Energy Reports 16 (2026) 109462

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

An interpretable and uncertainty-aware hybrid framework for short-term load forecasting in modern energy systems

Musa Terkes [a], Alpaslan Demirci [a], Idriss Dagal [b], Erdin Gokalp [a], Umit Cali [c,d,e,*]

[a] Department of Electrical Engineering, Yildiz Technical University, İstanbul, Türkiye
[b] Department of Electrical-Electronics Engineering, Istanbul Beykent University, Istanbul 34396, Türkiye
[c] Department of Electric Energy, Norwegian University of Science and Technology, Trondheim, Norway
[d] School of Physics, Engineering and Technology, University of York, York, United Kingdom
[e] American University of Sharjah, Artificial Intelligence, Smart Infrastructure and Robotics Research Center, Sharjah, United Arab Emirates

ARTICLE INFO

Keywords:
Short-term load forecasting
Hybrid deep learning
Stacking ensemble
Uncertainty quantification
Explainable AI

ABSTRACT

Accurate short-term load forecasting (STLF) is critical for the reliable operation of modern power systems, the integration of renewable energy resources, and informed participation in electricity markets. However, the increasing variability of residential, prosumer, and electric vehicle (EV) demand profiles introduces significant challenges related to non-stationarity, predictive uncertainty, and the limited interpretability of existing forecasting models. To address these challenges, this study proposes an interpretable and uncertainty-aware hybrid forecasting framework that integrates signal decomposition, deep temporal learning, and ensemble modeling within a unified decision-support pipeline. Load signals are decomposed using complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) to extract multi-scale temporal components, while a hybrid long short-term memory (LSTM)–Transformer architecture captures both short-term dynamics and long-range temporal dependencies. In parallel, a stacking ensemble combining gradient boosting, multilayer perceptron (MLP), and Ridge regression is employed to enhance robustness and generalization across heterogeneous load profiles. Predictive uncertainty is explicitly quantified through Monte Carlo (MC) Dropout, and model transparency is ensured via SHAP-based feature attribution. The proposed framework is evaluated on four real-world datasets encompassing residential consumption, prosumer demand with photovoltaic generation, EV charging, and heterogeneous household load profiles. Experimental results demonstrate that the framework achieves high forecasting accuracy while simultaneously providing well-calibrated uncertainty estimates and interpretable insights. Notably, the stacking ensemble attains root mean squared error (RMSE) and mean absolute error (MAE) values as low as 0.0359 and 0.0209, respectively, confirming improved forecasting performance under diverse and non-stationary conditions. The probabilistic forecasts are well-calibrated, with prediction interval coverage probability (PICP) values exceeding 95%, and the model maintains consistent performance across multiple load profiles.

1. Introduction

1.1. Background and motivation

Electric load forecasting plays a strategically critical role in ensuring the safe, economical, and sustainable operation of energy systems. Especially with the accelerating transition to smart grids, the integration of renewable energy sources, and the liberalization of energy markets, load forecasting has become an increasingly complex, multidimensional, and dynamic problem (Lee and Cho, 2022; Massaoudi et al., 2021; Sideratos et al., 2020). Load forecasting models act as decision support mechanisms in key processes such as supply-demand balancing, reserve capacity planning, pricing, and demand-side management. In recent years, the rising electricity consumption in residential (Alhussein et al., 2020; Sajjad et al., 2020; Syed et al., 2021), industrial (Wang et al., 2021), and commercial sectors has rendered consumption patterns more irregular and difficult to predict, introducing complex patterns such as temporal variability, non-linearity, non-stationarity, sudden

* Corresponding author at: Department of Electric Energy, Norwegian University of Science and Technology, Trondheim, Norway.
E-mail addresses: musa.terkes@yildiz.edu.tr (M. Terkes), ademirci@yildiz.edu.tr (A. Demirci), idrissdagal@beykent.edu.tr (I. Dagal), gokalp@yildiz.edu.tr (E. Gokalp), umit.cali@ntnu.no (U. Cali).

https://doi.org/10.1016/j.egyr.2026.109462
Received 27 January 2026; Received in revised form 16 April 2026; Accepted 20 June 2026
2352-4847/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

spikes, and time-varying dependencies (Bashir et al., 2022; Bouktif et al., 2020; Dai and Zhao, 2020; Dudek et al., 2022; Mounir et al., 2023; Shohan et al., 2022). Especially in decentralized and localized energy systems such as microgrids, prosumer systems, and building energy management systems (BEMS), short-term load forecasting (STLF) at minute-level resolution has become a critical requirement in terms of operational efficiency, supply security, and demand flexibility (Faraji et al., 2020; Moradzadeh et al., 2020; Tayab et al., 2020). Hybrid models such as convolutional neural network (CNN)-long short-term memory (LSTM) (Alhussein et al., 2020), gated recurrent unit (GRU)-CNN (Wu et al., 2020), and Transformer-based architectures (Li et al., 2023; Ran et al., 2023) have proven effective in capturing the spatiotemporal dependencies in load time series. Additionally, ensemble learning strategies (stacking, boosting) combine the strengths of different learners, thereby enhancing model robustness and generalization capabilities (Khan et al., 2020; Massaoudi et al., 2021; Xuan et al., 2021).

