Review

Electrical Load Forecasting in the Industrial Sector: A Literature Review of Machine Learning Models and Architectures for Grid Planning

Jannis Eckhoff [1,2,*], Simran Wadhwa [3], Marc Fette [1,2], Jens Peter Wulfsberg [2] and Chathura Wanigasekara [3,4]

1 Laboratory for Manufacturing Technologies, Helmut Schmidt University, 22043 Hamburg, Germany; marc.fette@hsu-hh.de
2 CTC GmbH (An AIRBUS Company), 21684 Stade, Germany; jens.wulfsberg@hsu-hh.de
3 Center for Industrial Mathematics, University of Bremen, 28359 Bremen, Germany; wsimran2000@gmail.com (S.W.); chathura.wanigasekara@dlr.de (C.W.)
4 Institute of Maritime Technologies and Propulsion Systems, German Aerospace Centre (DLR), 21502 Geesthacht, Germany
* Correspondence: jannis.eckhoff@hsu-hh.de

Abstract

The energy transition, driven by the global shift toward renewable and electrification, necessitates accurate and efficient prediction of electrical load profiles to quantify energy consumption. Therefore, the systematic literature review (SLR), followed by PRISMA guidelines, synthesizes hybrid architectures for sequential electrical load profiles, aiming to span statistical techniques, machine learning (ML), and deep learning (DL) strategies for optimizing performance and practical viability. The findings reveal a dominant trend towards complex hybrid models leveraging the combined strengths of DL architectures such as long short-term memory (LSTM) and optimization algorithms such as genetic algorithm and Particle Swarm Optimization (PSO) to capture non-linear relationships. Thus, hybrid models achieve superior performance by synergistically integrating components such as Convolutional Neural Network (CNN) for feature extraction and LSTMs for temporal modeling with feature selection algorithms, which collectively capture local trends, cross-correlations, and long-term dependencies in the data. A crucial challenge identified is the lack of an established framework to manage adaptable output lengths in dynamic neural network forecasting. Addressing this, we propose the first explicit idea of decoupling output length predictions from the core signal prediction task. A key finding is that while models, particularly optimization-tuned hybrid architectures, have demonstrated quantitative superiority over conventional shallow methods, their performance assessment relies heavily on statistical measures like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). However, for comprehensive performance assessment, there is a crucial need for developing tailored, application-based metrics that integrate system economics and major planning aspects to ensure reliable domain-specific validation.

Keywords: electrical grid planning; decoupled prediction; industrial load forecasting; adaptive horizon prediction; cycle-aware load learning; hybrid neural network

1. Introduction

Predicting the electrical load profile is crucial for effective electric grid planning, as it involves forecasting the electrical power needed over time and the peak amplitude [1–3].

Managing the power supply at regular intervals (e.g., hourly or daily) ensures reliability and fosters further growth [4,5]. As the world shifts towards cleaner energy sources, forecasting techniques support energy management by reducing waste and promoting sustainability [5,6]. The significance of this lies in maintaining grid stability, preventing blackouts, and supporting economic operations. However, challenges arise from factors such as increased consumer demand, unpredictable weather conditions that affect renewable energy generation, and the complexity of the datasets [7]. Due to the increasing integration of renewable energy sources, especially solar and wind, it introduces volatility and uncertainty into the grid [8]. In addition, real-time forecasting methods demand more precise and adaptable solutions to meet energy needs. These issues set the stage for exploring how advanced architectures can enhance load forecasting for grid planning [3,4,9].

The rise in population and the expansion of industries has increased the pressure on electrical grids to deliver consistent and reliable power, making efficient forecasting a priority [10]. The importance of load forecasting cannot be overstated in today's energy landscape, where environmental concerns drive the demand for sustainable solutions [1,3,11]. Load prediction models have several benefits, such as minimizing operational costs for utility companies by optimizing power generation and distribution, reducing energy waste, and better integration of renewable energy sources, which are essential for reducing reliance on fossil fuels but require careful management due to their intermittent nature [12]. The limitations of forecasting methods that exist include the complex patterns in sequential data and adapting to sudden changes in demand [4]. Thus, these gaps increase the importance of a comprehensive review to evaluate recent approaches, identify opportunities for improvement, and motivate this study to discuss advanced techniques, especially those that leverage machine learning.

This literature review concentrates on models and architectures for predicting sequential signals, including electrical load profiles, with a significance of their applications and the methodologies employed while also considering algorithms developed for other signals, such as wind or speed, that share similar predictive objectives and boundary conditions [9,13,14]. The review aims to explain the overview of different approaches, synthesize existing methods, compare their capabilities and limitations, and identify the trends in this field [15]. The diverse methodologies, as seen in most papers, include techniques such as Auto Regressive Integrated Moving Average (ARIMA), advanced deep learning models like LSTM networks, and hybrid model strategies that combine different architectures, driven by the need to optimize not only the model performance and accuracy but also their practical viability under real-world constraints such as computational complexity and limited data availability [1,10,11].

Although recent literature demonstrates that hybrid DL architectures, such as optimization-tuned LSTMs and CNNs, have successfully minimized forecasting errors through hyperparameter tuning, these models remain constrained by fixed forecasting horizons, which requires rigid sliding windows that cannot dynamically adjust to irregular load profiles without computationally expensive retraining [2]. A critical assessment of state-of-the-art methodologies reveals that even large time-series models treat the output sequence length as a static hyperparameter determined a priori, thereby lacking a framework to the treat the duration of a load event as a stochastic variable that needs to be predicted independently [15]. To address this point, the study proposes the first distinct strategy of decoupling output length prediction from the core signal estimation; this modernization is necessary to overcome the inherent limitations of single-step and fixed horizon strategies, providing a new architecture capable of handling dynamic length sequences without introducing synthetic noise through padding [4,9].

A logical framework supports the structure of this paper to provide a clear understanding of the topic. Section 2 outlines the methodology, which describes the systematic approach for the review process, including research questions, as well as inclusion and exclusion criteria. Section 3 represents the findings, organized by model types such as statistical, ML, DL, and hybrid, while also covering the insights into the inputs and outputs. Section 4 summarizes the key insights and highlights their significance for enhancing electrical grid planning.

2. Methodology of Literature Review

To determine models and architectures for predicting sequential load profiles, the study adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, as depicted in Figure 1.
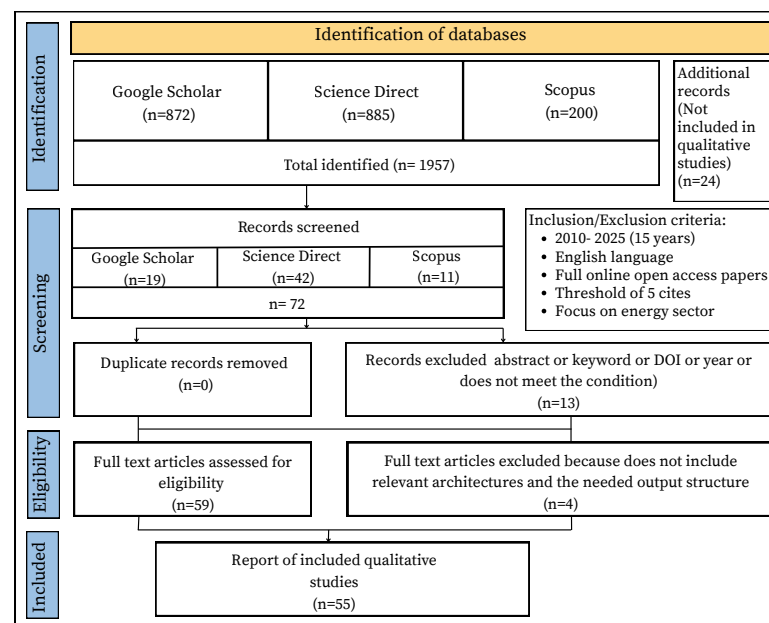
Figure 1. Flow diagram of systematic search results (based on PRISMA).

2.1. Research Questions

For this study, three research questions (RQs) have been formulated, as listed in Table 1. Using the RQ as a foundation, a structured search strategy is created to effectively gather relevant literature. The search method is carefully designed to find studies that investigate the input and output characteristics of electrical load prediction models (RQ1), assess their use in grid planning and management (RQ2), and analyze their performance based on accuracy and forecasting metrics (RQ3).

Table 1. Research questions guiding the systematic literature review.

ID	Research Question
RQ1	What are the primary input and output characteristics considered in the development of electrical load prediction models?
RQ2	How can identified models and frameworks be applied effectively to improve electrical grid planning and management?
RQ3	How do different models and architectures compare in terms of accuracy and performance metrics for the prediction of electrical loads?

2.2. Search Strategy

The search for the relevant literature was based on a primary search string:

'(predict * OR forecast * OR simulation * OR model) AND (power OR energy * OR load) AND (manufacture * OR machine * OR system OR industry * OR product *)'.

The search string was derived by decomposing the RQs into three primary keyword groups. The intersection of these groups ensure the literature discusses predictive modeling (Methodology) of electrical consumption (Target variable) within complex systems (Domain), which is essential for modern grid planning.

Group 1: Methodology

Keywords: (predict * OR forecast * OR simulation * OR model)
The search must retrieve studies that explicitly develop or test algorithms to include architectures such as LSTM and CNN in simulated environments before deployment, as seen in studies or probabilistic forecasting frameworks which link to RQ1 (identifying model characteristics) and RQ3 (comparing architectures and accuracy).

Group 2: Target variable

Keywords: (power OR energy * OR load)
The group directly addresses RQ1 by defining input/output characteristics. For instance, the studies may focus on "power load forecasting", "energy consumption", or specific "load profiles".

Group 3: Domain

Keywords: (manufacture * OR machine * OR system OR industry * OR product *)
To address RQ2, it is necessary to identify where these models are deployed. Consequently, keywords like "industry", "manufacture", and "machine" are crucial for locating studies where load forecasting facilitates grid planning for high demand industrial sectors or complex "systems" (e.g., integrated energy systems).

This search string was applied to the article titles to ensure a broad yet targeted retrieval of relevant publications. While residential and macro-grid studies exist, this review specifically targets the "industrial loads" subset of grid planning. We argue that accurate industrial forecasting is a critical yet distinct component of broader grid planning necessitated not only by high energy intensity of manufacturing systems but also by emerging challenges [6,16–22]. Many articles were removed based on a citation threshold of at least five, except for recent publications (2023–2025), where the criterion was relaxed due to the limited citation accumulation time. Three main criteria are used to filter the papers: (a.) whether the topics relate to energy domain, (b.) whether modeling techniques are a substantive part of the paper, and (c.) whether there are duplicate papers or papers discussing similar ideas or contents.

2.3. Inclusion and Exclusion Criteria

The following are the inclusion and exclusion criteria that have been applied to ensure the quality and relevance of the studies.

Inclusion Criteria

I1.   Papers published in the last 15 years (2010–2025) are taken into account to make sure we stick to the recent developments in the industry.

I2.   Papers written only in English are considered.

I3.   A minimum threshold of more than 5 cites is taken into account.

Exclusion Criteria

E1.   Non-peer reviewed works, studies on non-real time forecasting, or those lacking methodological detail.

E2.   Papers that did not focus primarily on the energy sector.

2.4. Retrieval of Results

The initial search retrieved 872 results from Google Scholar, 200 from Scopus, and 885 from Science Direct. They are filtered down to 72 relevant papers based on the inclusion and exclusion criteria detailed in Section 2.3. After eliminating duplication and applying quality criteria, a total of 58 papers are selected for review.

3. Results

Following the SLR, this section presents the comprehensive findings of the study, examining the information regarding inputs, outputs, and applications in predictive modeling.

3.1. Classification of Models and Methods

Developing robust models to predict load profiles typically involves three key steps: (i.) extracting related attributes; (ii.) analyzing influencing factors; and (iii.) selecting appropriate forecasting models. This literature review explores architectures employed in load profile prediction, categorized into statistical, ML, DL, and hybrid approaches. All the models listed in the Tables 2–5 employ fixed (static) length forecasting horizons. Approaches capable of dynamic output length prediction remain scarce and are thus discussed separately in Section 4.

