Frontiers in Computer Science and Artificial Intelligence
ISSN: 2978-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Hybrid Deep Learning and Machine Learning Approaches for Industrial Power Load Forecasting: A Review

Remon Das[1]✉, Md Abdul Ahad Juel[2], Md. Raisul Islam[3]
[1]MS in Engineering Technology, Western Carolina University, North Carolina, USA
[2]University at Buffalo, USA, abdulahadjuel@gmail.com
[3]Western Illinois University, USA, raisulislamjony@gmail.com
Corresponding Author: Remon Das **E-mail**: rdas1@catamount.wcu.edu

| ABSTRACT

Industrial Power Load in the United States is rising gradually due to rapid manufacturing facilities enhancement in the recent year. Due to this expansion, accurate forecasting of industrial electricity load plays a vital role for power generation, transmission, and distribution planning for a specified manufacturing zone. The statistical model of forecasting faces a significant challenge due to the nonlinearity and multi scale nature of industrial load data. But recent application of Hybrid Deep Learning and Machine Learning models has demonstrated superior performance in industrial power load forecasting over statistical forecasting model. This paper presents a structured and comparative review of the recent Hybrid Deep Learning and Machine Learning based Data driven approaches for Industrial power load forecasting. The findings of our study highlight the effectiveness of the hybrid approach for reliable and high precision industrial power load forecasting which enhances the intelligent energy management in modern industrial systems.

| KEYWORDS

Industrial power load forecasting, time-series forecasting, predictive analysis, machine learning, deep learning.

| ARTICLE INFORMATION

ACCEPTED: 01 January 2026 **PUBLISHED:** 11 January 2026 **DOI:** 10.32996/jcsts.2026.5.1.7x

1. Introduction

The effective and reliable power generation, transmission and distribution planning depends on the ability to accurately forecasting electricity demand. Forecasting is the foundational task for operational planning and strategic financial decision making [1]. Forecasting in the industrial sector is challenging and a high impact problem. In 2021, the industrial sector accounted for a substantial 33% of total U.S. energy consumption, making it a pivotal segment for grid management [2].

Historically, U.S. electricity consumption remained relatively flat for nearly two decades. Because of a major shift from manufacturing to service sectors, demand increases. This manufacturing-to-service trend has been shifted significantly since 2020. The U.S. Energy Information Administration (EIA) forecasts that total electricity consumption will grow at an average rate of 1.7% per year through 2026. Compare to the total power laod, the commercial and industrial sectors are growing even faster, at an average of 2.6% and 2.1% per year, repectively [3]. Industrial power demand is projected to climb by 43% from 2024 to 2050, fueled by automation and robotics in industrial production and manufacturing processes [4]. The dynamic power consumption pattern requires the development of more sophisticated and robust forecasting tools.

Electrical power load time series forecasting mostly depends on statistical methods such as ARIMA, Holt-Winters exponential smoothing, and regression-based models [5]. The regressive model provides a degree of interpretability but they often struggle to capture the complex, non linear relationship and long term temporal dependencies inherent in real-world energy and power data [5]. Various forecasting approches, fuzzy neural networks, gray algorithms, gray markov models and support vector regression algorithms [6-9], have been utilized for electricity generation forecasting. However, these approaches demonstrate limited efficiency when processing large-scale or highly complex datasets [10-13]. In photovoltaic (PV) forecasting, Artificial Neural Netowrks (ANNs) and Support Vector Machines (SVMs) [14-15] are among the most widely utilized machine learning models. Numerous studies indicate that, neural network based model outperform traditional time series models on nonlinear datasets, nontheless, they encounter difficulties in managing large datasets and intricate data mining tasks [15].

For that reason, the researchers have increasingly adopted deep learing-based approaches, particularly Recurrent Neural Networks (RNNs) and their advanced variants such as Long Short-Term Memory (LSTM) networks [16]. LSTM models have demonstrated strong potential in energy forecasting because LSTMs are able to handle the long term dependencies within sequential data [17]. Traditional LSTM architectures process information in a single (forward) direction, which may limit their ability to fully utilize temporal correlations between past and future data points. On the other hand, Bidirectional LSTM (Bi-LSTM) networks overcome this limitation by processing sequences in both forward and backward directions, enabling a more comprehensive representation of temporal dynamics [18]. Despite the effectiveness, Bi-LSTMs require large datasets and intensive computation.

