#)
- What Is Spot Power Trading? Available online: <https://www.kyos.com/faq/what-is-spot-power-trading/> (accessed on 15 August 2024).
- Ospina, R.; Gondim, J.A.M.; Leiva, V.; Castro, C. An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. *Mathematics* **2023**, *11*, 3069. [\[CrossRef\]](#)
- Noor, T.H.; Almars, A.M.; Alwateer, M.; Almaliki, M.; Gad, I.; Atlam, E.S. SARIMA: A Seasonal Autoregressive Integrated Moving Average Model for Crime Analysis in Saudi Arabia. *Electronics* **2022**, *11*, 3986. [\[CrossRef\]](#)
- Jónsson, T.; Pinson, P.; Nielsen, H.A.; Madsen, H. Exponential Smoothing Approaches for Prediction in Real-Time Electricity Markets. *Energies* **2014**, *7*, 3710–3732. [\[CrossRef\]](#)
- Queiroz, R.G.S.; David, S.A. Performance of the Realized-GARCH Model against Other GARCH Types in Predicting Cryptocurrency Volatility. *Risks* **2023**, *11*, 211. [\[CrossRef\]](#)
- Guido, R.; Ferrisi, S.; Lofaro, D.; Conforti, D. An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information* **2024**, *15*, 235. [\[CrossRef\]](#)
- Kim, S.J.; Bae, S.J.; Jang, M.W. Linear Regression Machine Learning Algorithms for Estimating Reference Evapotranspiration Using Limited Climate Data. *Sustainability* **2022**, *14*, 11674. [\[CrossRef\]](#)
- Zhang, J.; Wang, R.; Jia, A.; Feng, N. Optimization and Application of XGBoost Logging Prediction Model for Porosity and Permeability Based on K-means Method. *Appl. Sci.* **2024**, *14*, 3956. [\[CrossRef\]](#)
- Nadkarni, S.B.; Vijay, G.S.; Kamath, R.C. Comparative Study of Random Forest and Gradient Boosting Algorithms to Predict Airfoil Self-Noise. *Eng. Proc.* **2023**, *59*, 24. [\[CrossRef\]](#)
- Lazcano, A.; Jaramillo-Morán, M.A.; Sandubete, J.E. Back to Basics: The Power of the Multilayer Perceptron in Financial Time Series Forecasting. *Mathematics* **2024**, *12*, 1920. [\[CrossRef\]](#)
- Laitsos, V.; Vontzos, G.; Bargiotas, D.; Daskalopulu, A.; Tsoukalas, L.H. Enhanced Automated Deep Learning Application for Short-Term Load Forecasting. *Mathematics* **2023**, *11*, 2912. [\[CrossRef\]](#)
- Huang, D.; Liu, Z.; Wu, D. Research on Ensemble Learning-Based Feature Selection Method for Time-Series Prediction. *Appl. Sci.* **2024**, *14*, 40. [\[CrossRef\]](#)
- De Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean absolute percentage error for regression models. *Neurocomputing* **2016**, *192*, 38–48. [\[CrossRef\]](#)
- Cameron, A.C.; Windmeijer, F.A. An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.* **1997**, *77*, 329–342. [\[CrossRef\]](#)
- Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [\[CrossRef\]](#)
- Sethi, R.; Kleissl, J. Comparison of short-term load forecasting techniques. In Proceedings of the 2020 IEEE Conference on Technologies for Sustainability (SusTech), Santa Ana, CA, USA, 23–25 April 2020; pp. 1–6.