3.1.1. Statistical Models

A statistical model is a mathematical relationship between random variables, which can change unpredictably, and non-random variables that remain consistent or follow a deterministic pattern [5,7,23]. By employing statistical assumptions, these models make inferences about the fundamental mechanisms that generate the data and the relationships among the data points [23]. The primary goal is to optimally guide decision making by using this inferred information to maximize benefits or minimize negative consequences under uncertainty. Sophisticated models ensure that statistical inferences are logically consistent and can be extended to new situations, such as predicting future data or inferring properties across different experimental setups [23].

While statistical frameworks like the Seasonality Analysis of Electricity Consumption (SAECC) provide a baseline for identifying irregularities, comparative studies demonstrate that computational intelligence models utilizing Z domain normalization can achieve a relative improvement in false positive and false negative rates of over 42% compared to purely statistical approaches [24]. They promote transparency by quantifying uncertainty, such as through confidence intervals [2,25]. Statistical models explicitly represent the underlying physical processes; therefore, they are often referred to as "black-box" methods [13].

Common techniques within statistical modeling include the following:

1. **Linear regression (LR)**: Explains the linear relationship between dependent and independent variables. Elastic Net is a linear regression method that includes regularization [1,2].
2. **Auto Regressive (AR) and Moving Average (MA)**: The AR model is a linear regression of the current value based on one or more previous values, while the MA model is linear regression that regresses the current value against white noise or errors of one or more past values [2,9,26].
3. **Auto Regressive Integrated Moving Average (ARIMA)**: Combining AR and MA components, with ARIMA additionally handling non-stationary data. The common approach to establish ARIMA models is the Box–Jenkins methodology which best

fits the behavior of the time series [13]. It consists of model identification, parameter estimation, model validation, and forecasting [15].

4. **Exponential Smoothing**: Used to pre-process data by mitigating noise [27]. Other techniques such as Synchro squeezing Wavelet Denoising (SWT), are similarly employed to remove high frequency noise, which lays foundation for accurate forecasting [14].

5. **Spearman Rank Order Correlation Coefficient (SROCC)**: It is a non-parametric statistical method used to analyze the non-linear relationship between variables, with the primary purpose of quantifying the degree of influence of each parameter on another [14].

6. **Principal Component Analysis (PCA)**: A method used for dimensionality reduction and solving multicollinearity, and feature extraction/selection techniques aimed at identifying important variables [1,3].

Statistical models have been recognized for their suitability in very short-term forecasting (seconds to minutes), particularly in the prediction of wind power [13]. Min Yu et al. employ several techniques for analyzing and preparing data for building cooling load (CL) forecasting. The autocorrelation function (ACF) was used to calculate the temporal correlation within historical cooling load time series. The ACF analysis revealed distinct periodic patterns in the data, which helped to determine the appropriate time intervals for the historical load to be used as reference factors, increasing precision and reducing computational burden [14]. Table 2 presents various studies in the literature in which statistical models have been applied effectively. Statistical models have relatively low computational costs [2]. In addition, statistical methods perform well in short-term predictions. For example, an ARMA model showed a 20% error reduction in persistence for a 10 h wind speed forecast [13]. A proposed hybrid statistical method demonstrates small mean relative error in multi-step forecasting, which outperforms classical time-series and backpropagation network methods. Thus, statistical approaches, particularly hybrid ones, are robust in handling 'jumping data' [15].

Due to their reliance on simpler formulas, statistical models are considered prone to overfitting, failing to capture the complexity of data [1]. They struggle with limitations, such as the reliance on assumptions about the true distribution of the data, which might almost certainly be incorrect in reality. Over-reliance on 'goodness-of-fit' measures without ensuring that the model is statistically well-specified can lead to unreliable inferences [28]. For example, the persistence method is explicitly noted to be unsuitable for long predictions [2]. Although some statistical models can function with less data, others (categorized as the 'learning approach method') require a significant amount of historical time-series data to perform effectively [13]. Therefore, their limitations, especially with complex non-linear data and longer forecasting horizons, often lead researchers to explore more advanced machine learning approaches [2].

Table 2. Overview of identified literature for statistical models.

Year	Author	Input	Output	Accuracy	Model Type
2010	H Liu et al. [15]	Time-series wind speed wind power	Forecast wind speed and power	11.34% (MAPE) [short-term multi-step ahead]	Wavelet, ITSM, ARIMA, Box–Jenkins
2012	A.M. Foley et al. [13]	Wind and weather data, forecast	Wind power/Speed forecast	10–15% (MAE) [short term]	Persistence, AR, MA, ARMA, ARIMA, u.a.
2014	S. Tasnim et al. [26]	Past wind data, generated power	Wind power forecast	~0.16 (MAE) [short term]	Linear Regression (LR)
2015	A. K. Nayak et al. [29]	Historical wind data, air density, power coefficient	Wind forecast, ramps	136.103% (MAPE) [short term]	ARIMA

Table 2. *Cont.*

Year	Author	Input	Output	Accuracy	Model Type
2017	C. Deb et al. [27]	Energy consumption, occupancy, timestamp	Energy, cooling/heating load, temp., electricity price	1.05–2.59% (MAPE) [short, medium, and long term]	ES, ARMA, ARIMA
2020	M. Zaimi et al. [30]	Meteorological, PV-data, PV-parameters	PMPP, efficiency, I–V-curve, model parameters	<3% (NRMSE) [short term]	Non-linear fit, empirical, polynomial formula
2020	R. Ahmed et al. [31]	Meteorol. parameters, PV time series, timestamp	PV-output/radiation forecast	2–17% (NRMSE) [short, medium, and long term]	EWMA, ARMA, ARIMA
2022	D.V. Pombo et al. [3]	Power production, weather measurements	Short term PV-forecast	20.63% (MAPE, SPAR model), 32.07% (MAPE, Persistence model) [short term]	Persistence, SPAR
2022	C. Wang et al. [32]	Multi-load data, forecast date	Multi-energy load last 24 h	1.30 % (MAPE) [short term]	ARIMA, VAR, GRA, PCC
2022	W.H. Chung et al. [1]	Heat load variables, time, weather forecast	Short-term forecast load	2.60 (MAE) [short term]	Elastic Net
2023	M. Yu et al. [14]	Cooling load, time factors, meteorol. factors	Cooling load forecast (buildings)	statistical benchmark not included [short term]	ARIMA, SROCC, ACF
2024	T. G. Grandón et al. [33]	Electricity consumption, meteorol., economic., calendar	Electricity demand forecast	430.8 MW (MAE, LR) [medium term]	LR, ARIMA, ARMA
2025	W. Liao et al. [2]	Time series, load data	Load forecast	4.4% (MAPE, LR) [short term]	LR, Persistence

3.1.2. Machine Learning Models

Machine learning is a core component of artificial intelligence and a key pillar of the Fourth Industrial Revolution (4IR or Industry 4.0) that empowers algorithms to identify patterns in training data and deliver predictions on new data without explicit programming [34]. By optimizing a model's performance on a dataset that mirrors real world tasks through a process called model training [34,35].

The strength of ML lies in its ability to learn from large volumes of data, underpinning its applicability across diver energy domains [2,4,11]. This versatility extends to data scarce scenarios through frameworks like Physics-Informed Neural Networks (PINNs), which integrate physical constraints for robust handling of temporal data, while prototypes like reinforcement learning further strengthen decision making through feedback mechanisms [34,36,37].

Chung [1] considered short-term heat load forecasting for district heating systems, which is essential to optimize the operations of combined heat and power plants. They used a dataset containing weather information, time factors, and historical heat load data. ML models such as k-Nearest Neighbors (KNN), a simple learner that makes predictions by averaging nearby data points, serve as a baseline [1,9,31]. Advanced models included Support Vector Regression (SVR), known for its robustness in high-dimensional spaces, as well as Random Forest (RF) and XGBoost, which enhance accuracy [1]. Liao [2] particularly addressed scenarios with scarce historical data. Although the primary focus is on large time-series models, several established traditional and ML benchmarks are included, such as the persistence model (PM), which predicts by simply copying the value from the previous step, and linear regression, which is used to estimate future load values. The study also employs Regression Tree (RT) models and XGBoost, utilized for forecasting peak power demand and long-term electricity consumption. Dokur [38] presented an Extreme Learning Machine (ELM), which is characterized as a neural network with a single hidden layer. In ELM, the weights and biases of the input layer are randomly

initialized and remain fixed, whereas the output weights are determined by solving a linear system. This unique architecture contributes to the notable rapid learning speed, ease of implementation, and excellent generalization capabilities of ELM in forecasting tasks. To enhance performance, the study integrates it with the Single-Candidate Optimizer (SCO) to refine its initial parameters, resulting in the ELM-SCO model. The paper also compares this model against other ELM variants, including Meta-ELM, which uses several ELM's trained on distinct data subsets, and Multi-Layer Meta-Kernel-ELM (ML-Meta-KELM), which incorporates kernel functions to enhance stability and generalization. Additionally, the studies include MLP, a traditional feed-forward neural network, as a benchmark. These models utilize time series of power measurements from smart meter datasets to forecast node voltages in low-voltage networks with high penetration of low-carbon technologies. Faizan and Afgan [39] leveraged an MLP architecture to create a robust predictive model. The networks hidden layers utilize the Hyperbolic Tangent Sigmoid (HTS) function to capture non-linear relationships. To optimize the weights and biases of the MLP model and address multi-objective optimization problems, a Multi-Objective Genetic Algorithm (MOGA) is implemented. MOGA is a population-based search mechanism that uses genetic operations such as crossover, mutation, and selection to explore a broad solution space, with the aim of avoiding local optima and achieving global optimization by effectively maximizing and minimizing outputs. The study highlights that MOGA outperformed other methods, such as Bayesian optimization, in terms of error minimization and robustness. Alternative methods, such as decision trees, SVMs, and Particle Swarm Optimization (PSO), were considered less appropriate due to their limitations in regression, multi-objective optimization, and the handling of continuous variables.

Despite these advancements, ML is often limited by its interpretability and reproducibility as well as bias and noise introduced during model development. These constraints led to the design of DL models, by using automatic feature extraction, have achieved considerably better performance and scalability in sequential data modeling [35,40].

Table 3. Overview of identified literature for machine learning category.

Year	Author	Input	Output	Accuracy	Model Type
2012	Aoife M. Foley, Paul G. Leahy, Antonino Marvuglia, Eamon J. McKeogh [13]	Historical wind data, wind power production, NWP forecast values, Weather forecast data	Wind power generation/output/patterns, Forecasted wind speed	10–15% (MAE) [short term]	Multi-layer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Bayesian methods
2016	HA Azzeddine, Mustapha Tioursi, Djamel Eddine C, Brahim K [41]	Cell temperature, Solar irradiance	Current, Voltage, Maximum power point (MPP) of a photovoltaic panel	0.03 (MSE) [short term]	Radial Basis Function (RBF) neural network
2017	Weicong Kong et al. [42]	Household power consumption time series	Short-term residential electric energy forecast	8.18% (MAPE, aggregated) [short term]	BPNN, SVM, Extreme Learning Machine (ELM), Adaptive Neuro Fuzzy Inference System (ANFIS), Radial Basis Function (RBF), Decision Tree (DT), Bayesian Neural Network
2019	Tae-Young Kim, Sung-Bae Cho [43]	Household power consumption: time variables, sensors, submetering	Residential electric energy consumption (Global Active Power)	31.84% (MAPE) [short/medium term]	SVR, Random Forest (RF), Decision Tree (DT), Multi-Layer Perceptron (MLP)
2020	Liufeng Du, Linghua Zhang, Xu Wang [44]	Historical load data	Load forecasting	2.14% (MAE) [short term, one hour ahead]	SVR, Extreme Learning Machine (ELM), Stacked Denoising Autoencoder (SDAE)
2020	Kuihua Wu et al. [45]	Electric load, Temperature, Gas consumption, Cooling load	Short-term electric load forecasting	4.78% (MAPE) [short term]	BPNN, RF Regression (RFR), SVR
2020	PW Khan, Yung-Cheol Byun, Sang-Joon Lee, Dong-Ho Kang, Jin-Young Kang, Hae-Su Park [6]	Historical power consumption data, Time-series data	Energy consumption forecasting	5.77% (MAPE) [short term, one hour]	Support Vector Machine (SVR), Lasso, Ridge, GradientBoost, XGBoost, MLP Regressor, CatBoost
2022	Huafeng Xian, Jinxing Che [11]	Historical load data	Power load forecasting	3.37% (MAPE) [short term]	BPNN, SVR (RBF kernel), MSC framework for optimization
2022	Won Hee Chung, Yeong Hyeon Gu, Seong Joon Yoo [1]	Heat load variables, Time factors, Weather forecasts	Short-term load forecasting	0.587 (MAE) [short term]	k-Nearest Neighbors (k-NN), SVR, RF, XGBoost
2022	Chen Wang et al. [32]	Multi-energy load data, Forecast date/timestamp/type	Multi-energy load forecasting (24 h)	4.02% (MAPE) [short term]	Extreme Learning Machine (ELM), RF, SVR, Bagging-Boosting Neural Network
2022	Daniel Vázquez Pombo et al. [3]	Power production and meteorological measurements	Short-term PV power forecasting	11.77% (RMSE) [short term, five hour]	RF, SVR
2023	Ke Li et al. [46]	Multi-energy load data, Meteorological factors	Short-term multi-energy load forecasting	3.64% (MAPE) [short term]	SVM, Least Squares SVM (LSSVM), Generalized Regression Neural Network (GRNN)