Transformer-based models have also become prominent in time series forecasting, including for electricity load prediction, due to their ability to capture long-range temporal dependencies and handle multivariate inputs. A framework integrates Fourier Transform-based transformers with an enhanced optimization algorithm to achieve high-resolution energy forecasting, addressing the challenges of sequence redundancy and computational efficiency in energy systems [19]. Informer was developed precisely for Long Sequence Time-Series Forecasting (LSTF), and in experiments demonstrated strong performance on electricity consumption planning tasks [20]. The Autoformer model similarly enhances forecasting accuracy by introducing an auto-correlation mechanism, which helps identify periodic patterns and reduces the errors associated with long lookback windows [21]. *SmartFormer* (a graph-nested transformer) explicitly models inter-series dependencies for substations and shows improved performance over standard transformer baselines in multivariate load forecasting settings [22]. A relevant applied work is Power substation load forecasting using interpretable transformer-based temporal fusion neural networks, where the Temporal Fusion Transformer (TFT) is used to forecast 24-h and 48-h ahead loads, providing both accuracy and interpretable insights into which covariates (e.g. weather, temporal features) are most influential [23]. There is also work on hybrid approaches, fusing traditional statistical or machine learning models (e.g. LightGBM or multivariate linear regression) with transformer-based architectures to improve load forecasting accuracy, as seen in an "Improved Autoformer" approach [24]. Transformer-based models, while powerful, present several well-documented limitations. They typically require large volumes of historical and covariate data to effectively capture both periodic and irregular temporal patterns, which can be challenging in industrial contexts with limited or proprietary data sources [25].

With the rapid evolution of these Deep Learning (DL) and Machine Learning (ML) based techniques, it is necessary to compare techniques and their effectiveness. This review studies a range of hybrid deep learning and machine learning models proposed for industrial power load forecasting using data from multiple industrial load data, highlighting their methodologies, performance and real-world applicability.

2. Common DL and ML Methods Used in Power and Energy Forecasting

Power and energy load forecasting has been extensively studied using a wide range of machine learning (ML) and deep learning (DL) techniques to capture the complex, nonliner and time varying characteristics of electrical demand. Traditional ML methods have been widely adopted due to their somplicity, interpretability and effectiveness in short and medium term forecasting while DL models have gained incresing attention for their ability to automatically learn hierarchical temporal and spatial features from large scale time series data. This section will provide an overview of the most commonly used ML and DL approaches applied to power and energy load forecasting.

- Auto Regressive Integrated Moving Average (ARIMA) accounts for a changing mean in the dataset. This accounts for linear trends over time.
- SARIMA adds seasonality to the ARIMA technique. [26]
- Artificial Neural Networks (ANNs) are a class of Machine Learning models that are inspired by biological neural networks.[27]
- Autoencoders (AE) are a type of ANN that compress or encode data on its input to their essential features.[28]
- Convolutional Neural Networks (CNNs) use convolutional kernels applied to data. These are commonly used in 2D images as shown in figure 2 and 3 but can be applied to 1D data to find patterns in the data and for forecasting.[29]

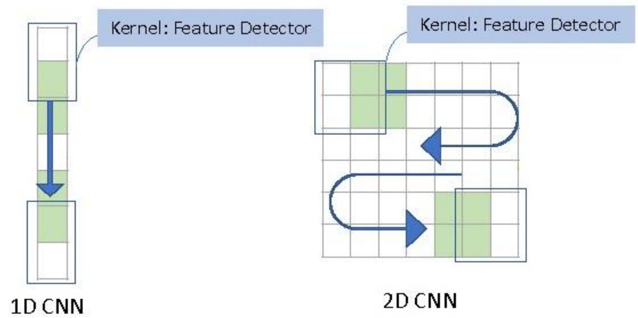

Figure 1: Feature detection of 1D CNN and 2D CNN. The shape of a convolutional filter is a vector form in 1D CNN and a Matrix in 2D CNN

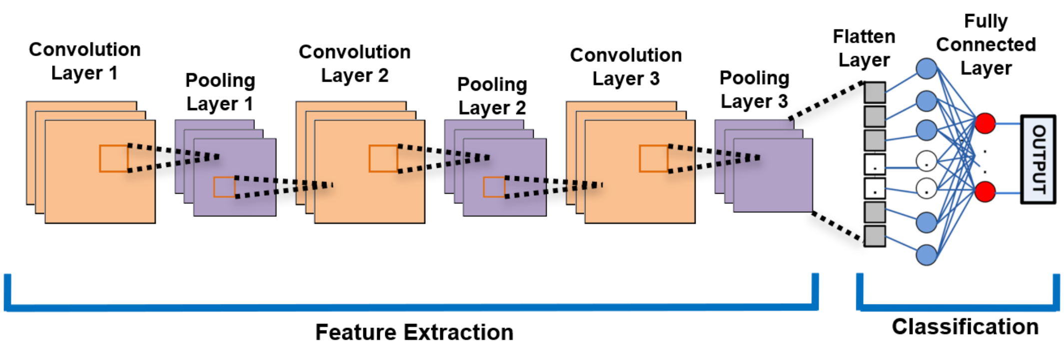

Figure 2: Typical building blocks of a CNN Model

- Recurrent Neural Networks (RNNs), shown in Figure 3, are neural networks that can process sequential data by using past information as inputs for subsequent outputs.[30]

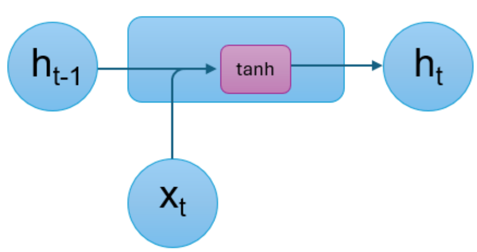

Figure 3. Typical model of a single cell of a generic RNN.