- Chou, J.S.; Tran, D.S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy* **2018**, *165*, 709–726. [\[CrossRef\]](#)
- Hadri, S.; Naitmalek, Y.; Najib, M.; Bakhouya, M.; Fakhri, Y.; Elaroussi, M. A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings. *Procedia Comput. Sci.* **2019**, *160*, 173–180. [\[CrossRef\]](#)

26. Vrablecová, P.; Bou Ezzeddine, A.; Rozinajová, V.; Šárik, S.; Sangaiah, A.K. Smart grid load forecasting using online support vector regression. *Comput. Electr. Eng.* **2018**, *65*, 102–117. [\[CrossRef\]](#)
27. Pelka, P. Analysis and Forecasting of Monthly Electricity Demand Time Series Using Pattern-Based Statistical Methods. *Energies* **2023**, *16*, 827. [\[CrossRef\]](#)
28. Elamin, N.; Fukushige, M. Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy* **2018**, *165*, 257–268. [\[CrossRef\]](#)
29. Mado, I.; Rajagukguk, A.; Triwiyatno, A.; Fadlullah, A. Short-Term Electricity Load Forecasting Model Based DSARIMA. *Int. J. Electr. Energy Power Syst. Eng.* **2022**, *5*, 6–11. [\[CrossRef\]](#)
30. Abumohsen, M.; Owda, A.Y.; Owda, M. Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms. *Energies* **2023**, *16*, 2283. [\[CrossRef\]](#)
31. Li, S.; Zhang, W.; Wang, P. TS2ARCformer: A Multi-Dimensional Time Series Forecasting Framework for Short-Term Load Prediction. *Energies* **2023**, *16*, 5825. [\[CrossRef\]](#)
32. Salami, M.; Movahedi Sobhani, F.; Ghazizadeh, M.S. Short-Term Forecasting of Electricity Supply and Demand by Using the Wavelet-PSO-NNs-SO Technique for Searching in Big Data of Iran's Electricity Market. *Data* **2018**, *3*, 43. [\[CrossRef\]](#)
33. Gao, X.; Li, X.; Zhao, B.; Ji, W.; Jing, X.; He, Y. Short-Term Electricity Load Forecasting Model Based on EMD-GRU with Feature Selection. *Energies* **2019**, *12*, 1140. [\[CrossRef\]](#)
34. Yuan, B.; He, B.; Yan, J.; Jiang, J.; Wei, Z.; Shen, X. Short-term electricity consumption forecasting method based on empirical mode decomposition of long-short term memory network. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *983*, 012004. [\[CrossRef\]](#)
35. Zhou, F.; Zhou, H.; Li, Z.; Zhao, K. Multi-Step Ahead Short-Term Electricity Load Forecasting Using VMD-TCN and Error Correction Strategy. *Energies* **2022**, *15*, 5375. [\[CrossRef\]](#)
36. Javed, U.; Ijaz, K.; Jawad, M.; Ansari, E.A.; Shabbir, N.; Kütt, L.; Husev, O. Exploratory Data Analysis Based Short-Term Electrical Load Forecasting: A Comprehensive Analysis. *Energies* **2021**, *14*, 5510. [\[CrossRef\]](#)
37. He, F.; Zhou, J.; kai Feng, Z.; 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. [\[CrossRef\]](#)
38. He, W. Load Forecasting via Deep Neural Networks. *Procedia Comput. Sci.* **2017**, *122*, 308–314. [\[CrossRef\]](#)
39. Shi, H.; Xu, M.; Li, R. Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. *IEEE Trans. Smart Grid* **2018**, *9*, 5271–5280. [\[CrossRef\]](#)
40. Amarasinghe, K.; Marino, D.L.; Manic, M. Deep neural networks for energy load forecasting. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017; pp. 1483–1488. [\[CrossRef\]](#)
41. Wang, M.; Yu, J.; Zhou, M.; Quan, W.; Cheng, R. Joint Forecasting Model for the Hourly Cooling Load and Fluctuation Range of a Large Public Building Based on GA-SVM and IG-SVM. *Sustainability* **2023**, *15*, 6833. [\[CrossRef\]](#)
42. Bessani, M.; Massignan, J.A.; Santos, T.M.; London, J.B.; Maciel, C.D. Multiple households very short-term load forecasting using bayesian networks. *Electr. Power Syst. Res.* **2020**, *189*, 106733. [\[CrossRef\]](#)
43. Caicedo-Vivas, J.S.; Alfonso-Morales, W. Short-Term Load Forecasting Using an LSTM Neural Network for a Grid Operator. *Energies* **2023**, *16*, 7878. [\[CrossRef\]](#)
44. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [\[CrossRef\]](#)
45. Imani, M.; Ghassemian, H. Electrical Load Forecasting Using Customers Clustering and Smart Meters in Internet of Things. In Proceedings of the 2018 9th International Symposium on Telecommunications (IST), Tehran, Iran, 17–19 December 2018; pp. 113–117. [\[CrossRef\]](#)
46. Raju, M.P.; Laxmi, A.J. IOT based Online Load Forecasting using Machine Learning Algorithms. *Procedia Comput. Sci.* **2020**, *171*, 551–560. [\[CrossRef\]](#)
