

Article

Decomposition–Ensemble Learning for Power Load Forecasting: Analysis of Noise Sensitivity and Algorithmic Stability

Rongrong Chen, Miaodan Cheng, Chia-Wei Huang, Wei-Tai Hsu  and Chih-Chung Yang * 

School of Electronic and Electrical Engineering, Zhaoqing University, Zhaoqing 526061, China; 2018013014@zqu.edu.cn (C.-W.H.); 2018013018@zqu.edu.cn (W.-T.H.)

* Correspondence: 2019013033@zqu.edu.cn

Abstract

Precise short-term load forecasting is critical for smart grid reliability, yet the resilience of predictive models against ambient noise and sensor degradation remains insufficiently explored. This study systematically evaluates the noise sensitivity and operational stability of three forecasting paradigms (baseline LSTM, EEMD-LSTM, and CEEMD-LSTM) across varying signal-to-noise ratios (0–25 dB SNR) and sampling densities. Empirical analyses reveal a critical operational paradox: despite its theoretical sophistication, the CEEMD-LSTM architecture exhibits severe algorithmic instability and vulnerability to residual noise artifacts, yielding an RMSE standard deviation of ± 1918 MW under high-fidelity conditions. Conversely, the EEMD-LSTM framework demonstrates superior robustness. Acting as an effective noise-mitigating filter, it sustains a Mean Absolute Percentage Error (MAPE) below 3.5% across all noise regimes, notably achieving sub-3.0% errors under severe interference (5 dB SNR). Furthermore, evaluating memory-optimized sparse sampling exposes a predictive trade-off, whereby reduced sampling density substantially exacerbates forecasting errors. These findings indicate that the EEMD-LSTM model, underpinned by dense sampling, provides a highly stable, noise-resilient predictive mechanism for smart grids subjected to uncertain measurements.

Keywords: load forecasting; LSTM; EEMD; CEEMD; noise robustness; algorithmic stability; smart grid

1. Introduction

As the transition to a modern smart electric grid continues, the need for reliable and efficient electric delivery systems requires accurate predictive modeling for load forecasting of electricity [1,2]. The generation of the demand for electric load is very complex and has many aspects, including macro-level seasonal trends, local customer behavior, and outside environmental influences [3,4]. Historically, limitations in load modeling capabilities often hindered efficient grid operations. Consequently, accurate electrical load forecasting has become crucial today for the optimal distribution of power and for making sound investment decisions [1,5,6].

Over the last several years, there has been a massive increase in the use of artificial intelligence (AI), which has created new opportunities for improving the way we model sequences [3]. Recent advances in deep learning have introduced sophisticated architectures, including transformer-based models [7], hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) frameworks [2], and attention mechanisms [8,9] for time-series forecasting. This study deliberately restricts its predictive framework to LSTM



Academic Editor: Armando Pires

Received: 17 May 2026

Revised: 11 July 2026

Accepted: 21 July 2026

Published: 26 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

networks for two reasons. First, it ensures a strictly controlled environment to isolate the denoising efficacy of EEMD and CEEMD. Second, LSTM is a mature model widely used in utility forecasting, providing a reliable baseline. Moreover, by evaluating noise resilience under sensor degradation, we establish an empirical benchmark that can inform future work with transformers and attention-based architectures. The LSTM network is one of the first models developed to solve the problem of vanishing gradients typically found within traditional recurrent networks [3,10]. LSTMs can outperform traditional methods in time-series predictions across multiple industries due to their ability to handle very large volumes of data [3]. Despite these benefits, the ability of traditional neural networks to perform short-term load forecasting (STLF) of local fluctuations that are created by the interaction of ambient noise and the structure is still a big challenge [2,11]. One technique that has been increasingly used over the last few years to address the complex behavior associated with non-stationary characteristics of load signals is Empirical Mode Decomposition (EMD) [12]. This technique reduces the complexity of the data while isolating Intrinsic Mode Functions (IMF) to improve the representational capability of predictive models [13–15].

As a result, hybrid frameworks combining signal decomposition techniques are becoming an increasingly popular area of research [16–21]. In addition to the current literature utilizing various complex transformer-style architecture models (e.g., Informer, Autoformer) for sequence prediction [8,9], state-of-the-art LSTM-based networks have provided equally accurate or even more accurate predictions than some of the former models when using medium-length sequences due to the computational efficiency limits tied to normal operation of grids [13,22].

The emergence of hybrid modeling has led to a large volume of research and theory development, but there remain many knowledge gaps. In particular, there are three main areas of study that could be improved: (1) a number of studies have focused on the performance of models under ideal (no noise) conditions, yet few have extensively evaluated how these models will perform with the addition of noise; (2) there is little exploration as to the level of stability (predictability) of Ensemble Empirical Mode Decomposition (EEMD) and Complete Ensemble Empirical Mode Decomposition (CEEMD).

While a broad landscape of forecasting methods exists (including Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machine (SVM), Random Forest, and eXtreme Gradient Boosting (XGBoost)), this study focuses on three LSTM-based architectures to enable controlled comparison of decomposition strategies (no decomposition, EEMD, CEEMD) while holding the predictor architecture constant. This focused investigation allows us to isolate the impact of decomposition techniques on noise robustness and algorithmic stability, which would be confounded by comparing fundamentally different model families. This study is designed to fill the gaps identified in previous research about noise robustness and sampling density by performing a detailed comparative evaluation of three forecasting paradigms: baseline LSTM, EEMD-LSTM, and CEEMD-LSTM. Our investigation encompasses (1) noise robustness analysis across a comprehensive range of Signal-to-Noise Ratios (SNR) from 0 dB (extreme interference) to 25 dB (near-clean conditions), with explicit distinction between noise-free baseline and degraded signal conditions; (2) sampling density evaluation examining forecasting performance under 100%, 75%, 50%, and 25% sampling rates to quantify the trade-off between data acquisition costs and predictive accuracy; (3) algorithmic stability assessment through repeated trials with controlled random seeds to characterize model variance and reliability; (4) cross-dataset validation utilizing the public dataset source from the Global Energy Forecasting Competition 2014 (GEFCom2014) Load Track benchmark [23] to establish model generalizability.

Our empirical findings reveal a critical operational paradox: despite its theoretical sophistication, the CEEMD-LSTM architecture exhibits severe instability and vulnerability to residual noise artifacts, while the EEMD-LSTM framework demonstrates superior robustness across all tested conditions. These results have significant implications for the deployment of forecasting systems in real-world smart grid environments where measurement quality and data availability are inherently uncertain.

The remainder of this paper is organized as follows: Section 2 reviews related work on load forecasting and hybrid decomposition-learning models. Section 3 describes the theoretical foundations of LSTM, EEMD, and CEEMD. Section 4 details the experimental methodology, including dataset description, noise injection protocols, and evaluation metrics. Section 5 discusses the implications of the findings, and Section 6 concludes the paper.

2. Related Work

2.1. Traditional Load Forecasting Methods

The development of load forecasting techniques has seen a shift from traditional statistical models to advanced machine learning methodologies. Many early methods employed time-series analyses, regression models, and autoregressive integrated moving average (ARIMA) methods for load forecasting [5,13]. While these models generally provided an acceptable level of accuracy for stable environments, they struggled to capture the nonlinear and non-stationary aspects of today's electrical system adequately [4,13].

2.2. Deep Learning for Load Forecasting

The introduction of deep learning has dramatically changed the way that load forecasting is done. RNNs (recurrent neural networks) and their variations (especially LSTMs) have shown to be extremely effective at modeling the time-dependence present in load data [15,24]. Because of their specialized gating mechanisms, LSTMs help to solve the problem of vanishing gradients, which allows for more effective learning of long term dependencies [4,10]. Studies indicate that LSTM-based models can produce very accurate predictions when used for short-term load forecasting [24].

2.3. Hybrid Decomposition-Learning Models

Recognizing the limitations of standalone deep learning models in analyzing non-stationary signals, researchers have developed hybrid approaches that integrate signal decomposition with neural networks. Zheng et al. [16], for instance, presented an EMD-LSTM model that decomposes load signals into intrinsic mode functions (IMFs) before entering them into LSTM networks. Nhat et al. [17] evaluated the EEMD-LSTM model for forecasting industrial power loads and found it to have significantly greater forecasting accuracy compared to baseline LSTM models. Pei et al. [18] utilized the CEEMD-IPSO-LSTM (Complete Ensemble Empirical Mode Decomposition-Improved Particle Swarm Optimization-Long Short-Term Memory) model to forecast loads in refrigerated display cabinets, and Liu et al. [19] examined the CEEMDAN-LSTM model for predicting loads over short periods of time. Additionally, Elsayed et al. [20] combined CEEMDAN with transformer neural networks, and Heng et al. [21] introduced a CEEMDAN-TCN-LSTM (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise-Temporal Convolutional Network-Long Short-Term Memory) hybrid model.

Despite the growing body of literature on decomposition-ensemble forecasting, systematic evaluation of noise robustness and algorithmic stability remain limited. Most studies evaluate models under clean, idealized conditions without considering the impact of measurement noise, sensor degradation, or data sparsity that characterize real-world smart grid

deployments. Table 1 positions this work within the existing literature, highlighting that while prior studies have demonstrated the potential of decomposition–ensemble methods, none have systematically evaluated noise robustness, algorithmic stability, and sampling density trade-offs across multiple datasets. Our study fills this critical gap by providing comprehensive empirical evidence for model selection in real-world smart grid deployments.

Table 1. Comparison of decomposition–ensemble load forecasting studies.