Table 3. *Cont.*

Year	Author	Input	Output	Accuracy	Model Type
2023	Jintao He, Lingfeng Shi, Hua Tian, Xuan Wang, Xiaocun Sun, Meiyan Zhang, Yu Yao, Gequn Shu [47]	State variables (temperature, pressure, mass flow rates), Input variables (cold source inlet mass flow rate, compressed rotational speed, pump rotational speed), Disturbance variables	Net power and refrigerating capacity of the CO_2 combined cooling and power cycle (CCP) system	0.0023 (RMSE) [short term]	Multilayer Feedforward Neural Network (MLFF), Radial Basis Function (RBF) neural network, Generalized Regression Neural Network (GRNN)
2023	Md Shazid Islam, A S M Jahid Hasan, Md Saydur Rahman, Jubair Yusuf, Md Saiful Islam S, Farhana Akter Tumpa [48]	Meteorological data: DNI, DHI, GHI, temperature, wind direction, wind speed	Solar power generation, predicted as a classification problem	81.02% (Classification) [short term]	Ensembles: Adaboost Classifier, Gradient Boosting Classifier, Random Forest Classifier
2023	Connor Scott, Mominul Ahsan, Alhussein Albarbar [49]	PV generation history, Time variable, Meteorological vars.	Forecasted PV power output	32.0 (RMSE) [short term]	Linear Regression (LR), RF, SVM, Neural Networks (NN)
2024	L.R. Visser et al. [50]	Meteorological vars., Lagged PV gen., Prob. forecasts, Market prices	Forecasted PV power generation	4.8 (CRPS) [short term, day ahead]	Multiple Linear Regression (MLR), RF, Smart Persistence (SP), Quantile Regression (QR), Quantile RF (QRF), Clear Sky Persistence Ensemble (CSPE)
2024	Lionel P. Joseph et al. [9]	Meteorological vars., Predictors, Ground level, Satellite climate vars.	Hourly wind speed forecasting	0.421 m/s (MAE) [short term, one hour]	RF, Decision Tree Regressor (DTR), Gradient Boosting
2025	Wenlong Liao et al. [2]	Time-series datasets, Scarce historical load data	Load forecasting	6.1 % (MAPE) [short term, 1 h–24 h]	Regression Tree (RT), XGBoost, MLP, Time-series large language model (TimeLLM)
2025	Emrah Dokur et al. [38]	Active/reactive power measurements, Past voltage	Forecasted node voltage	0.0019 (MSE) [short term]	Extreme Learning Machine (ELM) Variants, MLP, ANFIS

3.1.3. Deep Learning Models

Deep learning, a subset of ML, has become the dominant AI model architecture over the past decade, as demonstrated across diverse application domains [35]. The capability to extract latent features and demonstrate strong adaptability to complex non-linear relationships, which contributes to high forecast accuracy. A key advancement is the ability to leverage pre-trained knowledge from massive and diverse datasets, allowing models to perform effectively even with limited task-specific data through a fine-tuning process [2]. Some DL model types mentioned in the Table 4 are as follows:

1. **Artificial Neural Network (ANN)**: As a foundational element within the broader field of DL, ANNs are structured with input, hidden, and output layers [6,8,31]. They process input data through weighted connections and activation functions like ReLU, learning complex patterns via backpropagation and optimization (e.g., Adam solver). These models are increasingly employed for prediction tasks, such as determining if power delivery network design violates its target impedance, without needing additional simulations during optimization processes [51].

2. **LSTM**: A specialized type of Recurrent Neural Network (RNN) that is highly effective for processing sequential data and capturing temporal features from time series [2,52]. Studies have shown that LSTMs often demonstrate superior performance over conventional model such as MLPs in tasks like short-term load forecasting [2].

3. **Gated Recurrent Unit (GRU)**: Another well-known RNN model that is considered a simplified version of the LSTMs is GRU, possessing a more streamlined architecture with fewer gates (typically two: a reset gate and an update gate), which can lead to faster training times while still maintaining competitive performance [2].

4. **Convolutional Neural Network (CNN)**: It excels in extracting spatial and temporal features from data, identifying local patterns through convolutional and pooling layer architecture [2,53]. Originally popular in image processing, their application has expanded to time-series analysis, where they can identify local patterns and relationships within data segments [1,3,14]. In the context of load forecasting, CNNs are specifically used to depict spatial features between loads at different points in a system, such as various buses in a power network [1,2,53,54].

5. **Graph Neural Network (GNN)**: These are specialized models designed to operate on structured graph data, which are characterized by nodes and edges representing entities and their relationships, respectively. They are particularly adept at capturing both structural and temporal information within complex networks [2,53]. This makes them highly suitable for applications where data exhibit relational structures, such as heating networks, smart grids, or traffic flow systems [2].

6. **Transformer**: Unlike traditional RNNs which are inherently sequential and struggle with parallelization and long-term dependencies, the transformers are entirely based on attention mechanisms [2,32]. To manage information flow, prevent gradient degradation, and accelerate convergence, residual connections and layer normalization are integrated around each sub-layer. Since the model has no recurrence, positional encodings (e.g., sine-cosine functions) are added to the input embeddings to inject information about the order of token in the sequence [32,55].

Despite their power, large pre-trained time-series models such as Time GPT often generalized poorly to new domains when the pre-training data do not match the target distribution, exhibiting weak performance [2]. Thus, there is a necessity for hybrid models that can be developed to exploit the strengths of different techniques, for example, by combining large models with traditional forecasting models to mitigate these underlying distributional differences and increase overall predictive accuracy.

Table 4. Overview of identified literature for deep learning models.

Year	Author	Input	Output	Accuracy	Model-Type
2014	V. Lo Brano, G. Ciulla, M. Di Falco [8]	Air temp.; cell temp.; solar irradiance; wind speed; open-circuit voltage; short-circuit current	PV module power forecast	0.05–1% (Mean error) [short term]	RNN; Gamma Memory (GM)
2018	S. Bouktif, A. Fiaz, M.A. Serhani [4]	Energy-consumption data; time lags; weather; schedule vars	Short/medium-term load forecast	0.56% (RMSE) [short and medium term]	LSTM; RNN
2019	N. Jinil, S. Reka [56]	EV motor power req.; other component loads; driving conditions	EV power-req. prediction; distribution optimization	4.10% (MAPE) [long term]	Modular RNN (MRNN)
2019	C.M. Schierholz, K. Scharff, C. Schuster [51]	Simulated PCB-variation data	Binary TI-violation prediction	88% (Classification) [static design]	Multi-layer ANN
2019	L. Du, L. Zhang, X. Wang [44]	Historical load data	Load forecasting	2.14% (MAE) [short term, hour ahead]	3D-CNN-GRU
2019	T.Y. Kim, S.-B. Cho [43]	Household consumption dataset (time, sensors, submeter vars)	Residential energy-consumption prediction	31.83% (MAPE) [short term]	CNN; LSTM; GRU; Bi-LSTM; Attention LSTM
2021	H. Pariaman, G.M. Luciana, M.K. Wisyaldin, M. Hisjam [52]	Historical time-series sensor data	Reconstructed time-series patterns	93.36% (MAE) [short term]	LSTM-Autoencoder
2022	H. Xian, J. Che [11]	Historical load data	Power-load forecasting	3.025% (MAPE) [short term]	LSTM; RNN
2022	W.H. Chung, Y.H. Gu, S.J. Yoo [1]	Heat-load vars; time factors; weather forecasts	Short-term load forecast	94.2% (R^2) [short term]	DNN; RNN; LSTM; LSTM Attention
2022	Z. Gao, J. Yu, A. Zhao, Q. Hu, S. Yang [57]	Load data; internal/external disturbances	Short-term cooling forecast	3.25% (MAPE) [short and medium term]	ELM; GRNN; BP; WNN
2022	D. Niu, M. Yu, L. Sun, T. Gao, K. Wang [58]	Cooling, heat	electric load data; time features; external factors and Multi-energy-load forecast	5.44% (MAPE) [short term]	LSTM; BiGRU
2022	C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, K. Zhang [32]	Multi-energy load data; forecast date/timestamp/type	Next-24 h multi-energy forecast	1.037% (MAPE) [short term, day ahead]	Multiple-decoder Transformer
2022	Y. Guo, Y. Li, X. Xuebo [59]	Multi-energy load; meteorological; date info	Combined heating, cooling and electric forecast	1.76% (MAPE) [short term]	LSTM; BiLSTM; MTL
2022	D. Vázquez Pombo, P. Bacher, C. Ziras, H.W. Bindner, S.V. Spataru, P.E. Sørensen [3]	PV production and meteorological data	Short-term PV forecast	14.03% (RMSE) [short term, 5 h]	CNN; LSTM
2023	W. Cui, W. Yang, B. Zhang [60]	Time-series data (voltage magnitude, rotor angle, frequency deviation), system topology, fault locations/types, power injections (active and reactive)	Predicted trajectories; unstable-case identification	0.01% (Relative MSE) [transient, seconds]	DNN (Fourier-transform + filtering layers)

Table 4. *Cont.*

Year	Author	Input	Output	Accuracy	Model-Type
2023	J. He, L. Shi, H. Tian, X. Wang, X. Sun, M. Zhang, Y. Yao, G. Shu [47]	CCP system params (temp., pressure, mass-flow), disturbance vars (torque, speed, exhaust), inputs (mass-flow rate, compressor and pump speeds)	Net power; refrigerating capacity; state-variable prediction	0.13% (RMSE) [transient, seconds]	MLFF NN; RNN; LSTM; GRU
2023	Y. Huang, Y. Zhao, Z. Wang, X. Liu, H. Liu, Y. Fu [61]	District-heating consumption data	Multi-horizon district-heat forecast	31.2% (RMSE) [medium /long term]	TPA; LSTM; CRNN; Encoder; MSL
2023	M. Yu, D. Niu, J. Zhao, M. Li, L. Sun, X. Yu [14]	Cooling-load data; time	Short-term cooling forecast	Not specified	LSTM; Bi-LSTM; DNN; RNN; CNN; TTGAT-GTC
2024	X. Wang, H. Wang, S. Li, H. Jin [54]	Historical meter data	Real-time short-term load forecast	0.92% (MAPE) [short term]	LSTM; LSTM + Att; BiLSTM + Att
2024	Y. Huang, Y. Zhao, Z. Wang, X. Liu, Y. Fu [53]	Heat-load records; meteorological; exogenous factors	Future heat-load forecast	7.2% (MAE) [short term]	GNNs
2024	L.P. Joseph, R.C. Deo, D. Casillas-Pérez, R. Prasad, N. Raj, S. Salcedo-Sanz [9]	Meteorological predictors; ground and satellite data	Hourly wind-speed forecast	0.149 m/s [short term] (MAE)	LSTM; BiLSTM
2025	W. Liao, S. Wang, D. Yang, Z. Yang, J. Fang, C. Rehtanz, F. Porté-Agel [2]	Time series; scarce historical load data	Load forecasting	2.1% (MAPE) [short term]	Transformer (positional encoding, multi-head attention, CNN)

3.1.4. Hybrid Models

Hybrid models are computational frameworks that combine multiple methodologies such as statistical, ML, neural networks, or optimization algorithms to create a more accurate and robust prediction system [3,4,6,62]. These models leverage the complementary strengths of individual approaches, making them effective for complex problems where single-method solutions struggle to handle the variability and non-linearity present in real-world data [6].