47. Li, L.; Meinrenken, C.J.; Modi, V.; Culligan, P.J. Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features. *Appl. Energy* **2021**, *287*, 116509. [\[CrossRef\]](#)
48. Fekri, M.N.; Patel, H.; Grolinger, K.; Sharma, V. Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network. *Appl. Energy* **2021**, *282*, 116177. [\[CrossRef\]](#)
49. Andriopoulos, N.; Magklaras, A.; Birbas, A.; Papalexopoulos, A.; Valouxis, C.; Daskalaki, S.; Birbas, M.; Housos, E.; Papaioannou, G.P. Short Term Electric Load Forecasting Based on Data Transformation and Statistical Machine Learning. *Appl. Sci.* **2021**, *11*, 158. [\[CrossRef\]](#)
50. Alanazi, M.D.; Saeed, A.; Islam, M.; Habib, S.; Sherazi, H.I.; Khan, S.; Shees, M.M. Enhancing Short-Term Electrical Load Forecasting for Sustainable Energy Management in Low-Carbon Buildings. *Sustainability* **2023**, *15*, 16885. [\[CrossRef\]](#)
51. Jalali, S.M.J.; Ahmadian, S.; Khosravi, A.; Shafie-khah, M.; Nahavandi, S.; Catalão, J.P.S. A Novel Evolutionary-Based Deep Convolutional Neural Network Model for Intelligent Load Forecasting. *IEEE Trans. Ind. Inform.* **2021**, *17*, 8243–8253. [\[CrossRef\]](#)
52. Chitalia, G.; Pipattanasomporn, M.; Garg, V.; Rahman, S. Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Appl. Energy* **2020**, *278*, 115410. [\[CrossRef\]](#)
53. Hamilton, W.L. *Graph Representation Learning*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2020.
54. Vontzos, G.; Laitsos, V.; Charakopoulos, A.; Bargiotas, D.; Karakasidis, T.E. Estimating Spatio-Temporal Building Power Consumption Based on Graph Convolution Network Method. *Dynamics* **2024**, *4*, 337–356. [\[CrossRef\]](#)



55. Hu, Y.; Cheng, X.; Wang, S.; Chen, J.; Zhao, T.; Dai, E. Times series forecasting for urban building energy consumption based on graph convolutional network. *Appl. Energy* **2022**, *307*, 118231. [\[CrossRef\]](#)
56. Wei, M.; Wen, M.; Zhang, Y. A novel spatial electric load forecasting method based on LDTW and GCN. *IET Gener. Transm. Distrib.* **2024**, *18*, 491–505. [\[CrossRef\]](#)
57. Liao, X.; Cao, N.; Li, M.; Kang, X. Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days. In Proceedings of the 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 12–13 January 2019; pp. 675–678. [\[CrossRef\]](#)
58. Guo, Z.; Zhao, W.; Lu, H.; Wang, J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renew. Energy* **2012**, *37*, 241–249. [\[CrossRef\]](#)
59. Alobaidi, M.H.; Chebana, F.; Meguid, M.A. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Appl. Energy* **2018**, *212*, 997–1012. [\[CrossRef\]](#)
60. Du, P.; Jin, Y.; Zhang, K. A Hybrid Multi-Step Rolling Forecasting Model Based on SSA and Simulated Annealing—Adaptive Particle Swarm Optimization for Wind Speed. *Sustainability* **2016**, *8*, 754. [\[CrossRef\]](#)
61. Ma, X.; Jin, Y.; Dong, Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Appl. Soft Comput.* **2017**, *54*, 296–312. [\[CrossRef\]](#)
62. Sadaei, H.J.; de Lima e Silva, P.C.; Guimarães, F.G.; Lee, M.H. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy* **2019**, *175*, 365–377. [\[CrossRef\]](#)
63. Rafi, S.H.; Nahid-Al-Masood; Deeba, S.R.; Hossain, E. A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. *IEEE Access* **2021**, *9*, 32436–32448. [\[CrossRef\]](#)
64. Wang, J.; Liu, H.; Zheng, G.; Li, Y.; Yin, S. Short-Term Load Forecasting Based on Outlier Correction, Decomposition, and Ensemble Reinforcement Learning. *Energies* **2023**, *16*, 4401. [\[CrossRef\]](#)
65. Hammad, M.A.; Jereb, B.; Rosi, B.; Dragan, D. Methods and Models for Electric Load Forecasting: A Comprehensive Review. *Logist. Supply Chain. Sustain. Glob. Chall.* **2020**, *11*, 51–76. [\[CrossRef\]](#)
66. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [\[CrossRef\]](#)
67. Ozerdem, O.C.; Olaniyi, E.O.; Oyedotun, O.K. Short term load forecasting using particle swarm optimization neural network. *Procedia Comput. Sci.* **2017**, *120*, 382–393. [\[CrossRef\]](#)
68. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
69. Skopal, R. Short-term hourly price forward curve prediction using neural network and hybrid ARIMA-NN model. In Proceedings of the 2015 International Conference on Information and Digital Technologies, Zilina, Slovakia, 7–9 July 2015; pp. 335–338.