- Long Short-Term Memory (LSTM) models, shown in Figure 4, are a type of RNN that allows for both long-term and short-term historical values to affect the output.[31]

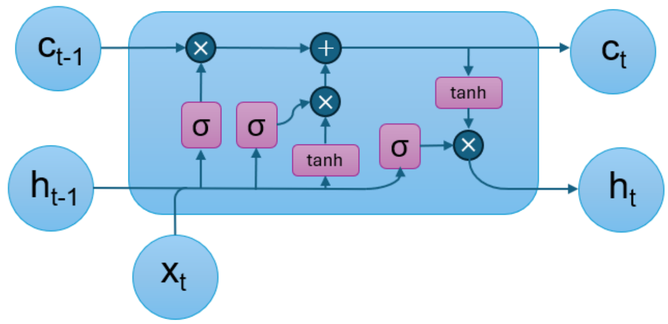

Figure 4: A single cell of a typical LSTM.

- Bidirectional LSTM (Bi-LSTM) as shown in figure 5 is a type of RNN that processes data both in forward directions as well as backward in the series to build stronger relationships between historical data and the next output. [32]

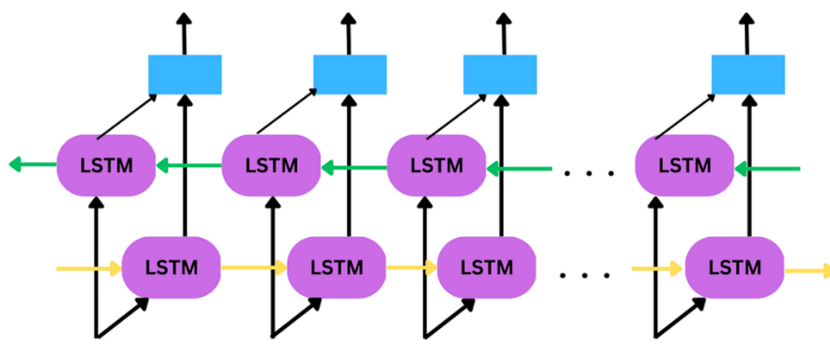

Figure 5: A building blocks of a typical BiLSTM.

- Gated Recurrent Units (GRU), shown in Figure 6, are like LSTM models but simpler. They use two "gates" to control the flow of data, named "update" and "reset."[33]

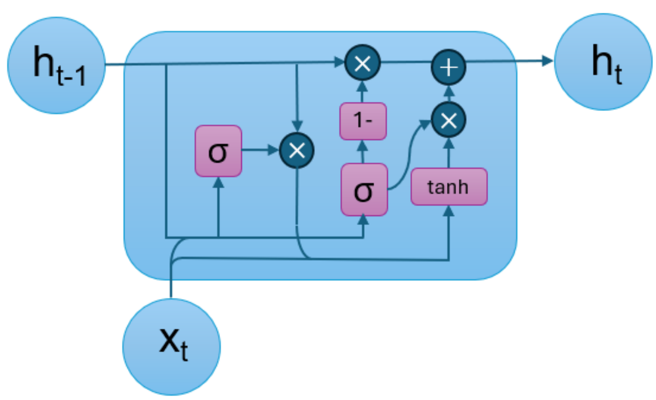

Figure 6: A single cell of a typical GRU.

- Attention-based models utilize a method to determine the importance of different components in a sequence and allow models to focus more on those parts.[34]
- Hybrid Models: While there are many parameters in each of these models that can be adjusted, often they are combined in hybrid configurations. For example, a study might compare LSTM with the hybrid model CNN-LSTM where the output of one model becomes the input for the next. This leverages the advantages of each model often creating a more accurate result.

3. Common Performance Evaluation Matrix for Power and Energy Forecasting Model

3.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is computed by taking the sum of the absolute difference between each predicted value and its true value and then dividing by the total number of data samples [35]. The MAE depicts as:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - f_i|$$

where y_i is the actual and f_i is the forecasted value for the power load and N is the number of data samples.

3.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is defined as the square root of the average squared difference of actual value and prediction value, in other words, the square root of Mean Squared Error (MSE) [35]. The RMSE is depicted as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - f_i)^2}$$

where y_i is the actual and f_i is the forecasted value for the power load and N is the number of data samples.