The key strength lies in their ability to integrate diverse techniques such as combining Support Vector Regression, neural networks, boosting methods, and feature selection, they handle both temporal and spatial patterns, account for physical constraints and deal with uncertainties [3]. They also incorporate feature engineering techniques, such as RF for parameter selection and improved optimization algorithms, such as the Improved Parallel Whale Optimization Algorithm (IPWOA) to tune network parameters [57]. For example, the parallel CNNs-LSTM Attention (PCLA) model extracts spatiotemporal characteristics and then intensively learns importance [1]. The synchro squeeze Wavelet Denoising (SWT), the Temporal Trend-aware Graph Attention Network (TTGAT), and the Gated Temporal Convolution layer (GTC) model are used for CL forecasting, considering spatiotemporal coupling and frameworks that utilize feature separation-fusion technology, with improved CNNs and multitask learning (MTL) for multienergy load forecasting [46]. These diverse architectures are designed to capture complex characteristics, local trends, and cross-correlations within the data more effectively. The SWT-TTGAT-GTC model shows significant improvements in R-squared and a significant reduction in RMSE compared to similar recurrent or convolutional networks [46]. Another example discussed by Joseph [9] is the three-phase hybrid Convolutional Bidirectional Long Short-Term Memory (3P-CBiLSTM) model framework, effectively captures both past and future long-term dependencies from historical sequential data. To enhance the capabilities, the model employs the Two-phase Mutation Grey Wolf Optimizer (TMGWO) for robust dimensionality reduction and feature selection. Additionally, a hybrid Bayesian Optimization and Hyper Band (BOHB) algorithm is utilized for hyperparameter optimization, which is crucial to fine-tuning complex "black-box" models to achieve optimal performance.

Individual deep learning models, in spite of their advanced capabilities, exhibit limitations that necessitate hybrid approaches. RNNs are susceptible to vanishing or exploding gradients, affects their ability to learn long-term dependencies, a problem that LSTMs aim to mitigate [1]. However, LSTMs can also struggle to simultaneously capture both long-term and short-term local dependence patterns or to fully account for cross-correlations in multivariate time-series data [14,46]. Integrated energy system forecasting often extends single-load approaches, which do not learn multi-energy coupling information or process disparate input features uniformly, thus introducing noise [46]. Hybrid models are therefore essential as they combine the strengths of various architectures and integrate mechanisms such as attention or advanced optimizers to overcome these individual limitations, resulting in more accurate, robust, and comprehensive forecasting solutions [1].

Based on the distinct yet synergistic advantages outlined in Table 6, it is evident from the data in Table 5 that CNN, LSTM, and BiLSTM are the most frequently utilized algorithms in hybrid frameworks, corresponding to the model types illustrated in Figure 2. Despite their strengths, hybrid models can be limited by increased complexity, demand for computational resources, and the need for careful parameter tuning and integration [3,11]. Sometimes, the model might also suffer from overfitting if not properly validated, and their interpretability can decrease as the architecture becomes more intricate, especially with extensive ensemble systems [4].

Table 5. Overview of identified literature for hybrid models (shortened for visualization).

Year	Author	Input	Output	Accuracy	Model Type
2018	S. Bouktif, Ali Fiaz, Mohamed Adel Serhani [4]	Electric energy consumption data, Time lags, Weather data, Schedule-related variables	Forecasted electric load/consumption (short- and medium-term horizons)	0.62% (RMSE) [short and medium term]	Genetic Algorithm (GA)- Enchanced LSTM-RNN
2020	PW Khan, Yung-Cheol Byun, Sang-Joon Lee, Dong-Ho Kang, Jin-Young Kang and Hae-Su Park [6]	Historical power consumption data, Time-series data	Energy consumption forecasting	4.29% (MAPE) [short term]	Ensemble thrree base models 1. CatBoost 2. SVR3. MLP
2020	Kuihua Wu, Jian Wu, Liang Feng, Bo Yang, Rong Liang, Shenquan Yang, Ren Zhao [45]	Historical electric load, Temperature, Gas consumption, Cooling load	Short-term electric load forecasting	99.1% (R^2) [short term]	Attention-based CNN-LSTM-BiLSTM model
2022	Huafeng Xian, Jinxing Che [11]	Historical load data	Power load forecasting	2.71% (MAPE) [short term]	MSC-PSO-SVR, Ensemble model (RF and XGBoost)
2022	Won Hee Chung, Yeong Hyeon Gu, Seong Joon Yoo [1]	Heat load-derived variables, Time factors, Weather forecasts	Short-term load forecasting	94.2% (R^2) [short term]	Parallel CNN-LSTM Attention (PCLA)
2022	Daniel Vázquez Pombo, Peder Bacher, Charalampos Ziras, Henrik W. Bindner, Sergiu V. Spataru, Poul E. Sørensen [3]	Basic dataset including power production and meteorological measurements	Short-term Photovoltaic (PV) power forecasting	18.65% (MAPE) [short term]	CNN-LSTM
2022	Dongxiao Niu, Min Yu, Lijie Sun, Tian Gao, Keke Wang [58]	Historical cooling, heat, and electrical load data, Time features, External influencing factors	Short-term multi-energy load forecasting	2.75% (MAPE) [short term]	CNN-BiGRU, BiGRU-Attention CNN-BiGRU-Attention- Multi Task Learning (MTL)
2023	Min Yu, Dongxiao Niu, Jinqiu Zhao, Mingyu Li, Lijie Sun, Xiaoyu Yu [14]	Historical cooling load (CL) data, Time factors, Meteorological factors	Short-term building cooling load (CL) forecasting	8.81% (MAPE) [short term]	SWT (Synchrosqueezing Wavelet Denoising), TTGAT (Temporal Trend-aware Graph Attention Network), GTC (Gated Temporal Convolutional Layer) SWT-TTGAT-GTC Model
2024	Ke Li a, Yuchen Mu a, Fan Yang a, Haiyang Wang a, Yi Yan b, Chenghui Zhang [63]	Uncertain variables in an Integrated Energy System (IES), Meteorological data	Joint source-load-price forecasting	4.10% (MAPE) [short and long term]	MCNN-SCAM-LSTM-MTL where, MCNN- Multi-column Convolutional Neural Network, SCAM- Sequential Convolution Attention Module MTL-BiLSTM, Radial Basis Function Deep Belief Network (RBF-DBN), MTL-LSSVM
2024	Sujan Ghimire, Ravinesh C. Deo, David Casillas-Pérez, Sancho Salcedo-Sanz [64]	Half-hourly electricity price sequences, Lagged values of the decomposed price series, Historical errors	Short-term, half-hourly electricity price forecasts	5.83% (sMAPE) [short term]	VMD-CLSTM-VMD-ERCRF model VMD: Variational Mode Decomposition CLSTM: combined of CNN and LSTM, ERCRF: Error compensation and Random Forest regresssion

Table 5. *Cont.*

Year	Author	Input	Output	Accuracy	Model Type
2024	Jungyeon Park, Estêvão Alvarenga, Jooyoung Jeon, Ran Li, Fotios Petropoulos, Hokyun Kim, Kwangwon Ahn [65]	Hourly electricity consumption series, Deseasonalized demand time series, Historical demand observations	Probabilistic load forecasts	60% (Error Red.) [short term]	ARMA-GARCH (Autoregressive Moving Average - Generalized Autoregressive Conditional Heteroskedasticity) model
2024	Yaohui Huang, Yuan Zhao, Zhijin Wang, Xiufeng Liu, Yonggang Fu [53]	Historical heat load records, Meteorological factors, Exogenous factors	Future heat oad values, Forecast of time steps ahead	23.5% (RMSE) [short term]	Sparse Dynamic Graph Neural Network (SDGNN)
2024	Lionel P. Joseph, Ravinesh C. Deo, David Casillas-Pérez, Ramendra Prasad, Nawin Raj, Sancho Salcedo-Sanz [9]	Meteorological variables, Attributes used as predictors, Ground level data, Satellite based climate variables	Hourly wind speed forecasting	99.5% (Index) [short term]	3 Phase hybrid model: 3P-CBiLSTM 1. TMGWO (Mutation Grey Wolf Optimizer) for feature selection 2. BOHB (Hybrid Bayesian Optimization and HyperBand) algorithm for hyperparameter optimization 3. CBiLSTM
2025	Ali Amini, Samuel Rey-Mermet, Steve Crettenand, Cécile Münch-Alligné [66]	High-frequency experimental data, Low-frequency SCADA data, Physics-based parameters, Engineered features	Instantaneous Power	99% (R^2) [real time]	physics-based analysis and data-driven (machine learning) approach
2025	Emrah Dokur, Nuh Erdogan, Ibrahim Sengor, Ugur Yuzgec, Barry P. Hayes [38]	Time series of active power measurements, Time series of reactive power measurements, Past voltage values	Forecasted node volatage	0.56% (Avg. Dec.) [near real time]	It combines Extreme Learning Machine (ELM) with Single Candidate Optimizer (SCO)
2025	Weikun Deng, Hung Le, Khanh T.P. Nguyen, Christian Gogu, Kamal Medjaher, Jérôme Morio, Dazhong Wu [67]	Raw sensor data, Continuous time-series data, Temperature data	Remaining Useful Life (RUL) prediction for fast-charging lithium-ion batteries	8.4% (MAPE) [long term]	Data driven branch: Dilated Convolutional Neural Network (D-CNN) Physics-informed branch: Neural network that uses physics-embedded algorithm structure Features from both branches are merged and processed by a Full Connectivity Neural Network (FCNN) in the final output layer
2025	Hui Song, Boyu Zhang, Mahdi Jalili, Xinghuo Yu [68]	Energy demand data, Temperature data	Energy demand forecasting	5.8% (RMSE) [short term]	Multi-swarm Multi-tasking Ensemble Learning (MSMTEL)
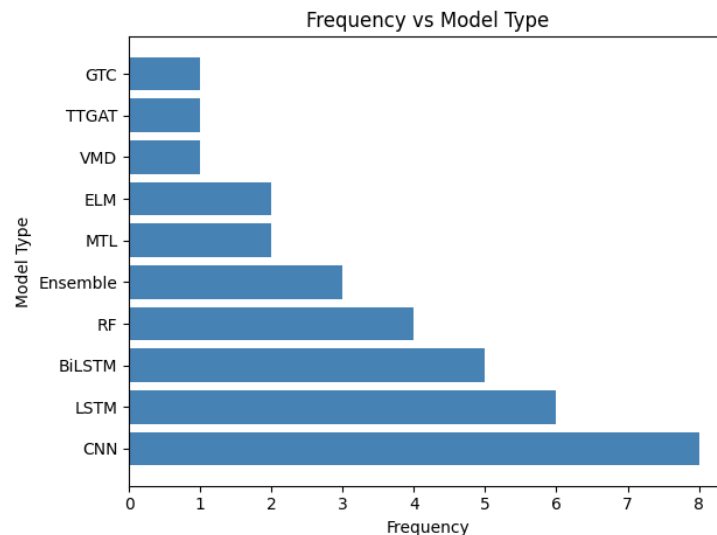
Figure 2. Frequency of top 10 hybrid model components used in load profile prediction literature (2011–2025).

Figure 2 summarized the ML models for load profile prediction. Table 6 summarizes the key strengths and weaknesses of the ML models to provide a structured comparison of the effectiveness in different applications. Convolutional Neural Networks (CNNs) offer high efficiency in extracting local features and inherent structures from time-series data, making them parameter-economical, though they are inherently restricted by their fixed temporal receptive field and rigidity to sequence length. The Long Short-Term Memory (LSTM) variants, including Bidirectional LSTMs (Bi-LSTMs), are crucial for time-series analysis, as they are specifically designed to learn and retain long-term dependencies, overcoming the vanishing gradient problem. Although Bi-LSTMs improve learning effectiveness by incorporating context from both past and future data, their complex "black-box" nature is a significant drawback, necessitating explicit explainable intelligence measures to ensure model transparency.

To understand composite methodologies, hybrid models are classified on two primary dimensions such as architecture and functionality.