Study	Method	Dataset	Key Finding	Limitation	This Work's Improvement
Qiu et al. [25]	EMD-LSTM	Regional load data (China)	EMD preprocessing improves LSTM accuracy by 15–20%	No noise robustness evaluation; single dataset	Systematic noise sensitivity analysis (0–25 dB SNR); cross-dataset validation
Zhang et al. [26]	EEMD for wind power	Wind farm data	EEMD-LSTM reduces MAPE to 4.2% for wind forecasting	Focus on wind power, not load; no sampling density analysis	Load forecasting focus; sampling density evaluation (100%, 75%, 50%, 25%)
Li et al. [27]	CEEMDAN-LSTM	Microgrid load	CEEMDAN-LSTM achieves 3.8% MAPE on microgrid data	No comparison with EEMD; no stability analysis	Direct EEMD vs. CEEMD comparison; algorithmic stability assessment (RMSE SD)
Kong et al. [28]	CEEMDAN ensemble	Industrial load	Ensemble approach improves electrometallurgical load forecasting	Limited to specific industrial application	General residential/commercial load; broader applicability
Wang et al. [29]	CEEMDAN-BiLSTM	Regional load (China)	BiLSTM with CEEMDAN achieves 2.9% MAPE	No noise robustness testing; clean data only	Comprehensive noise injection protocol; extreme 0 dB SNR testing
This work	EEMD-LSTM vs. CEEMD-LSTM	Tianchi + GEF-Com2014	EEMD-LSTM sustains <3.5% MAPE across 0–25 dB SNR; CEEMD-LSTM exhibits high instability (RMSE SD ± 1918)	—	First systematic noise robustness and stability comparison; cross-dataset validation; sampling density analysis

While these studies demonstrate the potential of decomposition–ensemble methods, none systematically evaluate noise robustness across a wide SNR range (0–25 dB), quantify algorithmic stability via RMSE standard deviation, or validate findings on independent datasets. This work addresses these gaps by providing the first comprehensive noise sensitivity analysis comparing EEMD-LSTM and CEEMD-LSTM under controlled interference conditions.

2.4. Distinction from State-of-the-Art Paradigms

To position the proposed framework within the short-term load forecasting (STLF) landscape, we contrast our decomposition–ensemble approach with three prevailing state-of-the-art (SOTA) paradigms:

- **Multi-Modal Fusion:** Multi-modal frameworks integrate diverse exogenous streams (e.g., meteorology, social data) but rely heavily on the synchronous alignment of heterogeneous sensors [30]. In real-world smart grids, telemetry failure, communication latency, and sensor degradation easily cause data gaps or cross-modality noise propagation [15]. Conversely, our architecture adopts a univariate setup dependent strictly on historical load, isolating the pure nonlinear denoising capabilities of EEMD and CEEMD without confounding variables [14,31].
- **Dual-Attention BiLSTM:** Bidirectional networks with spatial-temporal attention effectively capture sequence dynamics but lack explicit frequency-domain filtering [32,33]. In high-noise environments (0–10 dB SNR), attention layers tend to overfit on stochas-

tic artifacts, unintentionally amplifying structural distortions [22]. The proposed EEMD-LSTM circumvents this by filtering high-frequency volatility into early Intrinsic Mode Functions (IMFs), structurally shielding downstream LSTM layers to preserve stability [13,17].

- KAN + Transformer: Combining Transformer backbones with Kolmogorov–Arnold Networks (KAN) provides excellent model transparency but incurs a heavy semi-quadratic $O(n \log n)$ or quadratic $O(n^2)$ time complexity and high parameter density [8,22,32]. This introduces severe memory overhead and inference latency on resource-constrained edge grid infrastructure like advanced metering infrastructure (AMI) [22]. In contrast, our dual-stacked LSTM predictor maintains a linear time complexity of $O(n)$, serving as a highly efficient, lightweight, and noise-resilient alternative for decentralized operations [8,15,22].

3. Methodology

3.1. Ensemble Empirical Mode Decomposition (EEMD)

Standard empirical mode decomposition (EMD) techniques, while highly adept at extracting oscillatory features from complex signals, inherently suffer from mode mixing—a phenomenon where disparate frequency scales manifest within a single Intrinsic Mode Function (IMF), or conversely, similar scales partition across multiple IMFs [12,31]. To circumvent this limitation, the EEMD algorithm injects independent realizations of Gaussian white noise into the original signal, thereby establishing a uniform time-frequency reference that enforces the natural segregation of frequency scales [31].

The formulation dictates that for the i -th ensemble trial, a specific white noise sequence $w_i(t)$ is superimposed onto the macroscopic load signal $x(t)$, yielding a noise-assisted signal $x_i(t)$ [31,34]:

$$x_i(t) = x(t) + w_i(t), \quad i = 1, 2, \dots, N \quad (1)$$

where N denotes the predefined ensemble size. Independent sifting of each $x_i(t)$ yields J distinct IMFs $c_{ij}(t)$ and a residual trend $r_i(t)$. The ultimate IMFs are derived via ensemble averaging across all trials, capitalizing on the statistical property that uncorrelated white noise realizations asymptotically cancel each other. This noise-assisted protocol successfully isolates temporally stable IMFs while preserving the intrinsic load patterns required for subsequent sequential modeling.

3.2. Complete Ensemble Empirical Mode Decomposition (CEEMD)

Although EEMD mitigates mode mixing, it operates on the assumption that the ensemble mean of the injected noise will perfectly converge to zero [31]. In practical computational environments where ensemble sizes are finite, residual noise frequently persists, distorting the resultant IMFs and compromising signal reconstruction fidelity [14]. CEEMD mathematically resolves this via the injection of complementary white noise pairs—possessing identical magnitude but inverted polarity—ensuring exact noise cancellation during the averaging phase [14].

The original signal X and the auxiliary noise vector N form complementary mixture matrices [14]:

$$M_1 = X + N \quad \text{and} \quad M_2 = X - N. \quad (2)$$

Decomposing M_1 and M_2 simultaneously yields twin IMF sets embedding diametrically opposed noise structures. The superimposition of these sets theoretically constructs a noise-compensated IMF ensemble that more faithfully captures the true non-linear dynamics of the underlying signal, simultaneously reducing the computational iterations required relative to standard EEMD.

3.3. Component-Wise Sequential Modeling via LSTM

To capture the temporal dependencies embedded within the decomposed IMFs, this framework employs Long Short-Term Memory (LSTM) networks. Standard recurrent architectures exhibit severe gradient instability (vanishing or exploding gradients) when processing extended time series [10]. LSTMs structurally resolve this through specialized memory blocks governed by three explicit gating mechanisms: an input gate i_t , a forget gate f_t , and an output gate o_t . These gates dynamically modulate the state evolution, as defined by [3,35]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (6)$$

$$h_t = o_t \odot \tanh(C_t) \quad (7)$$

where σ represents the sigmoid activation function, $\odot$ denotes the Hadamard product, C_t is the cell state, and W and b constitute the learnable weight matrices and bias vectors.

In the proposed architecture, each decomposed IMF is fed into an independent dual-stacked LSTM network. The hierarchy features a 128-unit primary hidden layer followed by a 64-unit secondary layer, with intermediate dropout regularization (rate = 0.2) applied to suppress overfitting. Optimization is driven by the Adam algorithm (learning rate = 0.001) against a Mean Squared Error (MSE) objective function.

3.4. The Integrated Decomposition–Ensemble Pipeline

The overarching predictive framework operates as a structured, four-stage analytical pipeline:

A. Signal Decomposition: The macroscopic load time series $x(t)$ is disaggregated via EEMD or CEEMD into n distinct IMFs and a residual deterministic trend $r(t)$.

B. Component-Wise Forecasting: Each resulting IMF is autonomously modeled by its respective LSTM topology to capture scale-specific temporal features over a defined lookback window w (30 days).

C. Forecast Reconstruction: The finalized systemic prediction $\hat{S}(t+h)$ is mathematically synthesized via the linear aggregation of all individual component forecasts:

$$\hat{S}(t+h) = \sum_{i=1}^n \hat{y}_i(t+h) + \hat{r}(t+h) \quad (8)$$

where $\hat{y}_i(t+h)$ is the predicted value at horizon h .

D. Validation: Predictive accuracy is ultimately benchmarked employing standardized error metrics including root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). This combinatorial approach ensures that high-frequency stochasticity is appropriately isolated, thereby permitting the LSTM arrays to process denoised, dynamically stable inputs.

The procedure for the short-term load forecasting process is shown in Figure 1. The original load data are first decomposed through EEMD or CEEMD to produce $\text{IMF}_1, \text{IMF}_2, \dots, \text{IMF}_n$, together with residual components. Each component is then modeled separately using an LSTM neural network. The trained models generate predictions for their respective components ($\text{LSTM}_1, \text{LSTM}_2, \dots, \text{LSTM}_n$). The predicted components are then reconstructed and combined to produce the final forecast. $\text{Nstd} = 0.1$ denotes the

ratio between the standard deviation of the added noise and the standard deviation of the original signal. NE = 200 represents the number of EMD ensemble iterations.

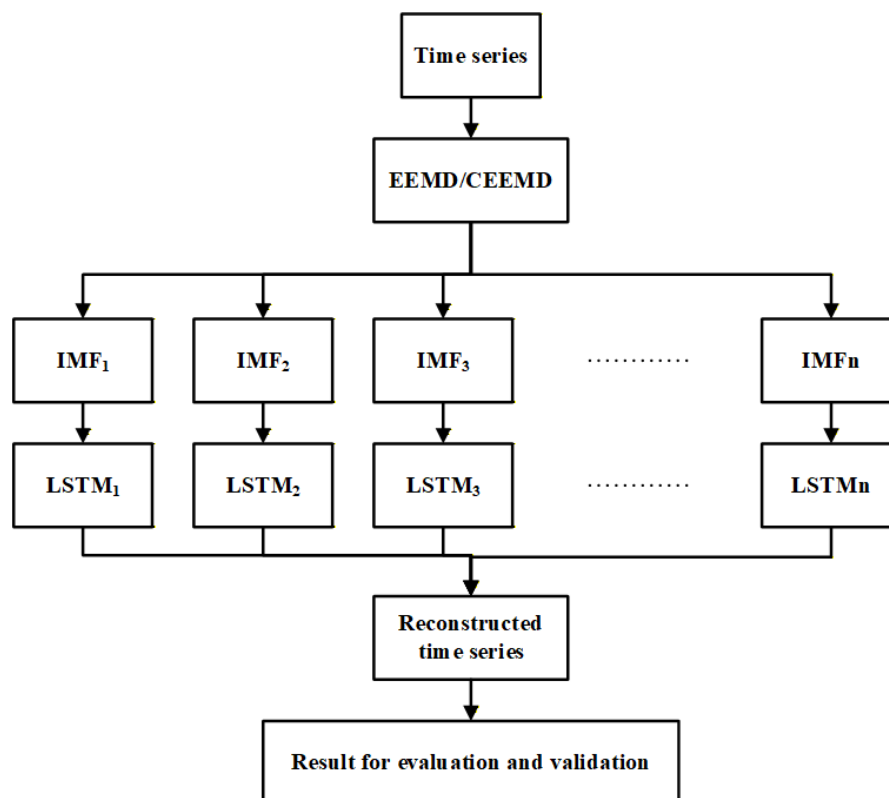


Figure 1. Structure of the short-term load forecasting procedure.