70. Liu, H.; Shi, J. Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Econ.* **2013**, *37*, 152–166. [\[CrossRef\]](#)
71. Rajan, P.; Chandrakala, K.V. Statistical Model Approach of Electricity Price Forecasting for Indian Electricity Market. In Proceedings of the 2021 IEEE Madras Section Conference (MASCOS), Chennai, India, 27–28 August 2021; pp. 1–5. [\[CrossRef\]](#)
72. Bissing, D.; Klein, M.T.; Chinnathambi, R.A.; Selvaraj, D.F.; Ranganathan, P. A Hybrid Regression Model for Day-Ahead Energy Price Forecasting. *IEEE Access* **2019**, *7*, 36833–36842. [\[CrossRef\]](#)
73. Marciusz, G.; Serafin, T.; Weron, R. Selection of Calibration Windows for Day-Ahead Electricity Price Forecasting. *Energies* **2018**, *11*, 2364. [\[CrossRef\]](#)
74. Sánchez de la Nieta, A.A.; González, V.; Contreras, J. Portfolio Decision of Short-Term Electricity Forecasted Prices through Stochastic Programming. *Energies* **2016**, *9*, 1069. [\[CrossRef\]](#)
75. Abunofal, M.; Poshia, N.; Qussous, R.; Weidlich, A. Comparative Analysis of Electricity Market Prices Based on Different Forecasting Methods. In Proceedings of the 2021 IEEE Madrid PowerTech, Madrid, Spain, 28 June–2 July 2021; pp. 1–6. [\[CrossRef\]](#)
76. Banitalebi, B.; Hoque, M.E.; Appadoo, S.S.; Thavaneswaran, A. Regularized Probabilistic Forecasting of Electricity Wholesale Price and Demand. In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, ACT, Australia, 1–4 December 2020; pp. 28–35. [\[CrossRef\]](#)
77. Shikhina, A.; Kochengin, A.; Chrysostomou, G.; Shikhin, V. Investigation of autoregressive forecasting models for market electricity price. In Proceedings of the 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), Palermo, Italy, 16–18 June 2020; pp. 570–575. [\[CrossRef\]](#)
78. Khairuddin, F.K.; Zaini, F.A.; Sulaima, M.F.; Shamsudin, N.H.; Abu Hanifah, M.S. A Single-Buyer Model of Imbalance Cost Pass-Through Pricing Forecasting in the Malaysian Electricity Supply Industry. *Electricity* **2024**, *5*, 295–312. [\[CrossRef\]](#)
79. Entezari, N.; Fuinhas, J.A. Quantifying the Impact of Risk on Market Volatility and Price: Evidence from the Wholesale Electricity Market in Portugal. *Sustainability* **2024**, *16*, 2691. [\[CrossRef\]](#)
80. Mišnić, N.; Pejović, B.; Jovović, J.; Rogić, S.; Đurišić, V. The Economic Viability of PV Power Plant Based on a Neural Network Model of Electricity Prices Forecast: A Case of a Developing Market. *Energies* **2022**, *15*, 6219. [\[CrossRef\]](#)
81. Gonzalez, V.; Contreras, J.; Bunn, D.W. Forecasting Power Prices Using a Hybrid Fundamental-Econometric Model. *IEEE Trans. Power Syst.* **2012**, *27*, 363–372. [\[CrossRef\]](#)
82. Menéndez Medina, A.; Heredia Álvaro, J.A. Using Generative Pre-Trained Transformers (GPT) for Electricity Price Trend Forecasting in the Spanish Market. *Energies* **2024**, *17*, 2338. [\[CrossRef\]](#)



83. Conte, T.; Oliveira, R. Comparative Analysis between Intelligent Machine Committees and Hybrid Deep Learning with Genetic Algorithms in Energy Sector Forecasting: A Case Study on Electricity Price and Wind Speed in the Brazilian Market. *Energies* **2024**, *17*, 829. [\[CrossRef\]](#)