Time-frequency decomposition techniques such as empirical mode decomposition (EMD), variational mode decomposition (VMD), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and wavelet transforms (WTs) (Li et al., 2023; Lv et al., 2022; Mounir et al., 2023) enable the extraction of more homogeneous and interpretable subcomponents from load signals, thereby facilitating pattern recognition and improving forecast accuracy. Heuristic algorithms used for hyperparameter optimization and feature selection (genetic algorithm (GA), Cuckoo Search, Harris Hawks optimization (HHO)) (Bouktif et al., 2020; Hafeez et al., 2020; Nie et al., 2020) provide more flexible and effective solutions compared to traditional methods, thereby boosting model performance. Attention mechanisms and explainable artificial intelligence (XAI) techniques improve transparency in deep learning (DL) models' internal decision-making processes and increase user trust (Chung et al., 2022; Shohan et al., 2022; Wan et al., 2023; C. Zhang et al., 2020).

1.2. Existing approaches and limitations

The most used methods for STLF include traditional machine learning (ML) algorithms (support vector machine (SVM), random forest (RF), k-nearest neighbors (k-NN)) and DL-based models (LSTM, GRU, CNN, Transformer). However, both categories suffer from notable limitations. Traditional ML algorithms are unable to effectively capture sequential dependencies in time series data, resulting in limited sensitivity and accuracy when handling highly volatile and complex load patterns (Hafeez et al., 2020; Moradzadeh et al., 2020; Tarmanini et al., 2023). Additionally, their computational cost increases with high-dimensional data, along with the risk of overfitting (Ahmad et al., 2020; Mounir et al., 2023). On the other hand, recurrent neural networks (RNN) such as LSTM and GRU can model temporal dependencies, yet due to their black-box nature, optimization challenges, and high sensitivity to hyperparameters, their applicability in real-world scenarios remains limited (Hafeez et al., 2020; Khan et al., 2021; C. Zhang et al., 2020). Similarly, CNN-based models are proficient in capturing local patterns but often fall short in modeling long-term dependencies (Khan et al., 2020; Rafi et al., 2021). Signal decomposition methods such as CEEMDAN, EMD, or VMD are often only used for noise reduction, while ignoring component-wise modelling, reconstruction, or their interdependencies (Hafeez et al., 2020; Li et al., 2023; Mounir et al., 2023; Ran et al., 2023). Furthermore, the integration of lag features, exogenous variables (weather conditions, holidays, social events), or physical system features into models is frequently inadequate (Li et al., 2023; Moradzadeh et al., 2020; Song et al., 2021). Although meta-heuristic algorithms offer advantages in hyperparameter tuning, they also introduce challenges such as premature convergence and high computational complexity, which can negatively affect model accuracy and training efficiency (Bouktif et al., 2020; Tayab et al., 2020; Z. Zhang et al., 2020). Additionally, many models are validated using only a single time frame, failing to assess performance across weekly, seasonal, or structural contexts (Chung et al., 2022; Ran et al., 2023). Recent studies highlight that incorporating industry-specific load characteristics through hybrid approaches combining classification indices with advanced learning architectures can significantly improve regional forecasting accuracy (Zhou et al., 2024), yet existing ML and DL models still struggle to fully capture the complex structure of time series data required for modern STLF.

1.3. Recent advances and research gap

Recent advances in STLF increasingly rely on hybrid AI frameworks to cope with the nonlinear, non-stationary, and multi-scale nature of electricity demand. Despite these methodological advances, current STLF literature remains largely fragmented and accuracy-centric, with several critical limitations. First, most hybrid forecasting approaches treat signal decomposition, temporal modelling, ensemble learning, uncertainty estimation, and explainability as independent methodological add-ons rather than as components of a unified forecasting pipeline. Second, uncertainty quantification remains weakly integrated into DL-based STLF, particularly for highly volatile demand profiles such as residential consumption, prosumer systems with photovoltaic (PV) generation, and electric vehicle (EV) charging loads. Third, while explainability has gained increasing attention, its joint consideration with predictive uncertainty remains largely unexplored. Finally, a systematic lack of cross-profile validation persists in the literature, as the majority of STLF studies focus on single or homogeneous datasets, thereby limiting the assessed generalizability of proposed models. Taken together, these limitations reveal the lack of a unified STLF framework that simultaneously and explicitly addresses non-stationarity, long-range temporal dependency, uncertainty quantification, explainability, and generalization across heterogeneous load profiles. In response, this study proposes a comprehensive hybrid framework that combines CEEMDAN-based signal decomposition with a fused LSTM–Transformer temporal learning structure, embeds predictive uncertainty through MC Dropout, and ensures transparency via SHAP-based explainability. By further enhancing robustness through a stacking ensemble that integrates gradient boosting regressor (GBR), Ridge regression, and multilayer perceptron (MLP) models, the proposed framework delivers a system-level, uncertainty-aware, and interpretable solution for STLF. Unlike existing hybrid approaches that loosely combine decomposition, DL, and ensemble techniques, the proposed framework introduces a structurally integrated pipeline in which these components are jointly optimized and functionally interdependent. This distinguishes the proposed approach from prior studies where such components are typically applied in isolation or as sequential add-ons.