3.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) calculates the average of the absolute percentage errors between actual and predicted values. [36] The MAPE is defined as:

$$MAPE = \frac{100}{N} \sum_{i=1}^{N} |\frac{y_i - f_i}{y_i}|$$

where y_i is the actual value and f_i is the forecasted value for the power load, and N is the number of data samples

3. Review on Industrial Load Forecasting Models

Over the past decades, various statistical methods have been developed for power and energy forecasting and for the planning of energy systems. But from those methods, very few are industrial power load forecasting [5-15]. Due to increase in complexity and variability in load patterns has led, researchers to gradually shift focus towards more advanced machine learning techniques like hybrid deep learning model [16-18]. Recently, different variants of transformer models were explored for power load forecasting and energy forecasting [19-25] [37]. This review studies a range of different models developed for industrial power load forecasting using data from multiple sources, highlighting their methodologies, performance, and real-world applicability.

Quanbo Ge et al. developed a hybrid intelligent forecasting model combining reinforcement learning, particle swarm optimization (PSO), and least squares support vector machines (LSSVM) for short-term industrial power load forecasting of a single industrial power consumer [38]. The author proposed the name of this hybrid model is Q-PSO-LSSVM. The model first applied an improved K-means clustering algorithm to classify historical load data into distinct production patterns and eliminate outliers using the local outliers factor (LOF) method. Then, a Q-learning enhanced PSO dynamically adjusts particle inertia weights to optimize

LSSVM parameters, preventing local minima and improving generalization. The model uses real industrial load data recoded every 15 minutes from various industries and regions in China. Experimental results demonstrate that the model Q-PSO-LSSVM has achieved significantly lower mean absolute percentage error (MAPE 3-5%) compared to traditional PSO-LSSVM and Prophet models, effectively capturing regional and industrial variations in load behavior and providing high-accuracy short-term forecasts suitable for industrial energy management applications.

Yuanyuan Wang et al. proposed a hybrid model named TCN-LightGBM [39] combining a temporal convolutional network (TCN) and light gradient boosting machine (LightGBM) for accurate short-term industrial power load forecasting. The model used a fixed length sliding time window to reconstruct electrical features (load, current, power etc.) along with meteorological and calendar date, enabling the TCN to extract long term temporal dependencies and hidden patterns. These extracted features passed through LightGBM, which efficiently performs the final load prediction using its gradient boosting and leaf-wise growth mechanisms. The study employs real smart meter data from industrial customers across China, Australia, and Ireland, representing sectors like medical, plastic, and coal mining industries. Experimental comparisons against models such as LSTM, Bi-GRU, CNN, PCA-LightGBM demonstrate that the proposed TCN-LightGBM achieves the lowest MAPE around 2-4%.

Felipe Leite Coelho da Silva et al. have developed a neural network auto regression (NNAR) and multilayer perceptron (MLP) models for forecasting monthly industrial electricity consumption in Brazil. A comparison study has been done to compare the developed model with traditional statistical models, Holt-winters, SARIMA, and Dynamic Linear Model (DLM). The author used historical data from 1979 to 2020, obtained from the central bank of Brazil, divided intro training (1979-2018) and testing (2019-2020) datasets. The performance was evaluated using mean absolute percentage error (MAPE) to assess short-term predictive accuracy. Results showed that while all models effectively captured seasonal and crisis related variations in industrial electricity use, the MLP model outperformed than the other models, achieving the lowest MAPE is 3.4% [40]

Sravani Parvathareddy et al. proposed a novel Fourier transform (FT) transformer and genetic algorithm (GA) hybrid model for accurate energy forecasting and load optimization has been developed in industrial and commercial energy systems [41]. The FT-transformer leverages self-attention mechanisms and Fourier based seasonality encoding to capture long term dependencies and temporal patterns in large-scale energy demand data, while the enhanced covariance matrix adaptation evolution strategy (CMA-ES) GA module optimizes load scheduling by minimizing costs, emissions, and peak demand fluctuations. The model was trained and validated using real industrial energy consumption data from facilities such as the Jwaneng Mine in Botswana, encompassing multiple years of monthly energy records. Experimental results demonstrate that the FT-Transformer achieved a mean absolute error (MAE) of 3.03×10^5 kWh and Root Mean Square Error (RMSE) of 3.31×10^5 kWh, outperforming RNN, PSO, and tree-based models by up to 48% in accuracy and reducing computational time by 38%.

Haoyu Jiang et al. established a hybrid forecasting model [42] that integrates a data completion network (DCN) with an enhanced temporal convolutional network (E-TCN) for ultra-short-term industrial load prediction. The model first employs the DCN to handle missing and incomplete industrial load data, ensuring data continuity and quality, and then uses the E-TCN to capture complex temporal dependencies and nonlinear variations in high frequency load patterns. The study utilizes real-time load data collected at 15 second intervals from multiple industrial enterprises in China, including sectors with irregular load fluctuations due to production dynamics. Results show that the DCN-E-TCN model significantly outperforms conventional TCN, Gru, and LSTM models, achieving up to 35% lower RMSE and 30% higher R^2, leading to highly stable and precise forecasts suitable for industrial demand response, production scheduling, and grid stability applications.