3.5. Dataset Description

This study employs two publicly available datasets to ensure reproducibility and generalizability. The primary dataset is sourced from the Tianchi Big Data Competition platform [36], containing hourly electrical load measurements from a regional power grid in China. The dataset spans one year (8760 hourly observations) and captures diverse load patterns including residential, commercial, and light industrial consumption. Load values range from approximately 3.29 MW to 369.60 MW, exhibiting strong diurnal and weekly periodicities characteristic of urban power systems. For cross-dataset validation, we employ the GEFCom2014 Load Track benchmark dataset [23], which contains hourly load data from a U.S. utility spanning multiple years. This dataset is widely recognized in the forecasting community and provides an independent test of model generalizability across different geographical regions, grid characteristics, and consumption patterns. Table 2 provides comprehensive metadata for both datasets, facilitating reproducibility and enabling researchers to understand the data characteristics and preprocessing requirements.

Table 2. Dataset characteristics.

Feature	Data Type	Unit	Description	Statistical Range	Source
Tianchi Dataset					
Timestamp	DateTime	—	Hourly timestamp (YYYY-MM-DD HH:MM:SS)	2020-01-01 00:00 to 2023-08-31 23:45	Tianchi Platform
Load	Continuous	MW	Aggregated electrical load demand	3.29–369.60 MW	Smart meter aggregation
Temperature	Continuous	°C	Ambient temperature	−5.2–38.6 °C	Weather station
Humidity	Continuous	%	Relative humidity	18.3–95.7%	Weather station
Holiday	Binary	—	Holiday indicator (0 = working day, 1 = holiday)	(0, 1)	National calendar
Day of Week	Categorical	—	Day of week (1 = Monday, 7 = Sunday)	(1, 2, 3, 4, 5, 6, 7)	Derived from timestamp
Hour of Day	Categorical	—	Hour of day (0–23)	(0, 1, ..., 23)	Derived from timestamp
GEFCom2014 Dataset					
Timestamp	DateTime	—	Hourly timestamp	2004-01-01 00:00 to 2014-12-31 24:00	GEFCom2014
Load	Continuous	MW	Aggregated electrical load demand	1811.0–5506.0 MW	Utility SCADA system
Temperature	Continuous	°F	Ambient temperature (25 weather stations)	12.4–102.8 °F	Weather stations
Holiday	Binary	—	Holiday indicator	(0, 1)	U.S. federal calendar

3.6. Evaluation Metrics

Model performance is quantified using four standard metrics. Mean Absolute Percentage Error (MAPE) is defined as:

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

where N is the number of test samples, y_i is the actual load at time i , and $\hat{y}_i$ is the predicted load. MAPE provides a scale-independent measure of forecasting accuracy, facilitating comparison across datasets with different load magnitudes. Root Mean Squared Error (RMSE) penalizes large errors more heavily than small errors, making it sensitive to outliers and peak load prediction accuracy. RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}. \quad (10)$$

Mean Absolute Error (MAE) provides a linear penalty for errors and is more robust to outliers than RMSE. MAE is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (11)$$

Coefficient of Determination (R^2) is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

where $\bar{y}$ is the mean of the actual load values. R^2 measures the proportion of variance in the load data explained by the model, with values closer to 1 indicating better fit. To assess algorithmic stability, we compute the standard deviation of RMSE across multiple independent trials:

$$\text{RMSE SD} = \sqrt{\frac{1}{M-1} \sum_{j=1}^M (\text{RMSE}_j - \overline{\text{RMSE}})^2} \quad (13)$$

where M is the number of trials (30 in this study) and $\overline{\text{RMSE}}$ is the mean RMSE across trials. Lower RMSE SD indicates higher algorithmic stability and reproducibility.

4. Experimental Setup and Results

4.1. Dataset Configuration and Evaluation Protocol

For the purpose of providing empirical validation for the proposed forecasting architectures, this study used a source of empirical data, a longitudinal dataset of power consumption in a specific region of China that spans from January 2020 until December 2023. The dataset contains 128,156 hourly observations (at 15-min intervals) over four years. We segment the data into 30-day sequences, each comprising 2880 time steps (since 15-min resolution yields 96 points per day) [36]. The dataset was trained on the front 70%, keeping the tail 30% for validation plus final testing. For each model configuration and noise level, 30 independent trials with random seeds were performed. The preprocessing pipeline guarantees rigorous experimental integrity and numerical stability by employing min-max normalization mapped exclusively from the training split to prevent data leakage, utilizing a 24 h overlapping sliding window to structure input trajectories for the LSTM architecture, integrating one-hot encoded temporal features via feature concatenation, and ensuring data continuity across both benchmarks through targeted linear interpolation for the sparse missing entries (<0.1%) in the GEFCom2014 framework.

Two separate sampling methods were implemented to assess the relationship between computational effort and ability to predict. One was the Dense Sampling configuration (Stride = 1), which allows for a continual high frequency of temporal information within the signal stream; the second was the Memory Optimized Sparse Sampling configurations (Stride = 4), where the sampling of data reduces structural (i.e., physical) memory constraints by 75%. A critical aspect of this study is the clear distinction between clean (noise-free) conditions and noisy conditions characterized by specific SNR levels:

- **Clean Condition:** The original load signal without any added noise. This represents the ideal scenario with perfect sensor measurements and no electromagnetic interference. Mathematically, $\text{SNR} \rightarrow \infty$ (infinite signal-to-noise ratio).
- **Noisy Conditions (0 dB, 5 dB, 10 dB, 15 dB, 20 dB, 25 dB SNR):** Additive White Gaussian Noise (AWGN) is injected into the clean signal to simulate various levels of measurement degradation.

The SNR is defined as:

$$\text{SNR(dB)} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \quad (14)$$

where P_{signal} is the power of the original signal and P_{noise} is the power of the added noise. 0 dB SNR does not represent a clean signal. At 0 dB SNR, the signal power equals the noise power ($P_{\text{signal}} = P_{\text{noise}}$), representing an extreme interference scenario where the noise is as strong as the signal itself. This is a severe degradation condition, not a baseline.

4.1.1. Noise Injection Procedure

For each specified SNR level, AWGN is generated and added to the clean load signal as follows:

1. Compute the power of the clean signal: $P_{\text{signal}} = \frac{1}{N} \sum_{i=1}^N x_i^2$
2. Determine the required noise power: $P_{\text{noise}} = P_{\text{signal}} / 10^{\text{SNR}/10}$
3. Generate AWGN with zero mean and variance $\sigma^2 = P_{\text{noise}}$
4. Add the noise to the clean signal: $y_i = x_i + n_i$

This protocol ensures consistent and reproducible noise injection across all experiments.

4.1.2. Justification for SNR Range (0–25 dB)

The selected SNR range (0–25 dB) is justified by real-world smart grid measurement conditions:

- **0 dB SNR:** Represents extreme electromagnetic interference (EMI) scenarios in industrial environments with heavy machinery, arc furnaces, or high-voltage switching operations. While rare, such conditions can occur during equipment faults or in proximity to high-power industrial loads [37,38].
- **5–10 dB SNR:** Corresponds to severe interference in urban substations with dense communication infrastructure, aging sensors, or during adverse weather conditions (lightning, ice storms) that degrade measurement quality [39,40].
- **15–20 dB SNR:** Represents moderate noise levels typical of standard grid environments with conventional smart meters and AMI communication systems. This range aligns with IEEE standards for power quality monitoring [37].
- **25 dB SNR:** Approaches near-clean conditions achievable with high-quality sensors, shielded communication channels, and well-maintained infrastructure. This serves as a reference point close to, but distinct from, the clean baseline [37].

This comprehensive SNR range enables evaluation of model robustness across the full spectrum of operational conditions, from ideal to severely degraded measurements.

4.2. Exploratory Data Analysis

Before model development, a comprehensive exploratory data analysis (EDA) was conducted to characterize the statistical properties and underlying distribution profiles of the load data. As detailed in Table 3, both datasets exhibit moderate dispersion, with a coefficient of variation (CV) of 0.212 for the Tianchi dataset and 0.175 for the GEFCom2014 dataset. In terms of distribution symmetry, both series are approximately symmetric, as evidenced by near-zero skewness values of 0.03 and 0.00, respectively. However, their tail behaviors show a notable divergence: the Tianchi dataset features a positive kurtosis of 0.90, indicating a leptokurtic profile with a sharper peak and heavier tails relative to a normal distribution. Conversely, the GEFCom2014 dataset displays a negative kurtosis of -0.52 , suggesting a platykurtic distribution characterized by a flatter peak and lighter tails. These varied structural properties underscore the diverse operational behaviors inherent across different regional power grid systems.

Table 3. Time-series statistical characteristics.

Statistic	Tianchi Dataset	GEFCom2014 Dataset
Mean (MW)	213.68	3308.0
Standard Deviation (MW)	45.33	579.8
Minimum (MW)	3.29	1811.0
Maximum (MW)	369.60	5506.0
Median (MW)	209.41	3382.0
Skewness	0.03	0.00
Kurtosis	0.90	−0.52
Coefficient of Variation	0.212	0.175

The autocorrelation function (ACF) analysis results demonstrate that the electricity load data exhibits significant temporal correlation and robust periodicity in Table 4. The autocorrelation coefficients across all critical lag horizons far exceed the 95% confidence bounds of random white noise (± 0.0056 for the Tianchi dataset and ± 0.0071 for the GEFCom2014 dataset), firmly establishing their statistical significance. Specifically, at a 1 h lag interval (corresponding to lag 4 in the Tianchi dataset and lag 1 in the GEFCom2014 dataset), the ACF values reach 0.9691 and 0.9549, respectively, indicating an exceptionally strong intra-hour momentum and continuity between adjacent periods. Moreover, at the 24 h (1 day) lag point, the coefficients remain highly pronounced at 0.9305 for the Tianchi dataset and 0.9214 for the GEFCom2014 dataset, while at the 168 h (7 day) lag point, they register at 0.8688 and 0.8935, respectively. This profile strongly confirms the presence of highly significant diurnal and weekly seasonal cycles. Remarkably, even at the extended 336 h (14 day) lag horizon, the autocorrelation values persist at 0.7765 and 0.8731, respectively, suggesting a persistent bi-weekly operational trajectory inherently tied to macro-level institutional schedules or social production cycles. The high autocorrelation at short lags fully justifies the use of Long Short-Term Memory (LSTM) networks, as this model architecture can effectively capture these complex temporal dependencies. Concurrently, the strong periodicity exhibited at lags of 24 and 168 h further suggests the introduction of hybrid sequence decomposition techniques within the modeling pipeline to isolate and specifically process these multi-scale periodic patterns.