84. Laitos, V.; Vontzos, G.; Bargiotas, D.; Daskalopulu, A.; Tsoukalas, L.H. Data-Driven Techniques for Short-Term Electricity Price Forecasting through Novel Deep Learning Approaches with Attention Mechanisms. *Energies* **2024**, *17*, 1625. [\[CrossRef\]](#)
85. Jin, A.; Lee, D.; Park, J.B.; Roh, J.H. Day-Ahead Electricity Market Price Forecasting Considering the Components of the Electricity Market Price; Using Demand Decomposition, Fuel Cost, and the Kernel Density Estimation. *Energies* **2023**, *16*, 3222. [\[CrossRef\]](#)
86. Baskan, D.E.; Meyer, D.; Mieck, S.; Faubel, L.; Klöpper, B.; Strem, N.; Wagner, J.A.; Koltermann, J.J. A Scenario-Based Model Comparison for Short-Term Day-Ahead Electricity Prices in Times of Economic and Political Tension. *Algorithms* **2023**, *16*, 177. [\[CrossRef\]](#)
87. Pinhão, M.; Fonseca, M.; Covas, R. Electricity Spot Price Forecast by Modelling Supply and Demand Curve. *Mathematics* **2022**, *10*, 2012. [\[CrossRef\]](#)
88. Heistrene, L.; Belikov, J.; Baimel, D.; Katzir, L.; Machlev, R.; Levy, K.; Mannor, S.; Levron, Y. An Improved and Explainable Electricity Price Forecasting Model via SHAP-based Error Compensation Approach. *IEEE Trans. Artif. Intell.* **2024**, *1*, 1–11. [\[CrossRef\]](#)
89. Heistrene, L.; Machlev, R.; Perl, M.; Belikov, J.; Baimel, D.; Levy, K.; Mannor, S.; Levron, Y. Explainability-based Trust Algorithm for electricity price forecasting models. *Energy AI* **2023**, *14*, 100259. [\[CrossRef\]](#)
90. Dong, J.; Dou, X.; Bao, A.; Zhang, Y.; Liu, D. Day-Ahead Spot Market Price Forecast Based on a Hybrid Extreme Learning Machine Technique: A Case Study in China. *Sustainability* **2022**, *14*, 7767. [\[CrossRef\]](#)
91. Saglam, M.; Lv, X.; Spataru, C.; Karaman, O.A. Instantaneous Electricity Peak Load Forecasting Using Optimization and Machine Learning. *Energies* **2024**, *17*, 777. [\[CrossRef\]](#)
92. Bhatia, K.; Mittal, R.; Varanasi, J.; Tripathi, M. An ensemble approach for electricity price forecasting in markets with renewable energy resources. *Util. Policy* **2021**, *70*, 101185. [\[CrossRef\]](#)
93. Guo, W.; Zhao, Z. A Novel Hybrid BND-FOA-LSSVM Model for Electricity Price Forecasting. *Information* **2017**, *8*, 120. [\[CrossRef\]](#)
94. Yang, H.; Schell, K.R. GHTnet: Tri-Branch deep learning network for real-time electricity price forecasting. *Energy* **2022**, *238*, 122052. [\[CrossRef\]](#)
95. Jahangir, H.; Tavarani, H.; Baghali, S.; Ahmadian, A.; Elkamel, A.; Golkar, M.A.; Castilla, M. A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2369–2381. [\[CrossRef\]](#)
96. Khajeh, M.G.; Maleki, A.; Rosen, M.A.; Ahmadi, M.H. Electricity price forecasting using neural networks with an improved iterative training algorithm. *Int. J. Ambient. Energy* **2018**, *39*, 147–158. [\[CrossRef\]](#)
97. Zhang, J.; Tan, Z.; Yang, S. Day-ahead electricity price forecasting by a new hybrid method. *Comput. Ind. Eng.* **2012**, *63*, 695–701. [\[CrossRef\]](#)
98. Manfre Jaimes, D.; Zamudio López, M.; Zareipour, H.; Quashie, M. A Hybrid Model for Multi-Day-Ahead Electricity Price Forecasting considering Price Spikes. *Forecasting* **2023**, *5*, 499–521. [\[CrossRef\]](#)

**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.