Mao Tan et al. developed a hybrid ensemble forecasting model [43] that combines long short-term memory (LSTM) networks with multiple machine learning algorithms, support vector regression (SVR), extreme gradient boosting (XGBoost), and Random Forest (RF) to enhance the accuracy and robustness of industrial power demand forecasting. The model used Bayesian optimization for parameter tuning and a stacking strategy to fuse the predictions from base learners, enabling effective capture of both liner and nonlinear temporal dependencies. The model is trained and tested on real industrial power load data collected a minute-level intervals from a steel manufacturing plant in China, characterized by high volatility and abrupt changes due to production dynamics. Experimental results showed that the proposed LSTM based hybrid ensemble achieved superior performance with a mean absolute percentage error (MAPE) below 1.5%, outperforming standalone deep learning and traditional models such as GRU, CNN and ARIMA.

Nam Nguyen Vu Nhat et al. a case study of Vietnam, introduced a hybrid forecasting model that combines ensemble empirical mode decomposition (EEMD) with a long-short term memory (LSTM) [44] network to improve the accuracy of short term industrial load prediction. The model first decomposes the complex and non-stationary load data into a set of simpler intrinsic mode functions (IMFs) using EEMD. These components are then individually predicted by the LSTM network before being aggregated into a final forecasting, a method that allows for more effective capture of intricate and nonlinear patterns across different time scales. The model is validated using hourly power load data from a Vietnamese industrial manufacturing plant, collected over approximately 18 months, and characterized by its nonlinear and highly fluctuating nature. Experimental results demonstrate that the proposed EEMD-LSTM model achieves superior performance with a mean absolute percentage error (MAPE) of approximately 1.3% for 1-step forecasting, significantly outperforming standalone models like linear regression (LR), artificial neural networks (ANN) and a standard LSTM.

Chaodong Fan et al. developed a novel deep learning model that leveraged a hybrid ensemble strategy and an error correction mechanism to accurately forecast industrial power demand. The model first employs an ensemble of gated recurrent unit (GRU) networks, where base learners are generated by perturbing GRU parameters using a novel Multi-Objective Molecular Dynamics Theory Optimization Algorithm (MMDTOA) and then integrated via kernel ridge regression stacking. Subsequently, a combined error correction strategy, utilizing both a dynamic Gaussian function and a separate GRU model, is applied to the residuals to further refine the prediction accuracy. The model is trained and evaluated on real 15-minute interval load data from a steel plant in South Korea, which includes multiple features and exhibits strong fluctuations. Experimental results demonstrate that the proposed model achieves superior performance with a Normalized Mean Absolute Error (NMAE) as low as 3.684%, outperforming eight other models including CNN and LSTM, thereby highlighting its effectiveness and robustness for complex industrial forecasting tasks [45].

Ziwei Zhu et al. a hybrid machine learning model designed for day-ahead industrial load forecasting utilizes an extreme learning machine (ELM) as its base predictor, with its initial weights and biases optimized by a firefly algorithm (FA) to enhance accuracy. This optimized FA-ELM is then integrated into an ensemble framework using the AdaBoost algorithm, which combines multiple weak learners to create a single strong predictor and correct for errors. A key innovation is the introduction of a "load change rate" (LCR) feature, which is calculated from historical load data to better capture nonlinear patterns, and feature selection is performed using the Spearman correlation coefficient to reduce input dimensionality. The model is applied to hourly electricity demand data from a furniture factory in China, including date, meteorological, and operational features. The results show that the proposed LCR-AdaBoost-FA-ELM model achieves a Mean Absolute Percentage Error (MAPE) of 3.01%, significantly outperforming standalone ELM and SVR models [46].

Thilo Walser et al. a Typical Load Profile (TLP)-supported convolutional neural network (CNN) framework [47] designed to improve short-term industrial load forecasting accuracy by incorporating typical daily and weekly load patterns into the learning process. The model first applies K-means clustering to historical industrial load data to generate representative TLPs, which are then fused with real-time load and meteorological features as multi-channel CNN inputs. This structure enables the CNN to extract both spatial and temporal correlations while maintaining sensitivity to periodic industrial operations. The study utilizes high-resolution industrial load data from multiple manufacturing enterprises in China, collected at 30-minute intervals over several months. Results demonstrate that the proposed TLP-CNN model outperforms baseline models such as LSTM, GRU, and SVR, achieving a mean absolute percentage error (MAPE) of 2.17%, thereby enhancing forecasting stability and adaptability across varying industrial processes.