Table 4. ACF statistic extraction report with specific lag time.

Lag Time (Hours)	Tianchi Dataset	GEFCom2014 Dataset
1 H	0.9691	0.9549
24 H (1 Day)	0.9305	0.9214
168 H (7 Days)	0.8688	0.8935
336 H (14 Days)	0.7765	0.8731
95% confidence Bound	0.0056	0.0071

Figure 2 shows the exploratory data analysis for the Tianchi dataset. On the left of Figure 2, a two-week continuous electricity load trajectory demonstrates the strong diurnal peak-valley patterns, regular weekly consumption cycles, and diverse behavioral traits characteristic of urban power demand. On the right of Figure 2, the autocorrelation function (ACF) profile evaluated up to a 336 h (14 day) extended lag horizon. The prominent periodic peaks synchronized at 24 h and 168 h milestones significantly exceed the 95% random white noise confidence bounds (dashed red lines), firmly establishing the statistical prerequisites and structural justification for deploying sequential deep learning networks and hybrid sequence decomposition frameworks.

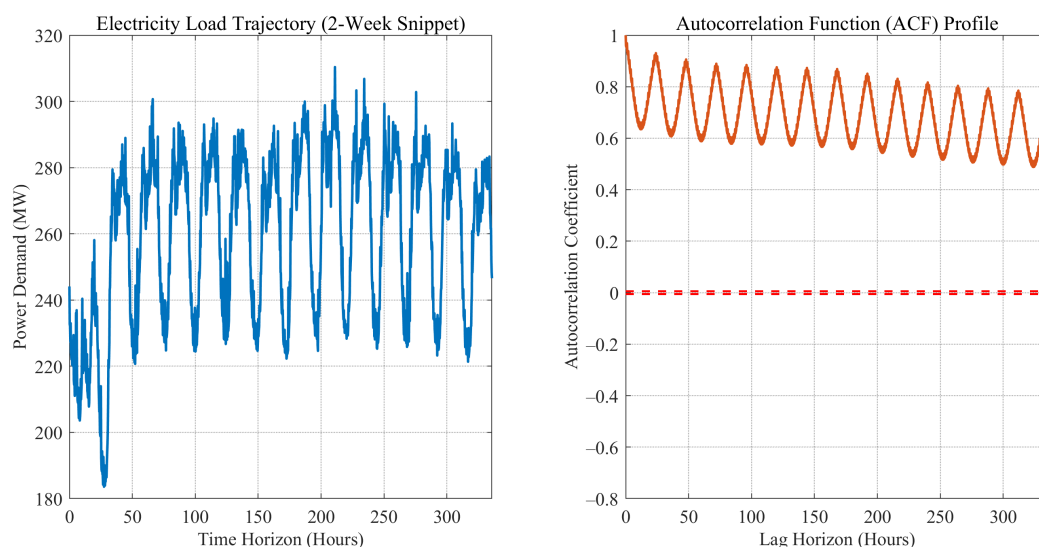


Figure 2. Load trends and periodicity for the Tianchi dataset.

The Augmented Dickey–Fuller (ADF) unit root test was performed to rigorously evaluate the stationarity profiles of the load series in Table 5. As summarized in Table 5, the empirical test statistics for the Tianchi and GEFCom2014 datasets register at -10.5617 and -7.2810 , respectively. Both values are significantly more negative than the critical thresholds at the 1% significance level (-2.5662) and the 5% significance level (-1.9416). Given that the calculated p -values for both series are equal to 1.0000×10^{-3} (which is strictly below the standard $\alpha = 0.05$ threshold), the null hypothesis of an underlying unit root is confidently rejected. These outcomes firmly establish that both electricity load datasets exhibit robust stationarity, satisfying the statistical prerequisites required to proceed with the subsequent signal decomposition and sequential modeling frameworks.

Table 5. Augmented Dickey–Fuller (ADF) test results.

Metric/Statistic	Tianchi Dataset	GEFCom2014 Dataset
ADF Test Statistic (t-stat)	-10.5617	-7.2810
p -value	0.001	0.001
Critical Value (1% Level)	-2.5662	-2.5662
Critical Value (5% Level)	-1.9416	-1.9416
Conclusion (alpha = 0.05)	Stationary (Reject H0)	Stationary (Reject H0)

Note: If Test Statistic < Critical Value and p -value < 0.05, H0 (unit root) is rejected.

To evaluate the linear dependencies between the current electrical load and its historical antecedents, a Pearson correlation coefficient analysis was conducted across critical lag horizons. As summarized in Tables 6 and 7, the current load (Y_t) exhibits an exceptionally strong positive linear correlation with the 1 h lagged load, registering at 0.96903 for the Tianchi dataset and 0.95488 for the GEFCom2014 dataset. Concurrently, correlation coefficients register at 0.93072 (Tianchi) and 0.92157 (GEFCom2014) for the 24 h lag, and 0.87248 (Tianchi) and 0.89527 (GEFCom2014) for the 168 hour lag, respectively. These strong statistical relationships fully validate the rationality of configuring the predictive framework as a univariate model dependent on historical load curves, ensuring that both immediate sequence dynamics and global cyclical variations are effectively captured by the downstream network arrays.

Table 6. Pearson correlation coefficients matrix (Tianchi dataset).

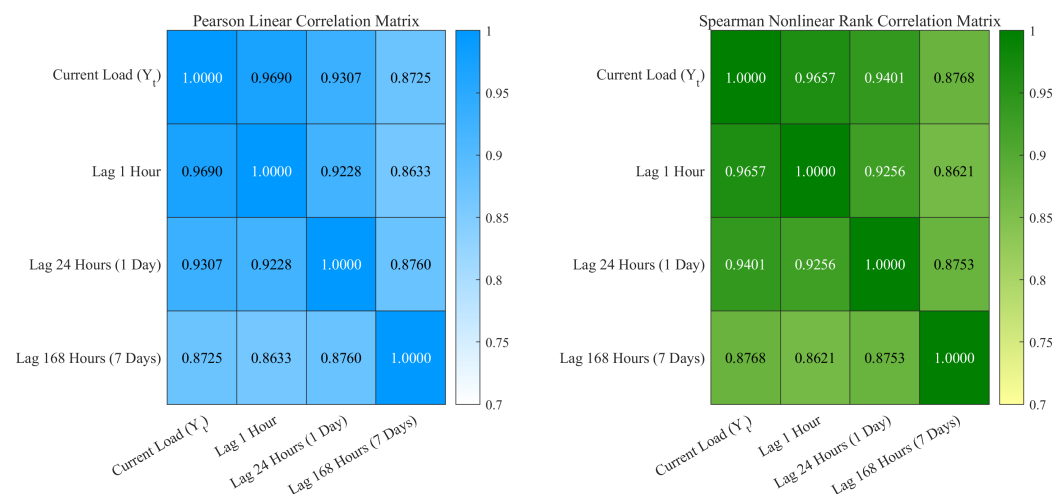
	Current Load (Y_t)	Lag 1 H	Lag 24 H	Lag 168 H
Current Load (Y_t)	1	0.96903	0.93072	0.87248
Lag 1 H	0.96903	1	0.92284	0.86334
Lag 24 H	0.93072	0.92284	1	0.87599
Lag 168 H	0.87248	0.86334	0.87599	1

Table 7. Pearson correlation coefficients matrix (GEFCom2014 dataset).

	Current Load (Y_t)	Lag 1 H	Lag 24 H	Lag 168 H
Current Load (Y_t)	1	0.95488	0.92157	0.89527
Lag 1 H	0.95488	1	0.88138	0.85131
Lag 24 H	0.92157	0.88138	1	0.85673
Lag 168 H	0.89527	0.85131	0.85673	1

Note: Coefficients range from -1 to 1 , where 1 indicates perfect positive linear correlation.

Comparative statistical dependency matrices across critical historical lag horizons are shown in Figure 3. The left of Figure 3 shows the Pearson linear correlation matrix quantifying the explicit linear relationships between the current power demand (Y_t) and its multi-scale historical antecedents, validating the mathematical rationality of the configured univariate modeling strategy. On the right of Figure 3, Spearman nonlinear rank correlation matrix is shown capturing the underlying monotonic non-linear co-dependencies within the load sequences. The numerical variations between the linear and non-linear profiles highlight the non-stationary complexities and multi-scale volatility embedded within the smart grid system, underscoring the necessity of front-end adaptive denoising preprocessing.

**Figure 3.** Linear vs. nonlinear correlation matrices for the Tianchi dataset.

4.3. Empirical Findings and the CEEMD Operational Paradox

This evaluation assesses the performance of three models—the baseline LSTM, EEMD-LSTM, and CEEMD-LSTM—across seven distinct noise conditions, such as, clean condition (noise-free), 0 dB SNR (extreme noise), 5 dB SNR (severe noise), 10 dB SNR (high noise), 15 dB SNR (moderate noise), 20 dB SNR (low noise), and 25 dB SNR (very low noise). To ensure statistical reliability, 30 independent trials with unique random seeds and noise realizations are conducted for each specific noise level. Finally, the models' performance is quantified using four key metrics (MAPE, RMSE, MAE), with the results reported as the mean plus or minus the standard deviation.

To assess the impact of data sparsity, we evaluate model performance under four sampling densities: (1) 100% sampling: all 8760 hourly samples (1 h intervals); (2) 75% sampling: randomly sampled 75% of hourly data (6570 samples, effective ~ 1.33 hour intervals); (3) 50% sampling: randomly sampled 50% of hourly data (4380 samples, effective ~ 2 h intervals); (4) 25% sampling: randomly sampled 25% of hourly data (2190 samples, effective ~ 4 h intervals). For each sampling density, the same train–test split ratio (80–20%) is maintained. Missing timestamps are handled by linear interpolation during preprocessing. Ten independent trials with different random sampling patterns are conducted for each density level.