1.4. Contributions and paper organization

The original contributions of this study are summarized as follows:

- A unified and structurally integrated STLF framework is proposed, in which signal decomposition, deep temporal learning, ensemble generalization, uncertainty quantification, and explainability are co-designed as interdependent components of a single forecasting architecture.
- Predictive uncertainty is explicitly embedded into a DL–based STLF framework through MC Dropout, enabling probabilistic load forecasts for highly volatile demand profiles and supporting risk-aware operational and market-level decision-making.
- Explainability is jointly integrated with uncertainty-aware modeling, allowing feature relevance and lag effects to be interpreted alongside predictive confidence, thereby enhancing the practical trustworthiness of AI-based load forecasting systems beyond post hoc point-forecast explanations.

- A systematic cross-profile generalization analysis is conducted across four heterogeneous real-world load profiles, addressing a key limitation of prior studies that focus on single or homogeneous datasets.
- Comprehensive comparative evaluations demonstrate that the proposed unified framework consistently outperforms state-of-the-art benchmark models in terms of forecasting accuracy and robustness across diverse consumption patterns.
- Practical deployment aspects are explicitly discussed, including computational complexity, scalability, and integration feasibility for real-time smart grid forecasting.

The remainder of this paper is organized as follows. Section 2 presents the datasets, preprocessing procedures, and the proposed methodology. Section 3 reports the results, while Section 4 provides the corresponding discussion. Section 5 concludes the paper and outlines future research directions.

2. Employed methods and model framework

2.1. Data and preprocessing

2.1.1. Dataset description

This study employs four distinct real-world load profiles to comprehensively evaluate the proposed hybrid AI model's performance in diverse operational contexts. The datasets represent varying consumer types, temporal resolutions, and geographical regions, ensuring broad applicability and robustness of the forecasting framework.

- **Profile 1:** Minute-level residential electricity consumption data collected from a typical urban household in Sceaux, France (December 2006 – November 2010). This dataset captures detailed active power, voltage, and sub-metering measurements, reflecting typical residential load variability. Missing values account for approximately 1.25% (Hebrail and Berard, 2006).
- **Profile 2:** Fifteen-minute resolution data from prosumers in Portugal, characterized by integrated PV generation and consumption patterns from 15 consumers throughout 2019 (Ramos et al., 2021). Seasonal variability, especially pronounced during summer months, adds complexity due to increased solar generation and consumption shifts.
- **Profile 3:** EV charging load profiles collected in Denmark, encompassing multi-phase charging behaviors and structured around day types (weekday vs. weekend), covering 3500 + households with hourly resolution for one year (Weckesser et al., 2024).
- **Profile 4:** High-resolution individual household consumption data from 346 homes in the United States (Pecan Street project, January – April 2016) (Zheng and Inc., 2016). This dataset features pronounced heterogeneity due to personalized consumption habits, voltage fluctuations, and diverse appliance usage patterns.

Each dataset underwent preprocessing steps including data cleaning, handling missing values, and synchronization of timestamps. Signal decomposition using CEEMDAN was applied to extract intrinsic mode functions (IMFs), improving the interpretability and modeling of non-stationary and nonlinear components. Table 1 summarizes the key attributes of each dataset, including time resolution, number of samples, and main features.

Table 1
Electric load profiles and dataset details used in this study.