4. Comparative Result Analysis of Industrial Power Forecasting Models

Direct comparison across these studies is challenging due to the variations in their input variables and evaluation criteria, which are compounded by inconsistent performance metrics. Table 1 outlines each study’s input variables, models tested, and the best-performing model of each study using the metrics of that study. From this table, the recent studies on industrial load forecasting [38-47] consistently demonstrated a shift toward hybrid machine learning and deep learning models that integrate advanced feature extraction, optimization, and ensemble learning. Hybrid approaches such as Q-PSO-LSSVM [38], TCN-LightGBM [39], FT-Transformer with GA [41], and DCN-E-TCN [42] are designed to capture nonlinear, non-stationary and multi-scale temporal characteristics of industrial load data. Ensemble and decomposition-based model, including LSTM-SVR-XGBoost-RF [43], EEMD-LSTM [44], MMDTOA-GRU ensemble [45], and LCR-AdaBoost-FA-ELM [46], further enhance forecasting robustness and accuracy. Overall, these methods report consistently low forecasting errors across diverse industrial application, confirming the effectiveness of hybrid DL-ML frameworks for short-term and ultra-short-term industrial power load forecasting [38-47].

TABLE 1: COMPARISONS OF INDUSTRIAL LOAD FORECASTING MODELS SURVEYED

Model / Reference	Key Feature & Description	Input Data Type	Application Domain	Reported Performance
Q-PSO-LSSVM [38]	Hybrid intelligent model combining Reinforcement Learning, PSO, and LSSVM; utilizes improved K-means clustering for data classification, LOF for outlier removal, and Q-learning-based PSO for dynamic parameter tuning.	15-min interval industrial load data (multiple industries, China)	Short-term industrial power load forecasting	Mean Absolute Percentage Error (MAPE) ≈ 3-5%
TCN-LightGBM [39]	Hybrid model integrating Temporal Convolutional Network (TCN) for temporal feature extraction and LightGBM for final load prediction; uses meteorological and calendar data for enhanced learning.	Smart meter data with electrical, meteorological, and calendar features	Industrial load forecasting across multiple regions	Mean Absolute Percentage Error (MAPE) around 2-4%
MLP vs. Statistical Models [40]	Comparative study between statistical (SARIMA, DLM, TBATS) and neural models (MLP, NNAR) for capturing seasonal and crisis-related variations in industrial electricity consumption.	Monthly industrial electricity data (Brazil, 1979–2020)	Industrial electricity demand forecasting	Mean Absolute Percentage Error (MAPE) ≈ 3.4%
FT-Transformer + GA [41]	Deep hybrid architecture using Fourier-based seasonality encoding with Transformer attention for temporal dependencies; Genetic Algorithm (GA) for load scheduling and optimization.	Monthly industrial energy data (Botswana, Jwaneng Mine)	Industrial and commercial energy forecasting and optimization	Mean Absolute Error (MAE) of 3.03×10^5 kWh and Root Mean Square Error (RMSE) of 3.31×10^5 kWh
DCN–E-TCN [42]	Hybrid framework combining Data Completion Network (DCN) for imputing missing values and Enhanced Temporal Convolutional Network (E-TCN) for capturing nonlinear temporal dynamics.	15-second industrial load data (China)	Ultra-short-term industrial load prediction	35% lower RMSE and 30% higher R^2 and the maximum error is 9.4%
LSTM + SVR + XGBoost + RF [43]	Ensemble hybrid model with LSTM feature extraction; predictions fused via stacking using SVR, XGBoost, and Random Forest; Bayesian optimization for parameter tuning.	Minute-level industrial load data (steel plant, China)	Industrial power demand forecasting	Mean Absolute Percentage Error (MAPE) below 1.5%
EEMD-LSTM [44]	Hybrid signal decomposition model integrating Ensemble Empirical Mode Decomposition (EEMD) with LSTM to handle non-stationary and nonlinear industrial load data.	Hourly industrial load data (Vietnam)	Short-term industrial load forecasting	Mean Absolute Percentage Error (MAPE) of approximately 1.3% for 1-step forecasting
MMDTOA-GRU Ensemble [45]	Optimized GRU-based ensemble model enhanced with Multi-Objective Molecular Dynamics Theory Optimization Algorithm (MMDTOA); includes kernel ridge regression stacking and dynamic error correction.	15-min interval industrial load data (steel plant, South Korea)	Industrial power demand forecasting	Normalized Mean Absolute Error (NMAE) as low as 3.684%
LCR-AdaBoost-FA-ELM [46]	Hybrid machine learning model using Firefly Algorithm (FA) to optimize Extreme Learning Machine (ELM); integrated with AdaBoost ensemble and load change rate (LCR) feature extraction.	Hourly industrial load data (furniture factory, China)	Day-ahead industrial load forecasting	Mean Absolute Percentage Error (MAPE) of 3.01%
TLP-CNN [47]	CNN-based model incorporating Typical Load Profiles (TLP) generated via K-means clustering; multi-channel input captures spatial-temporal industrial load correlations.	30-min interval industrial load data (China)	Short-term industrial load forecasting	Mean Absolute Percentage Error (MAPE) of 2.17%