1. **Architecture** defines the structural integration of components:

   (a) Series (Sequential): These models operate in a pipeline wherein the output of one module serves as the input for the next. This structure is designed to improve the data quality prior to learning [9,62].

   (b) Parallel (Ensemble): "Ensemble methods" or "ensemble learning", where multiple models run in parallel and their predictions are aggregated (bagging, boosting, stacking). These models emphasize on variance reduction, improved generalization, and combing multiple base learners via averaging, voting, or stacking meta-learners [1,6].

   (c) Embedded (Optimization): These models integrate a meta-heuristic optimization algorithm (e.g., GA, PSO) directly into the learning process of a predictor to automatically tune hyperparameter or weights [4,11].

2. **Functionality** describes the logic behind the combination:

   (a) Decomposition-based: Models that first decompose the original signal into a set of simpler subseries (e.g., modes, frequency bands, trends) using techniques such as WT or VMD and then learn based on these components or their recombination to handle the complex oscillatory behavior more constructively [14,62].

    (b)   Feature fusion-based: Models that integrate heterogeneous feature extractors and combine their latent representation through concatenations via attention or learned weighting into a unified feature space that is passed to a downstream predictor, aiming to capture complementary aspects of the data such as spatiotemporal structure (e.g., instead of standard CNN-LSTM combination, Graph Attention Networks (GAT) combined with TCN could be utilized) [14].

Table 6. Strengths and weaknesses of model types mostly used in hybrid setting for predicting sequential load profiles.

Model Type	Strength	Weakness
CNN	They excel in extracting hidden structures and inherent features from time-series data, to improve learning efficiency and reduce the number of parameters [9,31].	They are fundamentally limited to temporal receptive field, destructive flattening in hybrid setups, and rigidity with sequence lengths [1].
LSTM	It is designed to overcome the vanishing gradient problem of traditional RNNs and learn and retain long-term dependencies in temporal sequences [2,3,6,31].	While effective they are limited to pre-processing information in a single direction, meaning they have the capacity to miss out on pertinent information [9].
BiLSTM	They improve upon standard LSTMs by being able to process information in both forward and backward directions, which facilitates the effective learning of due to dual information flow characteristics [1,9,14].	Due to their complex architecture, they are non-interpretable "blackbox" models, which require explainable intelligence to increase model transparency [1,9,14].
RF	Pombo et al. (2022) explicitly state that RF required 1–3 h for training compared to 12–35 h for hybrid CNN-LSTM models [3]. Joseph et al. (2024) stated that tree-based models generally offer better short-term accuracy then physical or statistical models [9].	Tree-based models "perform poorly when extrapolating outside the range of the training data" [9].
Ensemble	Khan et al. (2020) argue that ensemble learning allows weak classifier to correct each other's mistake, resulting in a stronger supervised model [6]. Ensembles reduce the risk of selecting a single model with systematic bias errors [31].	Ensembles increase computational costs because multiple base models must be processed in parallel [31].

As an example, the Table 7 summarize the relative hybrid architectures and the functional reasoning that supports their integration.

Table 7. Hybrid architecture and functional logic.

Model Type	Hybrid Architecture	Functional Logic
PCLA (Parallel CNN-LSTM Attention [1]	Parallel (Ensemble)	Extracts spatial (CNN) and temporal (LSTM) features simultaneously in parallel branches to capture spatiotemporal characteristics before fusion.
Physics-Informed (PV model + RF/SVR/ANN) [3]	Series (Sequential)	Feature fusion with physics-based variables is achieved by concatenating raw meteorological data with derived quantities from a PV performance model prior to training ML models such as RF.
GA-LSTM [4]	Embedded (Optimization)	Integrates GA to optimize the number of time lags and hidden layers for LSTM model.
MSC-PSO-SVR [11]	Embedded (Optimization)	Integrates a new technique called multi-space collaboration (MSC) framework with Particle Swarm Optimization (PSO) to tune Support Vector Regression (SVR) parameters, preventing local optima.
SWT-TTGAT-GTC [14]	Series (Sequential)	Pipeline applies SWT for denoising and improving data quality, followed by graph attention (spatial) and gated temporal convolution (temporal).
ELM-SCO [38]	Embedded (Optimization)	Embed the SCO algorithm to optimize the initial weights and biases of an ELM to prevent overfitting.
Attention based CNN-LSTM-BiLSTM [45]	Series (Sequential)	Operates in a feature fusion pipeline such that CNN extracts features, an attention block assigns weights, and LSTM-BiLSTM forecasts the load.
SDGNN [53]	Series (Sequential)	A pipeline that constructs a sparse dynamic graph, enhances spatio temporal memory via convolution, and fuses features for global forecasting.
CNN-BiGRU Attention (Multi-task) [58]	Parallel (Ensemble)	Uses a hard weight sharing mechanism for multi task learning to share coupling information among cooling, heat, and electrical loads, combined with feature extraction.
VMD-CLSTM-VMD-ERCRF [64]	Series (Sequential)	Decomposition based pipeline where VMD decomposes data for a CNN-LSTM forecast; residual errors are then decomposed and corrected by RF.
ARMA-GARCH [65]	Embedded (Optimization)	Error mitigation using GA to optimize the portfolio (aggregation) of demand to minimize probabilistic forecast error.
MSMTEL [68]	Embedded (Optimization)	Error mitigation using multi swarm PSO to optimize knowledge transfer between tasks, followed by a PSO-optimized weighted ensembles.
MCNN-SCAM-LSTM-MTL [63]	Parallel (Ensemble)	A hybrid attention mechanism combined to enable separate extraction and unified fusion of features. Uses a Multi-Column CNN to extract features fused by SCAM.

3.2. Application in Grid Planning

Strategic integration of renewable sources in rural grid planning has proven essential for resilience; for instance, hybridizing grid connections with biomass generators utilizing local livestock waste can reduce grid dependencies by approximately 55% and lower carbon dioxide and sulfur dioxide emissions by roughly 30% [69]. This is crucial for sequential load

forecasting, which involves making continuous predictions for future time steps, mostly with overlapping horizons, to support dynamic grid operations [16,70].

Effective demand side management is crucial, as electricity demand is projected to increase substantially with electrification. Load forecasting models are essential for peak load management and support Demand Response (DR) strategies by predicting high demand periods [11,14]. This allows customers to shift their electricity needs to times when the power is more abundant or the demand is lower [16]. A 1% reduction in the MAPE value for short-term load forecast can save the utility company approximately USD 300,000 per year [16].

An example of sequential load forecasting that allows for comprehensive assessment of forecast accuracy and reflecting real-world scenarios is the simulated production of prediction of the next 24 h at each time step of the test data, rather than a single prediction every 24 h. Medium-Term Load Forecasting (MTLF) ranges from a few months to 1 year, useful for maintenance scheduling and financial evaluation [70]. Long-Term Load Forecasting (LTLF) covers 1 year to 10 years or more, essential for the planning of power grids and generators and future capacity expansion [70–72]. Therefore, the short-term to long-term continuous nature of these predictions forms a sequential chain of forecasts that supports the entire grid planning and operational cycle.

3.3. Adaptability Across Forecasting Horizons

The adaptability of forecasting models is intrinsically linked to the forecast horizon, as the dominant stochastic factor driving load and generation variance change significantly from minutes to years [13,31]. Forecast errors typically increase as the time horizon [13]. A critical synthesis of the literature reveals that no single architecture excels across all the time scales; rather, model selection must align with the specific temporal characteristics of the input features and the intended grid planning application [2,13]. An overview of relationship between time scales, dominant model inputs, and recommended architectures is summarized in Table 8.

1. **Short-term and intra-hour horizons (real-time operations):** In this domain, horizons range from minutes to 48 h, model adaptability relies heavily on capturing high-frequency fluctuations and rapid meteorological changes [31]. DL models, specifically LSTM networks and CNNs, demonstrate superior adaptability due to their ability to capture non-linear dependencies in volatile time-series data [1,9,13,14]. While LSTMs like Time GPT perform, well in short-look-ahead scenarios, their performance can degrade in longer horizons, often producing "conservative" forecast that fail to capture peaks and valleys necessary for granular operational planing [1,2,36].

2. **Medium-term horizons (Scheduling and maintenance):** Spanning one week to several months, in this horizon "catastrophic forgetting" of older patterns becomes a risk for standard neural networks [4,31]. Techniques such as GA is used to optimize time lags which has proven to be effective in prediction stability by identifying optimal historical windows [31]. Accurate medium forecasts allow utilities to optimize unit commitment and minimize reserve power requirements by predicting weekly load profiles with lower variance [2].

3. **Long-term horizons (Capacity planning):** For horizons extending from one year to decades, statistical and physical models often outperform pure ML approaches in this domain because they simulate atmospheric dynamics and physical boundary conditions, which may not account for long-term climate shifts [36]. However, recent trends advocate hybridizing physical models with ML error correction to enhance long-term validity [13,16,36].

Table 8. Mapping time scales to model adaptability and grid planning applications.

Forecast Horizon	Dominant Input Feature	Recommended Model Architecture	Key Challenges
Intra-hour/ Short-term (<1 h to 48 h)	Historical load, cloud motion, wind speed, temperature [31,71,72].	DL and hybrid models such as CNN-LSTM, BiL-STM, Ensembles [1].	Minimizing latency in data acquisition [3,45].
Medium-term (1 week to 1 year)	Seasonal indices, calendar events, temperature trends [4,70].	Optimized RNNs such as GA-LSTM and statistical methods like SARIMA [4,31].	Ensuring stability and avoiding overfitting to short-term noise [4].
Long-term (>1 year)	Macro economic indicators, demographics, climatological norms [13].	Physical statistical hybrid models [9,13].	Accounting for non-stationary trends [13,63].

4. Discussion

This section outlines the findings of the review with respective to the RQs. First, the input and output characteristics are analyzed, followed by evaluation of the applied ML models and architectures. Subsequently, the implications for sequential load profile prediction in electrical grid planning are derived, highlighting the current limitations and promising future direction. By synthesizing these aspects, the discussion aims to bridge theoretical insights with applicability in sustainable energy systems.

4.1. General Discussion

The research questions guiding the systematic literature review as mentioned in Table 1 are discussed. The key input drivers are categorized as time-series data, system specific parameters, meteorological variables, and operational parameters emerge as essential for enhancing prediction accuracy and robustness. By highlighting how these inputs are transformed yet often under exploit dynamic integration, the analysis underscores the need to decouple output length prediction from the core signal forecasting.

4.1.1. RQ1 (Input and Output Characteristics)

The inputs used can be categorized into four different data streams which serve as the foundations of these models. The following are the inputs identified across the articles:

1. **Time-series data**: This type of data is crucial because electrical load is inherently dynamic and influenced by factors that change over time. In the context of wind power, historical wind speed data is used, from which the one day ahead wind power target is derived using the power curve of specific turbines [9]. This historical wind speed itself forms a time-series input. The raw time-series data can be transformed using techniques such as Fast Fourier Transformation (FFT) to decompose the signal into frequency components [3]. This transformation also enhances the diversity of the inputs for ensemble learning [26]. In terms of historical energy consumption data both from renewable and non-renewable sources, the raw data can be aggregated into a single total energy consumption series, where the model leverages rich temporal features directly from the date time index of the collected data, allowing ML algorithms to learn and forecast energy consumption patterns based on time-specific variations like hour, day of the week, and year. This approach highlights the time-dependent context in which load occurred for accurate forecasting [6].

2. **System specific data**: This category encompasses parameters and specifications unique to the physical and operational characteristic of the power system and associated technologies. These involve core electrical variables such as voltage, current, active/reactive power, system topology, and fault data [8]. Inputs such as cell temperature, solar irradiance, and efficiency are crucial when modeling systems with

       solar integration. Instead of relying solely on physical measurements, some models use a mathematical representation (e.g., the single-diode model) to generate high-quality datasets. These datasets enable accurate offline training of ANNs to predict the maximum power point under varying conditions [2,30,41].

3.   **Meteorological data**: It plays a pivotal role, particular in systems that incorporate renewable energy sources such as solar and wind power. Meteorological data is not only diverse but also exhibits high spatial and temporal variability, meaning same model may perform differently across regions or over time due to changing climate conditions. Solar irradiance and Global Horizontal Irradiation (GHI) are critical variables for Photo Voltaic (PV) power forecasting, which provides insights into potential solar energy available at a given location and time [48].