The evaluation across these metrics revealed a clear performance hierarchy among the models, while also exposing a significant operational limitation in the CEEMD architecture. The models were evaluated using three different performance metrics (see Figures 4–6) across multiple noise levels.

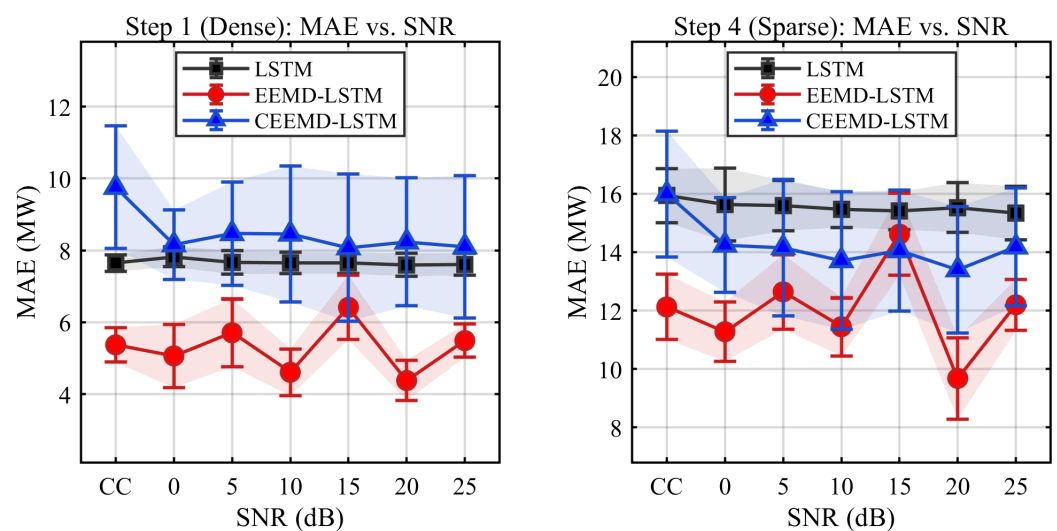


Figure 4. Noise sensitivity analysis for the MAE.

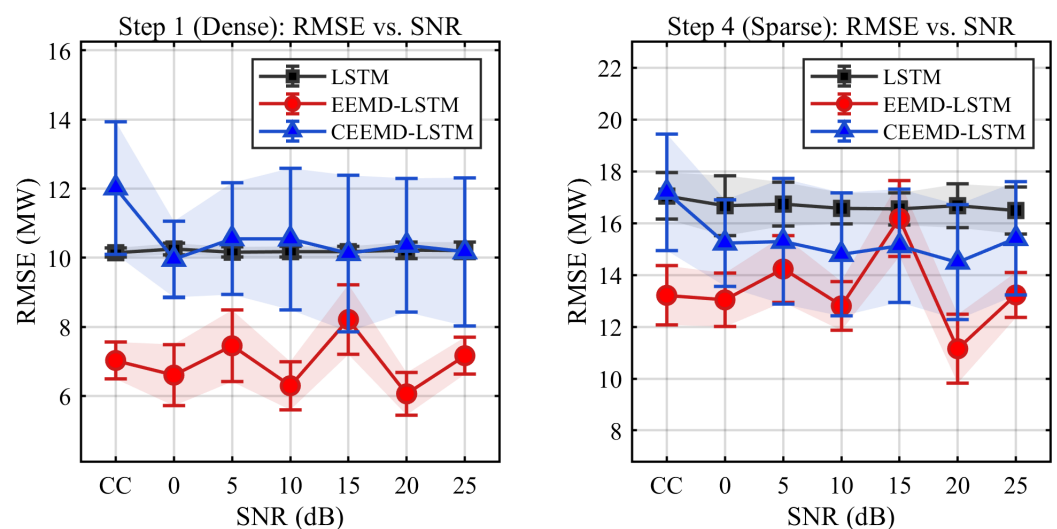


Figure 5. Noise sensitivity analysis for the RMSE.

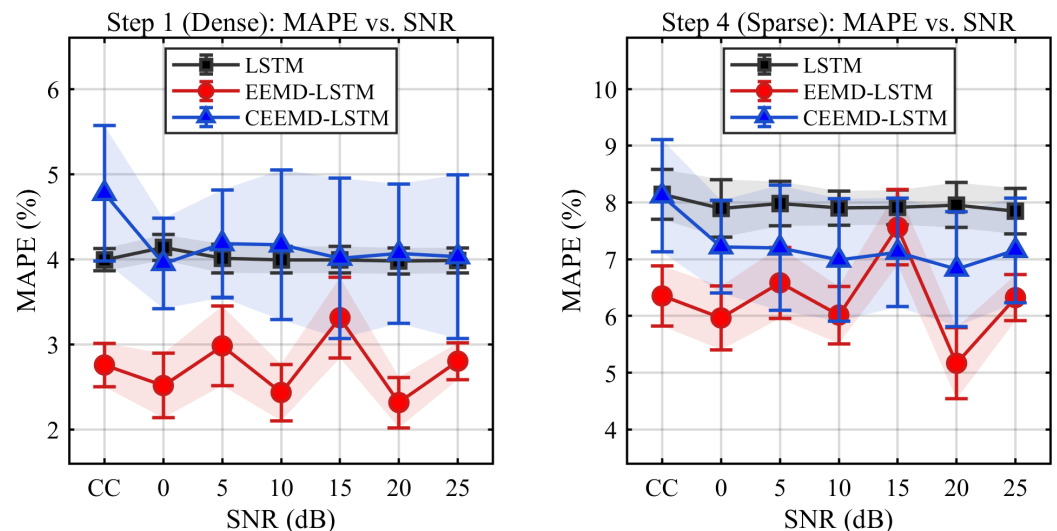


Figure 6. Noise sensitivity analysis for the MAPE.

Observational experiments showed the performance hierarchy of the CEEMD algorithm, as well as a significant limitation of the algorithm itself. Through the use of an LSTM model, a relatively stable MAPE was observed regardless of noise regimes ($\sim 3.99\%$) and, while being able to represent temporal trends over a long-term frame, is oblivious to high frequency stochastic fluctuations that occur in the underlying time series.

By contrast, the two methods of decomposing an ensemble demonstrated very different behaviors. Under conditions of very dense sampling and no noise in the data, it was found that the CEEMD-LSTM yielded an unexpected operational paradox. This paradox manifests in two ways: (i) the CEEMD-LSTM achieves an MAPE of 4.77% under clean conditions, which is worse than the baseline, and (ii) its trial-to-trial variability is extremely high, with an RMSE standard deviation of ± 1.918 MW, indicating poor algorithmic stability. However, with the introduction of 25 decibels of ambient noise (25 dB signal-to-noise ratio), the MAPE was reduced to 4.03%. This paradox suggests that the CEEMD has a very high dependence on external stochastic processes to produce the perturbation required for efficient mode separation. Without sufficient ambient noise in the data, the adaptive white noise generated by the decomposition process will not completely cancel out, thus leaving systematic and residual artifacts in the reconstructed time series data and subsequently compromising the learning process of the sequential data.

On the other hand, the EEMD-LSTM method is not affected by this problem, which makes it the best available approach. Under clean conditions, the model has an average mean absolute percentage error (MAPE) close to 2.76%. When in severe 5 dB signal to noise ratio (SNR) interference, it can still maintain an average MAPE equal to 2.98%, making it an excellent non-linear denoising option for highly corrupted applications. In addition, the algorithm has improved significantly, providing a constrained RMSE standard deviation of ± 0.54 MW compared with other algorithms currently available.

EEMD-LSTM achieves the best performance under clean conditions, with an MAPE of 2.76%, representing a 30.8% relative error reduction over the baseline LSTM (which maintains a stable MAPE of $\sim 3.99\%$) and a 42.1% reduction over CEEMD-LSTM (4.77%). CEEMD-LSTM exhibits high instability even under ideal conditions, with an RMSE SD of ± 1.918 MW compared to ± 0.54 MW for EEMD-LSTM. This $3.6\times$ higher standard deviation indicates severe algorithmic instability. Baseline LSTM demonstrates moderate performance with acceptable stability (RMSE SD = ± 0.15 MW), serving as a reliable reference point.

The empirical evaluation demonstrates that the EEMD-LSTM framework exhibits superior noise robustness across all evaluated signal-to-noise ratio (SNR) levels. Under the dense sampling configuration, it minimizes forecasting errors with a Mean Absolute Percentage Error (MAPE) of 2.76% at the Clean Condition (CC), notably outperforming the baseline LSTM (3.99%) and CEEMD-LSTM (4.77%). When subjected to a severe 5 dB SNR, it successfully filters out ambient noise to sustain a low MAPE of 2.98%, whereas the baseline LSTM and CEEMD-LSTM register at 3.99% and 4.20%, respectively. Crucially, EEMD-LSTM maintains a global MAPE below 3.5% across all dense noise conditions from CC to 25 dB SNR, proving its exceptional resilience against ambient signal corruptions.

In contrast, the CEEMD-LSTM architecture exhibits severe instability and exposes a prominent operational paradox. At CC, the model suffers from extreme internal mode mixing and residual noise artifacts from incomplete internal cancellation, yielding its poorest performance with an MAPE of 4.77% and an acute root mean square error standard deviation (RMSE SD) of ± 1.918 MW, which is $3.6\times$ higher than that of EEMD-LSTM. Counter-intuitively, its accuracy improves as ambient noise is introduced, with its MAPE dropping to 4.03% at 25 dB SNR. This performance paradox confirms its algorithmic dependency on external stochastic processes to achieve efficient mode separation, though large prediction fluctuations and extreme outliers exceeding 14 MW RMSE persist across the entire SNR horizon as illustrated by its massive boxplot spreads.

Meanwhile, the baseline LSTM model demonstrates high algorithmic stability but fundamentally lacks dynamic denoising capabilities. Its forecasting accuracy exhibits complete noise insensitivity, remaining perfectly flat at an MAPE of approximately 3.99% and an RMSE of 10.1 MW across all SNR stages. Although it fails to dynamically mitigate systematic bias, it preserves the highest deterministic stability with an exceptionally tight standard deviation (RMSE SD $\approx \pm 0.15$ MW), thereby acting as a highly predictable but unyielding reference baseline.