5. Conclusion

The United States Energy is undergoing an unprecedented transformation because of rapidly increasing power demand in manufacturing sector and the integration of renewable energy resources. In this environment, accurate and reliable prediction of energy needs is crucial for maintaining the generation planning, grid stability and optimizing resources. This paper has provided a structured and comparative review of recent Hybrid AI and ML approaches to forecast the industrial power load and highlighted their benefits over traditional statistical methods, which struggle with the non-liner and highly variable nature of modern power systems. This review also highlights the benefits of utilizing hybrid architectures such as a combination of CNN, LSTM, Transformer, and others to attain the closest predictions.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD https://orcid.org/0009-0009-4052-9279
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] U.S. Energy Information Administration (EIA), "Electricity Explained: Electricity in the United States," [Online] Available: https://www.eia.gov/energyexplained/electricity/

[2] Energy in the United States - Wikipedia, accessed September 22, 2025, Available: https://en.wikipedia.org/wiki/Energy_in_the_United_States

[3] U.S. Energy Information Administration (EIA), "After more than a decade of little change, U.S. electricity consumption is rising again" Available: https://www.eia.gov/todayinenergy/detail.php?id=65264

[4] Which Sources Will Dominate U.S. Electricity Use? (2024-2050) - Visual Capitalist, accessed September 22, 2025, Available: https://www.visualcapitalist.com/sp/tema-01-u-s-electricity-demand-by-source-2024-2050/

[5] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.

[6] P.C. Chang, C.Y. Fan, J.C. Hsieh, "A weighted evolving fuzzy neural network for electricity demand forecasting", 2009 First Asian Conference on Intelligent Information and Database System, vol.9, 2009, pp.330–350.

[7] Y. Xie, M. Li, "Application of gray forecasting model optimized by genetic algorithm in electricity demand forecasting", Second International Conference on Computer Modeling and Simulation, vol.10, 2010, pp.275–277.

[8] D.X. Niu, Y. Wei, J. Li, J. Wu, "Analysis of electricity demand forecasting in In-ner Mongolia based on gray Markov model", International Conference on E-Business and E-Government, vol.10, 2010, pp.5082–5085.

[9] A. Setaiwan, I. Koprinska, V.G. Agelidis, "Very short-term electricity load demand forecasting using support vector regression", International Joint Conference on Neural Network, Atlanta, Georgia, USA, vol.1, 2009, pp.2888–2894.

[10] Gigoni L, Betti A, Crisostomi E, Franco A, Tucci M, Bizzarri F, Mucci D (2017) "Day-ahead hourly forecasting of power generation from photovoltaic plants" IEEE Trans Sustain Energy 9(2):831–842

[11] Liu L, Zhan M, Bai Y (2019) "A recursive ensemble model for forecasting the power output of photovoltaic systems" Solar Energy 189:291–298

[12] Visser L, AlSkaif T, Van Sark W (2019) "Benchmark analysis of day ahead solar power forecasting techniques using weather predictions", 2019 IEEE 46th photovoltaic specialists conference (PVSC), pp 2111–2116

[13] H. Kong, D. Deng, W. Wu, "Short-term photovoltaic power forecasting based on LSTM neural network," in Proc. IEEE PES General Meeting, 2017, pp. 1–5.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), 2005, pp. 2047–2052.

[16] H. Kong, D. Deng, W. Wu, "Short-term photovoltaic power forecasting based on LSTM neural network," in Proc. IEEE PES General Meeting, 2017, pp. 1–5.

[17] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM networks,” in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), 2005, pp. 2047–2052.

[19] Parvathareddy, S., Yahya, A., Amuhaya, L., Samikannu, R., & Suglo, R. S. (2025). A hybrid machine learning and optimization framework for energy forecasting and management. Energy Reports, 26, 103456.

[20] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12),11106-11115.

[21] Litao TANG, Zhiyong ZHANG, Jun CHEN, Linna XU, Jiachen ZHONG, Peisen YUAN. Research on Autoformer-based electricity load forecasting and analysis[J]. Journal of East China Normal University (Natural Science), 2023, 2023(5): 135-146.

[22] F. Saeed, A. Rehman, H. A. Shah, M. Diyan, J. Chen, and J. M. Kang, “SmartFormer: Graph-based transformer model for energy load forecasting,” Sustainable Energy Technologies and Assessments, vol. 73, p. 104133, Jan. 2025, doi: 10.1016/J.SETA.2024.104133.

[23] A. B. A. Ferreira, J. B. Leite, and D. H. P. Salvadeo, “Power substation load forecasting using interpretable transformer-based temporal fusion neural networks,” Electric Power Systems Research, vol. 238, p. 111169, Jan. 2025, doi: 10.1016/J.EPSR.2024.111169.