4.   **Operational parameters**: It provides essential insights into the functioning, configuration, and temporal context of the system being modeled. An example where the inputs capture design specific and runtime characteristics of a Power Delivery Network (PDN) and in context of the ANN model, it includes Target Impedance (TI) which defines the PDN performance threshold, influencing how the ANN assesses design adequacy. Another input which represents the placement values of capacitors on the Printed Circuit Board (PCB), either integrated via ring or detailed grid and rectangular sector methods to enhance model performance is Decoupling Capacitor (Decap). Thus, the inputs are synthetically generated using physics-based simulations, enabling the ANN to predict whether a given PDN design will violate its target impedance. Abstracting spatial data through sector-based pre-processing significantly improves prediction accuracy, demonstrating the importance of a thoughtful input representation [51].

    In forecasting models, the output characteristics are fundamentally determined by how the prediction length is handled, primarily falling into two important concepts: static and dynamic output length. Table 9 provides a clear differentiation between these two approaches and Figure 3 illustrates the forecasting horizons. Especially the identified problem area is highlighted as gap. The Active Graph Recurrent Network (AC-GRN) model, as presented in [61], operates primarily within the static forecasting framework in its current implementation. The model is designed for multi-horizon multi-step district heat load forecasting. This means it predicts heat demands across various predefined future time periods, referred to as "horizons", and provides multiple distinct predictions within those periods, known as "steps". A significant highlight of the paper, which serves as a limitation and future scope towards dynamic forecasting, is that, despite its ability to handle multiple predefined horizons and steps, the model applies a uniform output length across all heat meters. This aligns directly with the "Fixed Length Model" characteristics of static forecasting as described in Table 9.

    The hybrid model addresses dynamic output length by engineering time-series features, from multi scale consumption data, and then training its ensemble (MLP, SVR, CatBoost) on flexible data partitions that reflect real operational scheduling, which enables the model to adaptively predict across varying time horizons without preset limits; this allows the system to continuously tailor and update its forecasting window purely based on process demands and available contextual information, overcoming static constraints through learned temporal flexibility [6].

Table 9. Forecasting comparison.

Aspect	Static Length Forecasting	Dynamic Length Forecasting
Definition	Predicts a predetermined, fixed number of future time steps (e.g., from time t to $t + L$, where L is constant across all instances).	Predicts a variable number of future time steps (e.g., from time t to $t + y$, where y adapts based on instance specific factors like component properties).
Parameter handling	Assumes uniform or averaged parameters (e.g., standard material sizes, processing times) across components.	Incorporate instance specific parameters (e.g., varying dimensions, material quantities, or process durations per component).
Data dependency	Relies on generalized, aggregated, or historical average data; lesser need for fine grained details.	Relies on detailed, instance specific data (e.g., from simulations or measurements for each component).
Training data	Require uniform length sequences (often achieved via padding truncation if raw data varies).	Natively handles variable length sequences (modern approaches use masking; classical method may still pad/truncate).
Model approach	Uses models with fixed output dimensions (e.g., direct multi step prediction with a static horizon).	Uses adaptive models (e.g., auto regressive with stop tokens, encoder-decoder with dynamic decoding, or masking in transformers) which adjust output length per instance.
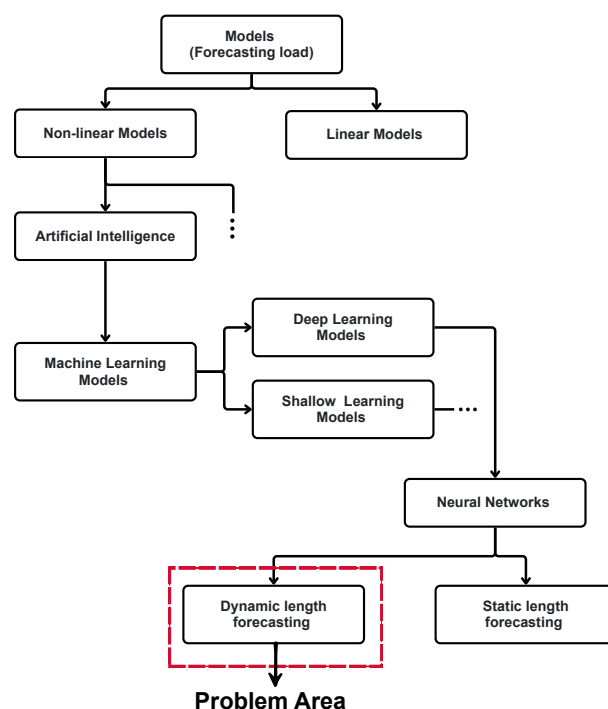
Figure 3. Hierarchical classification of load forecasting models, identifying "Dynamic length forecasting" domain as the primary research gap (highlighted in red).

Dong et al. [73] achieves dynamic output length by employing stochastic state space models and uncertain basis functions, which use recursive filtering techniques such as Kalman filtering and expectation maximization to continually update prediction parameters

and outputs in real time as new data is observed: this adaptive approach allows the forecast duration to shift and expand according to the behavior of solar radiation signals; thus, the horizon is not pre-set but evolves based on dynamic state and real measurement context, ensuring highly responsive and process specific forecasting for each time interval.

4.1.2. RQ2 (Improving Electrical Grid Planning)

The journey to improve electrical grid planning begins with understanding historical trends, a task in which statistical models lay the foundation. Studies like Liu et al. (2010) introduced ARIMA to forecast wind speed and wind power, providing operators with a reliable method to schedule maintenance and plan infrastructure upgrades based on seasonal patterns [15]. As grids grew complex with the integration of renewable energy, the need for adaptive predictions led to the rise of ML models; Hong (2016) pioneered the use of ANNs for probabilistic prediction of electrical loads [74].

The evolution continued with deep learning models designed to handle the intricate real-time data of smart grids. Bouktif et al. (2018) [4] leveraged LSTM networks to construct forecasting models for short to medium term aggregate load forecasting, and further enhanced the performance by using a Genetic Algorithm (GA) to optimize key hyper parameters such as time lags and the number of LSTM layers. An additional experiment has shown that a CNN-LSTM neural network, where the CNN layer can extract features between several variables affecting energy consumption, and the LSTM layer is appropriate for modeling temporal information of irregular trends in time-series components [4].

The pinnacle of this progression is reached with hybrid models, which synthesize the strengths of preceding approaches to tackle the multifaceted challenges of grid management. Liu et al. (2010) blended wavelet transforms with Improved Time-Series Methods (ITSM) and ARIMA to enhance wind power forecasting for small wind farms [15]. Osório et al. (2014) advanced this with a Hybrid Evolutionary Adaptive Approach (HEA), integrating Mutual Information (MI), Wavelet Transform (WT), Evolutionary Particle Swarm Optimization (EPSO), and neural-fuzzy systems for short-term electricity prices prediction and wind power prediction [62]. Khan et al. (2020) combined Multi-Layer Perceptron (MLP), SVM, and Categorical Boosting (CatBoost), using Shapely Additive Explanations (SHAP)-based feature analysis to optimize resource allocation for renewable and non-renewable energy consumption [6]. Finally, Bouktif et al. (2018) applied LSTM-RNN hybrids with statistical methods which have lower forecast errors in the challenging short to medium term electric load forecasting problem compared to the best machine learning [4]. This narrative flow from statistical foundations to hybrid innovations traces the progression and demonstrates a robust and scalable path to improve electrical grid planning and management.

The challenge of separating output length predictions from the core signal prediction is addressed not only by combining existing methodologies (hybrid models) but also by deploying foundation models (Time GPT) and Physics-Informed Neural Networks (PINT) that are designed to leverage learned knowledge or structural constraints to overcome data scarcity and varying lengths and frequencies [2,36].

4.1.3. RQ3 (Performance Metrics)

Forecasting electrical loads is not merely a data problem; it is a strategic imperative for grid reliability, sustainability, and operational efficiency [2,6,30]. This task is inherently complex, which requires not only accurate predictions, but also meaningful metrics to evaluate how well different models perform under various conditions. We focus on the performance metrics used to evaluate forecasting models, explaining their meanings, prevalence, and specific applications. We explore common metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R^2 score, and Mean Absolute Percentage Error

(MAPE), as well as rare metrics like binary classification accuracy, F1 score, and normalized stability indices, to provide a comprehensive comparison of model performance [4,6,9,14,31]

Performance metrics quantify how well a model's predictions align with actual load data and help identify trade-offs between accuracy and computational complexity. MAE measures the average absolute difference between predicted and actual load values, expressed in the same units as load (e.g., megawatts). It is intuitive and robust to outliers, making it widely used for assessing point forecasts. Khan et al. (2020) employ MAE to evaluate a hybrid ML model which forecasts dynamic energy consumption using historical load and time-series features [6]. MAE's interpretability is valuable for comparing model performance within a single dataset, but it does not account for the relative scale of errors, which can be problematic when forecasting loads across diverse systems, such as residential versus commercial grids [9]. RMSE is defined as the square root of the average squared differences between predictions and actual; it emphasizes larger error due to its quadratic formulation. This makes it sensitive to outliers, providing insight into model consistency across volatile load profiles. Osório et al. (2014) employ RMSE to assess a hybrid evolutionary adaptive model for short-term load-related forecasts, leveraging inputs like historical load and weather data [62]. However, RMSE's sensitivity to large errors can lead to the double penalty effect, where models are penalized twice for phase-shifted peak predictions, potentially discouraging accurate peak forecasting in volatile scenarios. R^2 score or the coefficient of determination quantifies the proportion of variance in the actual load explained by the model, ranging from 0 to 1. It is widely used for its scale-independent nature, enabling comparisons across datasets with different load magnitudes [6].

Binary classification accuracy measures the proportion of correct predictions (true positives and true negatives) in tasks where the output is categorical, such as predicting whether a power delivery network violates a target impedance threshold. This metric applies to evaluate an ANN model for power delivery network stability, using simulated PCB design data with inputs like target impedance and port positions [51]. While intuitive, accuracy can be misleading in imbalanced datasets, where a model may achieve high accuracy by predicting the majority class (e.g., stable conditions) while missing rare events (e.g., impedance violations). The F1 score, the harmonic mean of precision and recall, addresses this limitation by balancing the trade-off between correctly identifying positive cases (recall) and minimizing false positives (precision). It employs a related metric to assess an LSTM-autoencoder model for anomaly detection. Thus, F1 score is particularly valuable in imbalanced scenarios, such as detecting rare load anomalies or equipment failures, where high recall ensures most anomalies are identified, and high precision reduces false alarms [75].

Overall, the choice of model depends on the forecasting horizon, data complexity, and constraints including data availability, computational resources, and interpretability. Statistical models like ARIMA remain relevant for simpler linear datasets, but are less effective for complex scenarios [76]. Traditional ML models like SVM and ANN provide a practical balance of accuracy and efficiency, particularly when enhanced with pre-processing techniques like PCA, which reduce the number of parameters and compress the feature search space [3,31]. For short-term load forecasting, DL models, including LSTM, can achieve the highest accuracy but require significant computational resources [4,42]. Hybrid computational frameworks show versatility across various applications, offering robust performance for short- and medium-term forecasts. Future research should focus on optimizing the computational efficiency of deep learning models and developing adaptive hybrid frameworks to better handle the increasing data volatility and complexity inherent in modern power systems.

4.2. Trends and Advancements

Researchers are increasingly applying DL architectures such as Deep Neural Network (DNN), Deep Belief Network (DBN), and RNN to forecast energy consumption in various energy sectors [4,9,31]. An approach that allows a neural network's output to be computed via a black-box Ordinary Differential Equations (ODE) solver, propagating the network through a continuous block of computation with infinitesimally small time steps, and parameterizing hidden layer dynamics using an ODE is a key advancement known as Neural ODEs. Making continuous-time demand forecast model independent of time horizons and addressing data inconsistency problems [77].