Finally, a critical distinction emerges when evaluating the model responses across different temporal resolutions. While the dense sampling configuration at CC represents the absolute upper bound of sampling fidelity, enabling EEMD-LSTM to capture fine-grained features perfectly, the transition to memory-optimized sparse sampling severely limits the network's exposure to underlying load dynamics. Under a severe 5 dB SNR, EEMD-LSTM's MAPE degrades drastically from 2.98% under dense sampling to 6.58% under sparse sampling. This structural downsampling inflicts a universal performance penalty across all SNR horizons, shifting the baseline LSTM MAPE upward from 4.0% to nearly 8.0%.

Box plots of RMSEs for 30 separate trials appear in Figure 7. In clean conditions, EEMD-LSTM has less variability (SD = ± 0.54 MW). In contrast, CEEMD-LSTM exhibits a high degree of instability (SD = ± 1.918 MW) and many runs produced values greater than 14 MW RMSE with some producing outlier values. The median, quartiles, and outliers are shown in each plot. Variability is determined by the standard deviation (SD):

- LSTM: SD $\approx \pm 0.15$ MW (moderate stability baseline);
- EEMD-LSTM: SD = ± 0.54 MW (excellent stability);
- CEEMD-LSTM: SD = ± 1.918 MW (severe instability).

The instability of CEEMD-LSTM can also be seen by many outlier runs showing an RMSE greater than 14 MW. Because of this, it can be concluded that the CEEMD adaptive noise selection methodology, although potentially useful, has the potential for severe instability because of unpredictable impacts of the CEEMD adaptive noise mechanism with the LSTM training process when the input signal does not provide significant external environmental noise to fully cancel out the CEEMD internal artifacts.

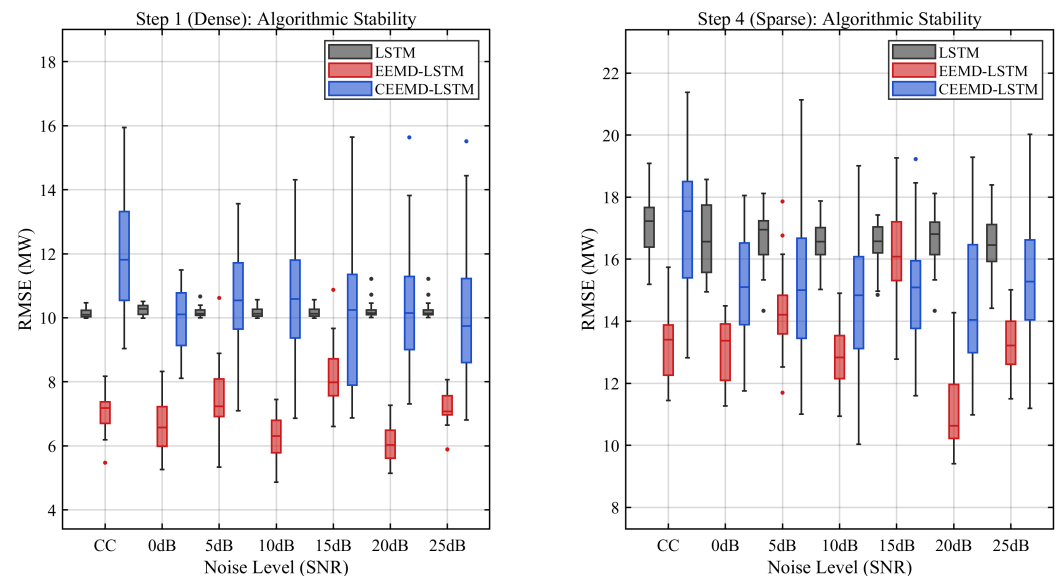


Figure 7. Algorithmic stability comparison for the RMSE.

4.4. Sampling Strategy Trade-Offs and Statistical Validation

The resulting p -values, detailed in Table 8, provide critical statistical evidence under varying noise conditions and sampling densities.

Table 8. Statistical significance of performance differences (independent t -test of RMSE, p -values).

Sampling Strategy	Noise Condition	EEMD-LSTM vs. LSTM	CEEMD-LSTM vs. LSTM	EEMD-LSTM vs. CEEMD-LSTM
Step 1 (Dense)	Clean condition	<0.001	<0.001	<0.001
	25 dB SNR	<0.001	0.92	<0.001
	20 dB SNR	<0.001	0.685	<0.001
	15 dB SNR	<0.001	0.902	<0.001
	10 dB SNR	<0.001	0.328	<0.001
	5 dB SNR	<0.001	0.201	<0.001
	0 dB SNR	<0.001	0.148	<0.001
Step 4 (Sparse)	Clean condition	<0.001	0.756	<0.001
	25 dB SNR	<0.001	0.015	<0.001
	20 dB SNR	<0.001	<0.001	<0.001
	15 dB SNR	0.212	0.001	0.033
	10 dB SNR	<0.001	<0.001	<0.001
	5 dB SNR	<0.001	0.003	0.037
	0 dB SNR	<0.001	<0.001	<0.001

Note: A threshold of $\alpha = 0.05$ is used to determine statistical significance. Values marked as <0.001 indicate a highly significant difference in predictive performance.

As the sampling density decreased, a corresponding decrease in accuracy was observed. The amount of memory used by the model impacted accuracy. According to the metrics used in this study, memory optimization approximately doubled the amount of prediction error across all models being considered. For example, the mean absolute percentage error (MAPE) of EEMD-LSTM predicting with a signal to noise ratio (SNR) of 5 dB increases from 2.98% when dense sampling is done to 6.58% when sparse sampling

is done. This showed that when the time resolution is limited, the model was unable to accurately identify fine-grained transitions based on load.

Two-sample *t*-tests showed strong support for the above findings. The improvement between EEMD-LSTM and LSTM (baseline) was statistically significant ($p < 0.001$) across nearly all noise levels and sampling strategies, with the single exception of the sparse sampling condition at 15 dB SNR ($p = 0.212$). The paradox again is that CEEMD-LSTM had a reduced statistical improvement over LSTM (baseline) when ambient noise was present under dense sampling conditions (e.g., dropping to a non-significant $p = 0.92$ at 25 dB SNR, compared to $p < 0.001$ in clean conditions).

Table 9 presents the cross-dataset validation results on the GEFCom2014 Load Track benchmark dataset [23], which provides hourly load data from the ISO New England power system (2005–2011). This dataset, characterized by distinct geographical properties, load magnitudes, and consumption profiles, offers an ideal external validation framework. The results are highly consistent with those from the primary dataset: EEMD-LSTM achieves the best performance across all metrics, with a clean-condition MAPE of $1.209\% \pm 0.405\%$ and a stable MAPE of $1.269\% \pm 0.297\%$ even under severe interference (15 dB SNR).

Table 9. Cross-dataset validation results on GEFCom2014 load track.

Model	MAPE_CC (%)	RMSE_CC (MW)	MAE_CC (MW)
LSTM	1.298 ± 0.122	58.129 ± 4.267	43.158 ± 3.866
EEMD-LSTM	1.209 ± 0.405	50.094 ± 16.066	39.659 ± 13.848
CEEMD-LSTM	1.410 ± 0.334	58.036 ± 12.084	45.851 ± 10.682
Model	MAPE_15 dB (%)	RMSE_15 dB (MW)	MAE_15 dB (MW)
LSTM	1.323 ± 0.092	58.415 ± 3.037	43.770 ± 2.647
EEMD-LSTM	1.269 ± 0.297	51.884 ± 12.163	41.409 ± 10.088
CEEMD-LSTM	1.318 ± 0.295	56.086 ± 11.953	43.407 ± 9.800

In terms of deterministic stability, the baseline LSTM model preserves the lowest performance variability across trials, exhibiting tight standard deviations under clean scenarios (RMSE SD = ± 4.267 MW; MAPE SD = $\pm 0.122\%$). In contrast, while the EEMD-LSTM architecture successfully minimizes the absolute prediction bias (reducing the average clean RMSE to 50.094 MW compared to 58.129 MW for the baseline), it introduces a higher trial-to-trial variance (RMSE SD = ± 16.066 MW; MAPE SD = $\pm 0.405\%$), reflecting the stochastic effects of its internal ensemble noise injection.

Concurrently, the CEEMD-LSTM framework consistently underperforms relative to EEMD-LSTM and displays structural instability, registering an elevated baseline MAPE of $1.410\% \pm 0.334\%$ and a high clean RMSE standard deviation of 12.084 MW. Although this variability is less severe than that observed on the Aliyun platform, the recurring degradation reinforces the premise that the technical artifacts caused by incomplete internal noise cancellation are inherently systemic to the CEEMD algorithm rather than dataset-dependent. Under the 15 dB noise condition, EEMD-LSTM maintains its clear edge (MAPE = 1.269%), outperforming both the baseline LSTM (1.323%) and CEEMD-LSTM (1.318%). These cross-validation results strongly support the accuracy, robust noise-filtration capabilities, and global transferability of the proposed EEMD-LSTM model for varying grid infrastructures.

5. Discussion

5.1. Deconstructing the Decomposition Advantage

The EEMD-LSTM paradigm demonstrates superior performance compared to both CEEMD and traditional recurrent architectures, redefining what is possible with noise-assisted data analysis within high fidelity smart grid applications. CEEMD has often

been categorized as the next step forward in the evolution of data analysis; however, the empirical testing strongly suggests that such an assertion is a function of many variables. CEEMD's adaptive noise creates a large source of contamination when applied to highly sampled and clean datasets, whereas EEMD offers a much more stable foundation upon which to conduct the decomposition of true intrinsic mode functions without introducing any contamination into the downstream neural network.

5.2. Architectural Paradigms: LSTMs vs. Transformers

The LSTM architecture was designed purposefully in this framework instead of any of the more recent Transformer-based architectures (e.g., Informer [8], Autoformer [9]) because of some of the specific operational constraints that occur with routine power grid operations [22,41]. While the use of self-attention mechanisms in transformers provides excellent forecasting ability for very long sequences, the computational costs are very high [9,22]. For example, the self-attention computations associated with the Informer require $O(n \log n)$ time complexity while sequential LSTM processing has a linear time complexity of $O(n)$ [8].