[24] Laiqi Zhao, Zhifeng Wu, Maoyu Du, “Improved Autuoformer Electricity load forecasting based on model fusion” Advances in Engineering Technology Research, Published: 2024-05-08, DOI: https://doi.org/10.56028/aetr.11.1.8.2024

[25] B. Lim, S. Arık, N. Loeff, and T. Pfister, “Temporal Fusion Transformers for interpretable multi-horizon time series forecasting,” International Journal of Forecasting, vol. 37, no. 4, pp. 1748–1764, Oct. 2021, doi: 10.1016/J.IJFORECAST.2021.03.012.

[26] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, “Time Series Analysis: Forecasting and Control”. in Wiley Series in Probability and Statistics. Wiley, 2015.

[27] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” Nature, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[28] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Networks, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[29] LeCun; Boser, B. E.; Denker, J. S.; Henderson, D.; Howard, R.; Hubbard, W. E.; Jackel, L. D. (1989). "Handwritten digit recognition with a back-propagation network". In Touretsky, David S. (ed.). Advances in Neural Information Processing Systems 2 (NIPS 1989).

[30] J. L. Elman, “Finding Structure in Time,” Cognitive Science, vol. 14, no. 2, pp. 179–211, 1990, doi: 10.1207/s15516709cog1402_1.

[31] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[32] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” Neural Networks, vol. 18, no. 5, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[33] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches,” in Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, D. Wu, M. Carpuat, X. Carreras, and E. M. Vecchi, Eds., Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111. doi: 10.3115/v1/W14-4012.

[34] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” May 19, 2016, arXiv: arXiv:1409.0473. doi: 10.48550/arXiv.1409.0473.

[35] C. Fan, S. Nie, L. Xiao, L. Yi, and G. Li, “Short-term industrial load forecasting based on error correction and hybrid ensemble learning,” Energy and Buildings, vol. 313, p. 114261, Jun. 2024, doi: 10.1016/J.ENBUILD.2024.114261.

[36] R. Das (2025), “Artificial Intelligence Based Fault Detection and Predictive Analysis for Smart Power Systems”, Western Carolina University

[37] A. Harris, R. Das and T. Kandil, "A Comparative Survey of AI-Driven Energy Demand Forecasting Across Diverse Sectors," *2025 57th North American Power Symposium (NAPS)*, Storrs, CT, USA, 2025, pp. 1-6, doi: 10.1109/NAPS66256.2025.11272390.

[38] Q. Ge et al., “Industrial Power Load Forecasting Method Based on Reinforcement Learning and PSO-LSSVM,” IEEE Transactions on Cybernetics, vol. 52, no. 2, 2022, doi: 10.1109/TCYB.2020.2983871.

[39] Y. Wang et al., “Short-term load forecasting for industrial customers based on TCN-LightGBM,” IEEE Transactions on Power Systems, vol. 36, no. 3, 2021, doi: 10.1109/TPWRS.2020.3028133.

[40] F. L. C. da Silva, K. da Costa, P. C. Rodrigues, R. Salas, and J. L. López-Gonzales, “Statistical and Artificial Neural Networks Models for Electricity Consumption Forecasting in the Brazilian Industrial Sector,” Energies, vol. 15, no. 2, 2022, doi: 10.3390/en15020588.

[41] Sravani Parvathareddy, Abid Yahya, Lilian Amuhaya, Ravi Samikannu, Raymond Sogna Suglo, “A hybrid machine learning and optimization framework for energy forecasting and management”, Results in Engineering, Volume 26, 2025, 105425, ISSN 2590-1230, https://doi.org/10.1016/j.rineng.2025.105425.

[42] H. Jiang, A. Wu, B. Wang, P. Xu, and G. Yao, “Industrial ultra-short-term load forecasting with data completion,” IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3017655.

[43] M. Tan, S. Yuan, S. Li, Y. Su, H. Li, and F. H. He, “Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning,” IEEE Transactions on Power Systems, vol. 35, no. 4, 2020, doi: 10.1109/TPWRS.2019.2963109.

[44] N. N. V. Nhat, D. N. Huu, and T. N. T. Hoai, “Evaluating the EEMD-LSTM model for short-term forecasting of industrial power load: A case study in Vietnam,” International Journal of Renewable Energy Development, vol. 12, no. 5, 2023, doi: 10.14710/ijred.2023.55078.

[45] C. Fan, S. Nie, L. Xiao, L. Yi, and G. Li, “Short-term industrial load forecasting based on error correction and hybrid ensemble learning,” Energy and Buildings, vol. 313, p. 114261, Jun. 2024, doi: 10.1016/J.ENBUILD.2024.114261.

[46] Z. Zhu et al., “A day-ahead industrial load forecasting model using load change rate features and combining FA-ELM and the AdaBoost algorithm,” Energy Reports, vol. 9, 2023, doi: 10.1016/j.egyr.2022.12.044.

[47] T. Walser and A. Sauer, “Typical load profile-supported convolutional neural network for short-term load forecasting in the industrial sector,” Energy and AI, vol. 5, 2021, doi: 10.1016/j.egyai.2021.100104.