The emphasis on incorporating domain knowledge and physical laws into model architectures has seen growth in forecasting models, leading to the development of Physics-Informed Neural Networks (PINNs). It represent an advancement in scientific machine learning by encoding model equations, such as Partial Differential Equations (PDEs), as components of the neural network's loss function. This allows them to solve forward and inverse problems within a unified framework, often without the need for extensive labeled data or fixed meshes [36]. Furthermore, Physics-Constrained Neural Networks (PCNNs), which implement initial or boundary conditions via custom architectures, apply domain decomposition. Continuous development focuses on optimizing activation functions, training procedures, and addressing theoretical challenges such as convergence and robustness to multiscale phenomena [78].

A major advancement in this area is the combination of MTL with GNNs, by interpreting transformers as nodes in the graph. GNNs leverage message passing between nodes to gain dependencies which standard neural networks might fail. The method proposed by Dominik [79] uniquely combines Bayesian Multi-Task Embedding with the GNN architecture, allowing a single model to predict for different transformers while capturing their individual latent characteristics (e.g., load behaviors). Thus, this represents a trend towards data-driven models to learn from topological and interactional complexities.

4.3. Gap and Challenges

Firstly, existing methods restricts the use of neural networks in dynamic prediction because there is no established framework that successfully separates output length prediction from core signal prediction. This paper introduces the first methodology to explicitly decouple these two tasks, proposing the development of specialized dedicated models. Secondly, the development of tailored, application-based accuracy metrics is crucial for reliable integration of processes in industries. This enables targeted quality assurance, enhances end-user trust, and supports regulatory compliance, particularly as industries adopt more complex, digital, and data-driven workflows. Developing forecasting architectures introduces several challenges, a few which are outlined below:

1. **Need for continuous adaptability**: Models require continuous integration feedback, handling variation in parameters, applications, and modalities in real time. When the output length varies, the complexity is greatly amplified [80].
2. **Handling irregularity**: Models must handle these irregular length sequences without relying on padding or truncation, which can obscure critical information or introduce noise [81].
3. **Balancing accuracy**: Sliding window is one common approach in handling sequential data in dynamic forecasting [81]. The comparative optimizations reveal that it is a critical hyperparameter; for instance, expanding the window from 22 to 30 h in hybrid RF and LSTM models was proven necessary to capture specific multi day dependencies in non-stationary solar data [82].

5. Conclusions

This review examines the evolution of electrical load forecasting, which traces the trajectory from foundational statistical models to newly proposed DL architectures, identifying RNNs as those with the best performance in temporal modeling. On the other hand, they are still limited by the vanishing gradient problem as well as their inflexibility when data is multivariate on different time horizons, and thus progressively evolved towards the hybrid architectures, the most common being the use of CNNs to learn the spatial features and the BiLSTM to learn from time-series data, which enables the modeling of the larger range of complexities in modern energy systems.

Despite these advances, a major limitation of the state-of-the-art models is their inability to generalize to other forecasting horizons without wide-ranging retraining. Most proposed models fix the forecasting horizon as a hyperparameter that can be optimized a priori and they are not easily adaptable to variable production and load duration in practice. To address these limitations and enhance grid planning flexibility, future research must focus on two specific actionable domains:

- **Adapting to dynamic signal lengths:** There are currently no algorithms for dynamically adjusting the signal length in a manufacturing environment. Existing models assume constant clock intervals regardless of varying load profiles. Additionally, MAE or RMSE maybe unsuitable and misleading for these results when signals are irregularly sampled. Therefore, further work is needed to examine industrially relevant metrics to best depict the results and their relevance for managing real-time industrial grid applications, and demand response.
- **Decoupling output length from signal prediction:** In future work, the length of the output of the model can be decoupled from the signal prediction task. The model is able to predict how long the future predictions will be based on the length of the forecast horizon. Therefore, two sub-modules for duration prediction and load shape prediction can strengthen the overall algorithm.

Abbreviations

The following abbreviations are used in this manuscript:

SLR	Systematic Literature Review		ANN	Artificial Neural Network
ML	Machine Learning		kNN	k-Nearest Neighbor
DL	Deep Learning		LLMs	Large Language Models
LSTM	Long Short-Term Memory		FFT	Fast Fourier Transform
GHI	Global Horizontal Irradiation		TI	Target Impedance
PDN	Power Delivery Network		LR	Linear Regression
PCB	Printed Circuit Board		Decap	Decoupling Capacitor
SVM	Support Vector Machine		PCA	Principal Component Analysis
GA	Genetic Algorithm		CL	Cooling Load
FCNN	Fully Connected Neural Network		ACF	Autocorrelation Function
MI	Mutual Information		ReLU	Rectified Linear Unit
ELM	Extreme Learning Machine		TCN	Temporal Convolutional Network
GTC	Gated Temporal Convolution		DES	Discrete Event Simulation
CatBoost	Categorical Boosting		RBF	Radial Basis Function
MLP	Multi-Layer Perceptron		MTL	Multi-Task Learning
RNN	Recurrent Neural Network		GNN	Graph Neural Network
SHAP	SHapley Additive Explanations		DBN	Deep Belief Network
ES	Exponential Smoothing		ODE	Ordinary Differential Equations
MA	Moving Average		RMLP	Recurrent Multi-Layer Perceptron
AR	Auto Regressive		WT	Wavelet Transform
SVR	Support Vector Regression		AC-RNN	Active Graph Recurrent Network
FNOs	Fourier Neural Operators		NWP	Numerical Weather Prediction
ITSM	Improved Time-Series Methods		CNN	Convolutional Neural Network
BP	Back Propagation		DNN	Deep Neural Network
MAPE	Mean Absolute Percentage Error		SPAR	Semi-Parametric Auto Regressive
DR	Demand Response		VAR	Vector Auto Regression
GRU	Gated Recurrent Unit		GRA	Grey Relational Analysis
MOGA	Multi-Objective Genetic Algorithm		PCC	Pearson Correlation Coefficient
PINNs	Physics-Informed Neural Networks		MAE	Mean Absolute Error
EWMA	Exponentially Weighted Moving Average		MSE	Mean Squared Error
GRNN	Generalized Regression Neural Network		RMSE	Root Mean Square Error
PCNNs	Physics-Constrained Neural Networks		VMD	Variational Mode Decomposition
HEA	Hybrid Evolutionary Adaptive Approach		ARMA	Auto Regressive Moving Average
EPSO	Evolutionary Particle Swarm Optimization		MTLF	Medium-Term Load Forecasting
SROCC	Spearman Rank Order Correlation Coefficient		LTLF	Long-Term Load Forecasting
SDGNN	Sparse Dynamic Graph Neural Network		ARIMA	Auto Regressive Integrated Moving Average
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses		BiLSTM	Bidirectional Long Short-Term Memory

References

1. Chung, W.H.; Gu, Y.H.; Yoo, S.J. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. *Energy* **2022**, *246*, 123350. [CrossRef]
2. Liao, W.; Wang, S.; Yang, D.; Yang, Z.; Fang, J.; Rehtanz, C.; Porté-Agel, F. TimeGPT in load forecasting: A large time series model perspective. *Appl. Energy* **2025**, *379*, 124973. [CrossRef]
3. Pombo, D.V.; Bacher, P.; Ziras, C.; Bindner, H.W.; Spataru, S.V.; Sørensen, P.E. Benchmarking physics-informed machine learning-based short term PV-power forecasting tools. *Energy Rep.* **2022**, *8*, 6512–6520. [CrossRef]
4. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* **2018**, *11*, 1636. [CrossRef]
5. Ugbehe, P.O.; Diemuodeke, O.E.; Aikhuele, D.O. Electricity demand forecasting methodologies and applications: A review. *Sustain. Energy Res.* **2025**, *12*, 19. [CrossRef]

6. Khan, P.W.; Byun, Y.C.; Lee, S.J.; Kang, D.H.; Kang, J.Y.; Park, H.S. Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energies* **2020**, *13*, 4870. [CrossRef]

7. Kuster, C.; Rezgui, Y.; Mourshed, M. Electrical load forecasting models: A critical systematic review. *Sustain. Cities Soc.* **2017**, *35*, 257–270. [CrossRef]

8. Lo Brano, V.; Ciulla, G.; Di Falco, M. Artificial neural networks to predict the power output of a PV panel. *Int. J. Photoenergy* **2014**, *2014*, 193083. [CrossRef]

9. Joseph, L.P.; Deo, R.C.; Casillas-Pérez, D.; Prasad, R.; Raj, N.; Salcedo-Sanz, S. Short-term wind speed forecasting using an optimized three-phase convolutional neural network fused with bidirectional long short-term memory network model. *Appl. Energy* **2024**, *359*, 122624. [CrossRef]

10. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]

11. Xian, H.; Che, J. Multi-space collaboration framework based optimal model selection for power load forecasting. *Appl. Energy* **2022**, *314*, 118937. [CrossRef]

12. Ma, L.; Luan, S.; Jiang, C.; Liu, H.; Zhang, Y. A review on the forecasting of wind speed and generated power. *Renew. Sustain. Energy Rev.* **2009**, *13*, 915–920. [CrossRef]

13. Foley, A.M.; Leahy, P.G.; Marvuglia, A.; McKeogh, E.J. Current methods and advances in forecasting of wind power generation. *Renew. Energy* **2012**, *37*, 1–8. [CrossRef]

14. Yu, M.; Niu, D.; Zhao, J.; Li, M.; Sun, L.; Yu, X. Building cooling load forecasting of IES considering spatiotemporal coupling based on hybrid deep learning model. *Appl. Energy* **2023**, *349*, 121547. [CrossRef]

15. Liu, H.; Tian, H.Q.; Chen, C.; Li, Y.f. A hybrid statistical method to predict wind speed and wind power. *Renew. Energy* **2010**, *35*, 1857–1861. [CrossRef]

16. Wang, Q.; Li, Y.; Li, R. Integrating artificial intelligence in energy transition: A comprehensive review. *Energy Strategy Rev.* **2025**, *57*, 101600. [CrossRef]

17. Wanigasekara, C.; Swain, A.; Nguang, S.K.; Prusty, B.G. Neural network based inverse system identification from small data sets. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–6.

18. Wanigasekara, C.; Oromiehie, E.; Swain, A.; Prusty, B.G.; Nguang, S.K. Machine learning based predictive model for AFP-based unidirectional composite laminates. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2315–2324. [CrossRef]

19. Wanigasekara, C.; Oromiehie, E.; Swain, A.; Prusty, B.G.; Nguang, S.K. Machine learning-based inverse predictive model for AFP based thermoplastic composites. *J. Ind. Inf. Integr.* **2021**, *22*, 100197. [CrossRef]

20. Oromiehie, E.; Prusty, B.G.; Rajan, G.; Wanigasekara, C.; Swain, A. Machine learning based process monitoring and characterisation of automated composites. In Proceedings of the SAMPE, Seattle, WA, USA, 22–25 May 2017; pp. 1–6.

21. Wanigasekara, C.; Swain, A.; Nguang, S.K.; Prusty, B.G. Improved learning from small data sets through effective combination of machine learning tools with VSG techniques. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–6.

22. Witharama, W.; Bandara, K.; Azeez, M.; Bandara, K.; Logeeshan, V.; Wanigasekara, C. Advanced genetic algorithm for optimal microgrid scheduling considering solar and load forecasting, battery degradation, and demand response dynamics. *IEEE Access* **2024**, *12*, 83269–83284. [CrossRef]

23. McCullagh, P. What is a statistical model? *Ann. Stat.* **2002**, *30*, 1225–1310. [CrossRef]

24. Luna-Romero, S.F.; Serrano-Guerrero, X.; de Souza, M.A.; Escrivá-Escrivà, G. Enhancing anomaly detection in electrical consumption profiles through computational intelligence. *Energy Rep.* **2024**, *11*, 951–962. [CrossRef]

25. King, G.; Tomz, M.; Wittenberg, J. Making the most of statistical analyses: Improving interpretation and presentation. *Am. J. Political Sci.* **2000**, *44*, 347–361. [CrossRef]

26. Tasnim, S.; Rahman, A.; Shafiullah, G.; Oo, A.M.T.; Stojcevski, A. A time series ensemble method to predict wind power. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA, 9–12 December 2014; pp. 1–5.

27. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]

28. Spanos, A. Foundational issues in statistical modeling: Statistical model specification and validation. *Ration. Mark. Morals* **2011**, *2*, 146–178.

29. Nayak, A.K.; Sharma, K.C.; Bhakar, R.; Mathur, J. ARIMA based statistical approach to predict wind power ramps. In Proceedings of the 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA 26–30 July 2015; pp. 1–5.