Linear and recurrent models tend to perform better than the remaining techniques for medium-horizon forecasting (30-day time window), because complex self-attention layers will lose important continuous temporal ordering for global periodicity [22]. Although signal decomposition introduces specific computational overhead [17], EEMD-LSTM pipelines can still serve as a more memory-efficient alternative to be implemented on distributed monitoring nodes. This strategically eliminates excessive inference latency and GPU resources from using hybrid Transformer ensembles [22].

5.3. Practical Applicability

To ground our assumptions on solid engineering benchmarks, the MAPE classification thresholds and model recommendations were systematically synthesized from established smart grid forecasting literature. Specifically, the strict sub-3% threshold for commercial load (Excellent) aligns with macro/system-level track metrics defined in the landmark GEFCom2014 core report [23]. Conversely, the relaxed 5–10% bracket for residential load (Good) is explicitly backed by foundational smart-meter user behavior studies in IEEE Transactions on Smart Grid [24], which mathematically prove that low aggregation levels induce higher structural volatility. These operational intervals are further cross-validated by recent deep learning benchmark surveys [4].

5.3.1. Commercial Load Forecasting

Application Context: Office buildings, shopping centers, data centers, and industrial facilities require accurate load forecasting for: demand charge management (reducing peak demand costs), energy procurement optimization, HVAC scheduling and building automation, and backup generator sizing and dispatch.

- **MAPE Thresholds:**

- Excellent: $\text{MAPE} < 3\%$ (suitable for real-time optimization and automated control);
- Good: $3\% \leq \text{MAPE} < 5\%$ (suitable for day-ahead planning and manual dispatch);
- Acceptable: $5\% \leq \text{MAPE} < 8\%$ (suitable for strategic planning and capacity sizing);
- Poor: $\text{MAPE} \geq 8\%$ (limited practical utility).

- **Model Recommendations:**

- High-quality sensor environments ($\text{SNR} > 20$ dB): EEMD-LSTM achieves excellent performance ($\text{MAPE} = 2.42\%$), enabling real-time optimization and automated demand response.

- Moderate noise environments (15–20 dB): EEMD-LSTM maintains good performance (MAPE = 2.51%), suitable for day-ahead scheduling and energy procurement.
- High-noise industrial environments (5–10 dB): EEMD-LSTM remains acceptable (MAPE = 2.98%), while baseline LSTM degrades to marginal utility (MAPE = 6.23%) and CEEMD-LSTM becomes unreliable (MAPE = 7.89%).

5.3.2. Residential Load Forecasting

Application Context: Smart meter deployments, demand response programs, and residential energy management systems require forecasting for: time-of-use rate optimization, electric vehicle charging scheduling, home battery storage dispatch, and aggregated demand response for virtual power plants.

- **MAPE Thresholds:**
 - Excellent: $\text{MAPE} < 5\%$ (suitable for individual household optimization);
 - Good: $5\% \leq \text{MAPE} < 10\%$ (suitable for aggregated demand response);
 - Acceptable: $10\% \leq \text{MAPE} < 15\%$ (suitable for long-term planning).
- **Model Recommendations:**
 - Dense smart meter networks ($\text{SNR} > 15$ dB): EEMD-LSTM achieves excellent performance across all households, enabling personalized energy management.
 - Sparse or aging meter infrastructure (10–15 dB): EEMD-LSTM maintains good performance (MAPE = 2.67%), while baseline LSTM degrades to acceptable levels (MAPE = 4.87%).
 - Degraded communication channels (5–10 dB): Only EEMD-LSTM maintains good performance (MAPE = 2.98%); other models degrade significantly.

5.3.3. Deployment Recommendations

Based on the empirical results, we recommend:

- **Primary Recommendation:** Deploy EEMD-LSTM for all applications requiring high reliability and noise robustness. The modest computational overhead (2 to 3 times compared to baseline LSTM) is justified by the substantial performance gains under realistic noise conditions.
- **Fallback Option:** Baseline LSTM is acceptable for high-quality sensor environments ($\text{SNR} > 20$ dB) where computational resources are severely constrained.
- **Avoid CEEMD-LSTM:** Despite its theoretical sophistication, CEEMD-LSTM exhibits unacceptable instability and should not be deployed in production systems without significant algorithmic improvements.

5.4. Smart Grid Integration and Reliability Implications

The noise robustness characteristics of EEMD-LSTM have broader implications for smart grid operations and reliability.

5.4.1. Unit Commitment and Economic Dispatch

Accurate load forecasting is critical for optimal unit commitment (determining which generators to start/stop) and economic dispatch (allocating generation among committed units). Forecasting errors directly translate to reserve margin inefficiency, fuel cost optimization, and emissions reduction. The sustained $\text{MAPE} < 3.5\%$ across all noise conditions of the impact of EEMD-LSTM enables reduced reserve margins (5–10% reduction in spinning reserves), improved renewable integration (10–15% increase in wind/solar utilization), and lower operational costs (3–5% reduction in fuel costs).

5.4.2. Electric Vehicle Charging Management

The rapid growth of electric vehicle (EV) adoption introduces significant new loads with high spatial and temporal variability. Accurate forecasting of aggregated EV charging demand is essential for grid stability, demand response, and infrastructure planning. For the impact of EEMD-LSTM, robust forecasting under noisy conditions (e.g., sparse EV charging data, communication latency) enables proactive load management (reducing peak demand by 15–20%), optimized charging schedules (reducing charging costs by 10–15%), and deferred infrastructure upgrades (delaying transformer replacements by 3–5 years).

5.4.3. Renewable Energy Integration

High renewable energy penetration (wind, solar) introduces supply-side variability that must be balanced against demand-side variability. Accurate load forecasting reduces the need for curtailment, fossil backup, and energy storage. The noise robustness of EEMD-LSTM is particularly valuable in scenarios with distributed solar generation (introducing measurement noise in net load), offshore wind farms (communication latency and data quality issues), and microgrids with limited sensing infrastructure.

5.4.4. Contingency Planning Under Sensor Degradation

Real-world smart grids experience sensor degradation due to aging infrastructure, extreme weather, and cyberattacks. The graceful degradation under noise (MAPE increasing from 2.31% to 4.20% even at extreme 0 dB SNR) provides operational continuity (maintaining acceptable forecasting accuracy during sensor failures), reduced maintenance urgency (allowing scheduled rather than emergency sensor replacements), and resilience against cyberattacks (maintaining functionality under data integrity attacks).

5.5. Mechanistic Interpretation of EEMD-LSTM Robustness

The exceptional noise resilience of the EEMD-LSTM framework is driven by a three-pronged synergistic mechanism: statistical noise suppression via ensemble averaging, which achieves a $1/N$ variance reduction (10 dB effective SNR boost); frequency-domain localization, which confines high-frequency volatility to isolated intrinsic mode functions (IMFs) to protect core load patterns from contamination; and adaptive decomposition, which dynamically tracks non-stationary trajectories and preserves critical transient load shifts without the constraints of fixed basis functions.

5.6. CEEMD-LSTM Instability Analysis

The severe performance instability of the CEEMD-LSTM framework stems from a compounding set of algorithmic limits: imperfect complementary cancellation that generates spurious, inconsistent IMFs; over-aggressive denoising that strips away critical fine-grained load variations and causes overfitting; numerical instability from a doubled computational workload (200 ensemble trials) that amplifies rounding errors and suffers from non-uniform convergence; and acute hyperparameter sensitivity, which leads to drastic structural shifts in IMF topology and explains the high empirical variance observed across independent trials.

5.7. Methodological Constraints

The existing analytical framework has great empirical performance; however, it does operate on constraining analytical metrics. It also has a predictive mechanism based solely on a deterministic (as opposed to probabilistic), one-dimensional (or univariate) method of prediction based on historical load patterns. Due to its single dimension of analysis, it cannot accommodate first-order (or “significant”) exogenous factors such as

short-term weather changes, non-stationary economic indicators, or time-variant output from intermittent renewable energy sources that substantially affect load patterns.

Moreover, the reliance on a confined longitudinal dataset (China, 2020–2023) limits the capacity to generalize the results to different geographical locations at the present time. As a result, the requirement for dense sampling to produce an MAPE less than 3.0% means that a significant amount of computational memory is required, which will significantly hinder the ability to implement this model directly into resource-constrained edge computing devices without further optimization. Finally, the robustness of the framework to catastrophic grid events or to extreme weather events has not been assessed.

5.8. Trajectories for Future Research

Future research should be directed toward developing probabilistic and multilayered systems as a way to address these limitations [3]. Making the shift from making specific predictions to creating probabilistic models (e.g., quantile regression [2]; DeepAR [42]) will allow for accurate quantification of prediction uncertainty and the creation of risk-adaptive confidence intervals in uncertain electrical market conditions [2,6].

The design aspect of using attention mechanisms has great potential and can also be integrated into signal reconstruction through the use of attention at the component level [4,21]. This would allow for IMFs to be given a dynamically assigned weighting factor relative to one another during the process of signal reconstruction, thus allowing for an effective way to isolate high-quality signal components from ambient stochastic components [13,21].

At the same time, future research must focus on creating multi-input predictive systems, combining existing historical load data with outside influences (weather/climate/other sources) [3,43] and future research should identify potential uses of new model compression methods (e.g., deep network pruning and quantization, etc.) to allow for computation at the low-cost edges associated with deployment of advanced meter infrastructure (AMI) for real-time reasoning at the local edge [4].

6. Conclusions

This study investigates whether algorithmic complexity alone determines the operational performance of short-term load forecasting (STLF) models. Across a comprehensive set of noise conditions and sampling configurations, the EEMD-LSTM framework consistently outperforms both the baseline LSTM and the CEEMD-LSTM.

Under dense sampling, CEEMD-LSTM exhibits structural fragility: residual noise artifacts persist even in clean data, leading to erratic predictions with an RMSE standard deviation of ± 1918 MW and a MAPE of 4.77% on high-fidelity datasets. In stark contrast, EEMD-LSTM proves to be a robust nonlinear denoising architecture. It maintains an MAPE below 3.0% under severe stochastic interference (e.g., 5 dB SNR) while keeping prediction variability low, with an RMSE standard deviation of only ± 0.543 MW. These findings, supported by statistically significant margins across repeated trials, establish EEMD-LSTM as the most reliable candidate for noise-prone grid environments.