30. Zaimi, M.; El Achouby, H.; Zegoudi, O.; Ibral, A.; Assaid, E. Numerical method and new analytical models for determining temporal changes of model-parameters to predict maximum power and efficiency of PV module operating outdoor under arbitrary conditions. *Energy Convers. Manag.* **2020**, *220*, 113071. [CrossRef]

31. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [CrossRef]

32. Wang, C.; Wang, Y.; Ding, Z.; Zheng, T.; Hu, J.; Zhang, K. A transformer-based method of multienergy load forecasting in integrated energy system. *IEEE Trans. Smart Grid* **2022**, *13*, 2703–2714. [CrossRef]

33. Grandón, T.G.; Schwenzer, J.; Steens, T.; Breuing, J. Electricity demand forecasting with hybrid classical statistical and machine learning algorithms: Case study of Ukraine. *Appl. Energy* **2024**, *355*, 122249. [CrossRef]

34. Sarker, I.H.; Hoque, M.M.; Uddin, M.K.; Alsanoosy, T. Mobile data science and intelligent apps: Concepts, AI-based modeling and research directions. *Mob. Netw. Appl.* **2021**, *26*, 285–303. [CrossRef]

35. Bergmann, D. What is Machine Learning? IBM—ibm.com. Available online: https://www.ibm.com/think/topics/machine-learning (accessed on 15 November 2026).

36. Park, K.; Kim, J.; Seo, J. Pint: Physics-informed neural time series models with applications to long-term inference on weatherbench 2m-temperature data. *arXiv* **2025**, arXiv:2502.04018.

37. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M.; et al. Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manag.* **2023**, *71*, 102642. [CrossRef]

38. Dokur, E.; Erdogan, N.; Sengor, I.; Yuzgec, U.; Hayes, B.P. Near real-time machine learning framework in distribution networks with low-carbon technologies using smart meter data. *Appl. Energy* **2025**, *384*, 125433. [CrossRef]

39. Faizan, M.; Afgan, I. Dynamic Assessment and Optimization of Thermal Energy Storage Integration with Nuclear Power Plants Using Machine Learning and Computational Fluid Dynamics. *Appl. Energy* **2025**, *391*, 125939. [CrossRef]

40. INT Global. 5 Major Limitations of Machine Learning. 2025. Available online: https://intglobal.com/blogs/5-major-limitations-of-machine-learning/ (accessed on 15 November 2026).

41. Azzeddine, H.A.; Tioursi, M.; Chaouch, D.E.; Khiari, B. An offline trained artificial neural network to predict a photovoltaic panel maximum power point. *Rev. Roum. Sci. Techn. Énerg* **2016**, *61*, 255–257.

42. 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* **2017**, *10*, 841–851. [CrossRef]

43. Kim, T.Y.; Cho, S.B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [CrossRef]

44. Du, L.; Zhang, L.; Wang, X. Spatiotemporal feature learning based hour-ahead load forecasting for energy internet. *Electronics* **2020**, *9*, 196. [CrossRef]

45. Wu, K.; Wu, J.; Feng, L.; Yang, B.; Liang, R.; Yang, S.; Zhao, R. An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e12637. [CrossRef]

46. Li, K.; Mu, Y.; Yang, F.; Wang, H.; Yan, Y.; Zhang, C. A novel short-term multi-energy load forecasting method for integrated energy system based on feature separation-fusion technology and improved CNN. *Appl. Energy* **2023**, *351*, 121823. [CrossRef]

47. He, J.; Shi, L.; Tian, H.; Wang, X.; Sun, X.; Zhang, M.; Yao, Y.; Shu, G. Applying artificial neural network to approximate and predict the transient dynamic behavior of CO_2 combined cooling and power cycle. *Energy* **2023**, *285*, 129451. [CrossRef]

48. Islam, M.S.; Hasan, A.J.; Rahman, M.S.; Yusuf, J.; Sajol, M.S.I.; Tumpa, F.A. Location agnostic source-free domain adaptive learning to predict solar power generation. In Proceedings of the 2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG), Wollongong, Australia, 3–6 December 2023; pp. 1–6.

49. Scott, C.; Ahsan, M.; Albarbar, A. Machine learning for forecasting a photovoltaic (PV) generation system. *Energy* **2023**, *278*, 127807. [CrossRef]

50. Visser, L.; AlSkaif, T.; Khurram, A.; Kleissl, J.; van Sark, W. Probabilistic solar power forecasting: An economic and technical evaluation of an optimal market bidding strategy. *Appl. Energy* **2024**, *370*, 123573. [CrossRef]

51. Schierholz, C.M.; Scharff, K.; Schuster, C. Evaluation of neural networks to predict target impedance violations of power delivery networks. In Proceedings of the 2019 IEEE 28th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), Montreal, QC, Canada, 6–9 October 2019; pp. 1–3.

52. Pariaman, H.; Luciana, G.; Wisyaldin, M.; Hisjam, M. Anomaly detection using lstm-autoencoder to predict coal pulverizer condition on coal-fired power plant. *J. Nov. Carbon Resour. Sci. Green Asia Strategy* **2021**, *1*, 89–97. [CrossRef]

53. Huang, Y.; Zhao, Y.; Wang, Z.; Liu, X.; Fu, Y. Sparse dynamic graph learning for district heat load forecasting. *Appl. Energy* **2024**, *371*, 123685. [CrossRef]

54. Wang, X.; Wang, H.; Li, S.; Jin, H. A reinforcement learning-based online learning strategy for real-time short-term load forecasting. *Energy* **2024**, *305*, 132344. [CrossRef]

55. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.

56. Jinil, N.; Reka, S. Deep learning method to predict electric vehicle power requirements and optimizing power distribution. In Proceedings of the 2019 Fifth International Conference on Electrical Energy Systems (ICEES), Chennai, India, 21–22 February 2019; pp. 1–5.

57. Gao, Z.; Yu, J.; Zhao, A.; Hu, Q.; Yang, S. A hybrid method of cooling load forecasting for large commercial building based on extreme learning machine. *Energy* **2022**, *238*, 122073. [CrossRef]

58. Niu, D.; Yu, M.; Sun, L.; Gao, T.; Wang, K. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Appl. Energy* **2022**, *313*, 118801. [CrossRef]

59. Guo, Y.; Li, Y.; Qiao, X.; Zhang, Z.; Zhou, W.; Mei, Y.; Lin, J.; Zhou, Y.; Nakanishi, Y. BiLSTM multitask learning-based combined load forecasting considering the loads coupling relationship for multienergy system. *IEEE Trans. Smart Grid* **2022**, *13*, 3481–3492. [CrossRef]

60. Cui, W.; Yang, W.; Zhang, B. A frequency domain approach to predict power system transients. *IEEE Trans. Power Syst.* **2023**, *39*, 465–477. [CrossRef]

61. Huang, Y.; Zhao, Y.; Wang, Z.; Liu, X.; Liu, H.; Fu, Y. Explainable district heat load forecasting with active deep learning. *Appl. Energy* **2023**, *350*, 121753. [CrossRef]

62. Osório, G.J.; Matias, J.C.; Catalão, J.P. Hybrid evolutionary-adaptive approach to predict electricity prices and wind power in the short-term. In Proceedings of the 2014 Power Systems Computation Conference, Wroclaw, Poland, 18–22 August 2014; pp. 1–7.

63. Li, K.; Mu, Y.; Yang, F.; Wang, H.; Yan, Y.; Zhang, C. Joint forecasting of source-load-price for integrated energy system based on multi-task learning and hybrid attention mechanism. *Appl. Energy* **2024**, *360*, 122821. [CrossRef]

64. Ghimire, S.; Deo, R.C.; Casillas-Pérez, D.; Salcedo-Sanz, S. Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting. *Appl. Energy* **2024**, *353*, 122059. [CrossRef]

65. Park, J.; Alvarenga, E.; Jeon, J.; Li, R.; Petropoulos, F.; Kim, H.; Ahn, K. Probabilistic forecast-based portfolio optimization of electricity demand at low aggregation levels. *Appl. Energy* **2024**, *353*, 122109. [CrossRef]

66. Amini, A.; Rey-Mermet, S.; Crettenand, S.; Münch-Alligné, C. A hybrid methodology for assessing hydropower plants under flexible operations: Leveraging experimental data and machine learning techniques. *Appl. Energy* **2025**, *383*, 125402. [CrossRef]

67. Deng, W.; Le, H.; Nguyen, K.T.; Gogu, C.; Medjaher, K.; Morio, J.; Wu, D. A Generic physics-informed machine learning framework for battery remaining useful life prediction using small early-stage lifecycle data. *Appl. Energy* **2025**, *384*, 125314. [CrossRef]

68. Song, H.; Zhang, B.; Jalili, M.; Yu, X. Multi-swarm multi-tasking ensemble learning for multi-energy demand prediction. *Appl. Energy* **2025**, *377*, 124553. [CrossRef]

69. Motevakel, P.; Roldán-Blay, C.; Roldán-Porta, C.; Escrivá-Escrivá, G.; Dasí-Crespo, D. Hybrid energy solutions for enhancing rural power reliability in the Spanish municipality of Aras de los Olmos. *Appl. Sci.* **2025**, *15*, 3790. [CrossRef]

70. Diaz-Iglesias, A.; Belaunzaran, X.; Florez-Tapia, A.M. Short-Term Power Demand Forecasting for Diverse Consumer Types to Enhance Grid Planning and Synchronisation. *arXiv* **2025**, arXiv:2506.04294. [CrossRef]

71. Pinheiro, M.G.; Madeira, S.C.; Francisco, A.P. Short-term electricity load forecasting—A systematic approach from system level to secondary substations. *Appl. Energy* **2023**, *332*, 120493. [CrossRef]

72. Hasan, M.; Mifta, Z.; Papiya, S.J.; Roy, P.; Dey, P.; Salsabil, N.A.; Chowdhury, N.U.R.; Farrok, O. A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications. *Energy Convers. Manag.* **2025**, *26*, 100922. [CrossRef]

73. Dong, J.; Olama, M.M.; Kuruganti, T.; Melin, A.M.; Djouadi, S.M.; Zhang, Y.; Xue, Y. Novel stochastic methods to predict short-term solar radiation and photovoltaic power. *Renew. Energy* **2020**, *145*, 333–346. [CrossRef]

74. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]

75. Benato, A.; Bracco, S.; Stoppato, A.; Mirandola, A. LTE: A procedure to predict power plants dynamic behaviour and components lifetime reduction during transient operation. *Appl. Energy* **2016**, *162*, 880–891. [CrossRef]

76. Suganthi, L.; Samuel, A.A. Energy models for demand forecasting—A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1223–1240. [CrossRef]

77. Xie, X.; Parlikad, A.K.; Puri, R.S. A neural ordinary differential equations based approach for demand forecasting within power grid digital twins. In Proceedings of the 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Beijing, China, 21–23 October 2019; pp. 1–6.

78. Cuomo, S.; Di Cola, V.S.; Giampaolo, F.; Rozza, G.; Raissi, M.; Piccialli, F. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *J. Sci. Comput.* **2022**, *92*, 88. [CrossRef]

79. Beinert, D.; Holzhüter, C.; Thomas, J.M.; Vogt, S. Power flow forecasts at transmission grid nodes using graph neural networks. *Energy AI* **2023**, *14*, 100262. [CrossRef]

80. Azeem, A.; Ismail, I.; Jameel, S.M.; Romlie, F.; Danyaro, K.U.; Shukla, S. Deterioration of electrical load forecasting models in a smart grid environment. *Sensors* **2022**, *22*, 4363. [CrossRef]

81. Jagait, R.K.; Fekri, M.N.; Grolinger, K.; Mir, S. Load forecasting under concept drift: Online ensemble learning with recurrent neural network and ARIMA. *IEEE Access* **2021**, *9*, 98992–99008. [CrossRef]

82. Díaz-Bedoya, D.; González-Rodríguez, M.; Clairand, J.M.; Serrano-Guerrero, X.; Escrivá-Escrivá, G. Forecasting Univariate Solar Irradiance using Machine learning models: A case study of two Andean Cities. *Energy Convers. Manag.* **2023**, *296*, 117618. [CrossRef]