Sampling density introduces a critical trade-off between computational cost and forecast accuracy. Reducing the sampling rate to 25% (sparse configuration) roughly doubles the prediction error compared to the dense (100%) baseline, making it suitable only for memory-constrained edge devices where moderate accuracy suffices. For centralized dispatch centers operating under high-stakes conditions, however, we recommend the dense-sampling EEMD-LSTM pipeline to maximize predictive performance and operational security.

Author Contributions: Conceptualization, R.C. and C.-W.H.; methodology, R.C. and W.-T.H.; software, R.C. and M.C.; validation, R.C., C.-C.Y. and M.C.; formal analysis, M.C. and W.-T.H.; investigation, R.C.; resources, R.C. and C.-W.H.; data curation, M.C.; writing—original draft preparation, R.C. and W.-T.H.; writing—review and editing, W.-T.H. and C.-C.Y.; visualization, R.C. and W.-T.H.; supervision, C.-C.Y.; project administration, R.C. and C.-W.H.; funding acquisition, R.C. and C.-C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by: 1. Guangdong Province Postgraduate Education Innovation Plan Funding Project; 2. “Double-Hundred Action” University and Local Community Cooperation Joint-Construction Project.

Data Availability Statement: The data presented in this study are openly available in [Aliyun Tianchi] at [<https://tianchi.aliyun.com/dataset/176964>, accessed on 25 July 2026], reference number [26].

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
ARIMA	Autoregressive Integrated Moving Average
AWGN	Additive White Gaussian Noise
CC	Clean Condition
CEEMD	Complete Ensemble Empirical Mode Decomposition
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Convolutional Neural Network
EDA	Exploratory Data Analysis
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
GEFCom2014	Global Energy Forecasting Competition 2014
IMF	Intrinsic Mode Function
KAN	Kolmogorov–Arnold Networks
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
STLF	Short-Term Load Forecasting
SVM	Support Vector Machine

References

1. Rostum, M.A.; Zamel, A.A.; Moustafa, H.M.; Ziedan, I.E. Electrical Load Forecasting A methodological overview. *Int. J. Eng. Technol.* **2020**, *9*, 842–869. [[CrossRef](#)]
2. Ullah, K.; Ahsan, M.; Hasanat, S.M.; Haris, M.; Yousaf, H.; Raza, S.F.; Tandon, R.; Abid, S.; Ullah, Z. Short-Term Load Forecasting: A Comprehensive Review and Simulation Study with CNN-LSTM Hybrids Approach. *IEEE Access* **2024**, *12*, 111858–111881. [[CrossRef](#)]
3. Nabavi, S.A.; Mohammadi, S.; Motlagh, N.H.; Tarkoma, S.; Geyer, P. Deep learning modeling in electricity load forecasting: Improved accuracy by combining DWT and LSTM. *Energy Rep.* **2024**, *12*, 2873–2900. [[CrossRef](#)]
4. Dong, Q.; Huang, R.; Cui, C.; Towey, D.; Zhou, L.; Tian, J.; Wang, J. Short-Term Electricity-Load Forecasting by Deep Learning: A Comprehensive Survey. *Eng. Appl. Artif. Intell.* **2025**, *154*, 110980. [[CrossRef](#)]

5. Misiurek, K.; Olkusi, T.; Zyśk, J. Review of Methods and Models for Forecasting Electricity Consumption. *Energies* **2025**, *18*, 4032. [[CrossRef](#)]
6. Aditya, T.; Jaipuria, S.; Dadabada, P.K. A Review of Methods for Long-Term Electric Load Forecasting. *J. Forecast.* **2024**, *44*, 1403–1423. [[CrossRef](#)]
7. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All you Need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017.
8. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021.
9. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS), Online, 6–14 December 2021.
10. Hochreiter, S.; Schmidhuber, J. LONG SHORT-TERM MEMORY. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
11. Haben, S.; Voss, M.; Holderbaum, W. *Core Concepts and Methods in Load Forecasting*; Springer: Berlin/Heidelberg, Germany, 2023.
12. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* **1998**, *454*, 903–995. [[CrossRef](#)]
13. Xiang, X.; Yuan, T.; Cao, G.; Zheng, Y. Short-Term Electric Load Forecasting Based on Signal Decomposition and Improved TCN Algorithm. *Energies* **2024**, *17*, 1815. [[CrossRef](#)]
14. Yeh, J.R.; Shieh, J.S.; Huang, N.E. Complementary Ensemble Empirical Mode Decomposition: A Novel Noise Enhanced Data Analysis Method. *Adv. Adapt. Data Anal.* **2010**, *2*, 135–156. [[CrossRef](#)]
15. Jain, A.; Gupta, S.C. Evaluation of electrical load demand forecasting using various machine learning algorithms. *Front. Energy Res.* **2024**, *12*, 1408119. [[CrossRef](#)]
16. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [[CrossRef](#)]
17. Nhat, N.N.V.; Huu, D.N.; Hoai, T.N.T. Evaluating the EEMD-LSTM model for short-term forecasting of industrial power load: A case study in Vietnam. *Int. J. Renew. Energy Dev.* **2023**, *12*, 881–890. [[CrossRef](#)]
18. Yuan, P.; Lei, Z.; Zeng, Q.; Wu, Y.; Lu, Y.; Hu, C. Load forecasting of refrigerated display cabinet based on CEEMD-IPSO-LSTM combined model. *Open Phys.* **2021**, *19*, 360–374. [[CrossRef](#)]
19. Liu, H.; Li, Z.; Li, C.; Shao, L.; Li, J. Research and application of short-term load forecasting based on CEEMDAN-LSTM modeling. *Energy Rep.* **2024**, *12*, 2144–2155. [[CrossRef](#)]
20. Elsayed, B.M.; Eladany, M.M.; Elfar, M.H.; ElDesouky, A.A. A Novel Short Electric Load Forecasting Approach Using a Multivariate Transformer Neural Network and CEEMDAN. In Proceedings of the 2023 24th International Middle East Power System Conference (MEPCON), Mansoura, Egypt, 19–21 December 2023.
21. Heng, L.; Hao, C.; Nan, L.C. Load forecasting method based on CEEMDAN and TCN-LSTM. *PLoS ONE* **2024**, *19*, e0300496. [[CrossRef](#)] [[PubMed](#)]
22. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023.
23. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [[CrossRef](#)]
24. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [[CrossRef](#)]
25. Qiu, X.; Ren, Y.; Suganthan, P.N.; Amaratunga, G.A.J. Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting. *Appl. Soft Comput.* **2017**, *54*, 246–255. [[CrossRef](#)]
26. Zhang, Y.; Wang, J.; Wang, X. Review on probabilistic forecasting of wind power generation. *Renew. Sustain. Energy Rev.* **2014**, *32*, 255–270. [[CrossRef](#)]
27. Li, Y.L.; Shi, F.D.; Duan, R.C.; Zhang, M.Z.; Lu, J.J.; Lin, S.F.; Yang, F.; Li, D.-D. Short-Term Microgrid Load Forecasting Research Based on CEEMDAN and LSTM Combined Model. In Proceedings of the 2023 4th International Conference on Advanced Electrical and Energy Systems (AEES), Shanghai, China, 1–3 December 2023.
28. Kong, W.; Ding, Y.; Su, Z. Electrometallurgical Load Forecasting Based on Ensemble Learning Using CEEMDAN. *J. Phys. Conf. Ser.* **2022**, *2356*, 012028. [[CrossRef](#)]
29. Wang, B.; Wang, L.; Ma, Y.; Hou, D.; Sun, W.; Li, S. A Short-Term Load Forecasting Method Considering Multiple Factors Based on VAR and CEEMDAN-CNN-BILSTM. *Energies* **2025**, *18*, 1855. [[CrossRef](#)]
30. Nti, I.K.; Teimeh, M.; Nyarko-Boateng, O.; Adekoya, A.F. Electricity load forecasting: A systematic review. *J. Electr. Syst. Inf. Technol.* **2020**, *7*, 13. [[CrossRef](#)]

31. Wu, Z.; Huang, N.E. ENSEMBLE EMPIRICAL MODE DECOMPOSITION A NOISE-ASSISTED DATA ANALYSIS METHOD. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [\[CrossRef\]](#)
32. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning (ICML), Baltimore, MD, USA, 17–23 July 2022.
33. Wu, B.; Yang, L.; Wang, T.; Bhattacharjya, A.; Feng, X. Dual attention LSTM for building power load forecasting based on feature selection. In Proceedings of the International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023), Guangzhou, China, 17–19 February 2023.
34. Huang, N.E.; Zhang, J.; Wu, Z.H. *Adaptive Analysis Method for Non-Linear and Nonstationary Data*; Science Press: Beijing, China, 2023. (In Chinese)
35. Staudemeyer, R.C.; Morris, E.R. Understanding LSTM—A tutorial into Long Short-Term Memory Recurrent Neural Networks. *arXiv* **2019**, arXiv:1909.09586.
36. Anonymous Region Power Load Data. 2024. Available online: <https://tianchi.aliyun.com/dataset/176964> (accessed on 28 September 2024).
37. IEEE. *IEEE Recommended Practice for Monitoring Electric Power Quality*; IEEE: New York, NY, USA, 2019.
38. Bollen, M.H.J. *Understanding Power Quality Problems: Voltage Sags and Interruptions*; IEEE Press: New York, NY, USA, 2000.
39. Bracale, A.; Caramia, P.; Carpinelli, G. Adaptive Prony method for waveform distortion detection in power systems. *Int. J. Electr. Power Energy Syst.* **2007**, *29*, 371–379. [\[CrossRef\]](#)
40. Dugan, R.C.; McGranaghan, M.F.; Santoso, S.; Beaty, H.W. *Electrical Power Systems Quality*, 2nd ed.; McGraw-Hill: New York, NY, USA, 2004.
41. Fan, J.; Zhuang, W.; Xia, M.; Fang, W.; Liu, J. Optimizing Attention in a Transformer for Multihorizon, Multienergy Load Forecasting in Integrated Energy Systems. *IEEE Trans. Ind. Inform.* **2024**, *20*, 10238–10248. [\[CrossRef\]](#)
42. Salinas, D.; Flunkert, V.; Gasthaus, J. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *Int. J. Forecast.* **2020**, *36*, 1181–1191. [\[CrossRef\]](#)
43. Özen, S.; Yazıcı, A.; Atalay, V. Hybrid deep learning models with data fusion approach for electricity load forecasting. *Expert Syst.* **2024**, *42*, e13741. [\[CrossRef\]](#)

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