Load	Time res.	Sample size	Main features	Ref.
Profile 1	1 min.	2075,259	Residential consumption (France). Parameters: global_active_power, global_reactive_power, voltage, global_intensity, three sub-meterings for kitchen, laundry, and heating/cooling loads. Missing values ~1.25%.	(Hebrail and Berard, 2006)
Profile 2	15 min.	35,040	Prosumer data (Portugal): Includes PV production (3 producers), consumption of 15 users, common services, and total load.	(Ramos et al., 2021)
Profile 3	1 h	~8760	EV charging profiles (Denmark): Multi-phase charging (1, 2, 3 phases), various charger types and percentile aggregation levels. Covers weekdays, weekends, summer/winter.	(Weckesser et al., 2024)
Profile 4	1 min.	~4months	Household consumption (USA, Pecan Street): 346 homes, kWh, voltage, high granularity.	(Zheng and Inc., 2016)

2.1.2. Data cleaning and missing value handling

The accuracy and robustness of hybrid AI models for STLF heavily depend on meticulous data preparation. In this study, four different load datasets with varying temporal and user granularity were utilized, requiring tailored preprocessing strategies to ensure data integrity and model reliability. Raw data exhibited typical real-world challenges, such as missing timestamps, duplicated records, irregular sampling intervals, and NaN values. For datasets with timestamped entries (Profiles 3 and 4), all time columns were first converted to a standardized datetime format, with invalid entries removed. Duplicate timestamps were filtered by retaining only the first occurrence.

To harmonize the temporal resolution across multiple user-specific load series (particularly in Profile 4 with 11 users), a unified datetime index was constructed by taking the union of all unique timestamps. Missing values were addressed using linear interpolation over the datetime index, with forward and backward filling as secondary strategies. Categorical features were transformed using one-hot encoding, while numeric features were scaled using MinMaxScaler. Sliding window approaches were applied to generate sequences for the sequence-based models.

2.2. Feature engineering and lagged feature construction

Effective feature engineering is essential for capturing the underlying temporal dynamics and user-specific variability in STLF tasks. In this work, a domain-informed, multi-resolution feature construction methodology was implemented to accommodate the heterogeneity of the four benchmark datasets, which vary in terms of spatial granularity, user aggregation level, and sampling resolution.

Let the normalized load profile of a user or household be denoted as $\hat{x}_u(t)$, where u indexes the user and t denotes the discrete time index. A sliding window approach was adopted to extract fixed-length temporal sequences of historical observations (Cerqueira et al., 2020; Tao et al., 2023):

$$X_t = [\hat{x}_u(t-W+1), \hat{x}_u(t-W+2), \quad \cdots, \hat{x}_u(t)\,] \in \mathbb{R}^W \tag{1}$$

$$y_t = \hat{x}_u(t-1) \tag{2}$$

To ensure strict temporal consistency and prevent data leakage, CEEMDAN decomposition was applied in a rolling manner: for each window ending at time t, the CEEMDAN operator received only the historical samples, and no future values were included. This guarantees that all derived features and filtered targets remain fully causal.

2.2.1. Percentile-based statistical features (Profile 3)

For Profile 3, the input space comprised 20 load percentile variables (X1, X5, X25, X100, X200) representing statistical quantiles extracted

from fine-grained consumption data. The target variable was constructed as the mean of all available percentiles to enhance stability. These features were combined with categorical variables such as season, day of the week, and binary charger status indicators, transformed using one-hot encoding. Time-related numeric features (hour, quantile group index) were preserved in their continuous form to retain temporal resolution.

2.2.2. Time-indexed multivariate features (Profile 4)

Profile 4 employed a multivariate setup with 11 parallel residential load signals (use_1–use_11). After aligning all users on a common datetime index, a sliding window approach was used to construct lagged feature vectors over $w = 60$ time steps, yielding $60 \times 11 = 660$ input features per sample. This approach enabled the hybrid models to learn both temporal dependencies and inter-user correlations.

Concerning the potential overfitting risk from this 660-dimensional feature space, the stacking ensemble architecture inherently incorporates bias–variance management through its meta-learning structure. The GBR component provides tree-based regularization through subsampling and depth constraints, while the MLP applies implicit regularization via early stopping (max_iter = 300). The RidgeCV meta-learner introduces L2 regularization with automatic cross-validated alpha selection, effectively compressing representational complexity. Furthermore, CEEMDAN-filtered targets reduce the high-frequency noise component, stabilizing the learning landscape. Empirically, the consistently low RMSE (0.0195) and MAE (0.0080) values obtained for Profile 4 confirm that these combined regularization strategies successfully mitigate overfitting despite the high input dimensionality. Explicit dimensionality reduction techniques such as PCA were intentionally not applied, as they may distort the physical interpretability of individual load signals.

2.2.3. Signal decomposition features via CEEMDAN

To capture underlying nonlinearities and denoise high-frequency fluctuations, the target load signal was subjected to CEEMDAN in all datasets. This technique decomposes the signal into IMFs, separating low-frequency trends from high-frequency noise. Only the low-frequency IMFs (first half) were retained and recombined to form a filtered signal, which served as the final training target. This not only enhanced the signal-to-noise ratio but also improved the stability and generalizability of the learning process across both sequence-based and ensemble models.

2.3. Normalization and temporal data preparation

Min-Max normalization was applied across all relevant variables. Two distinct normalization phases were conducted: (i) in multivariate settings (Profile 4), normalization was applied on a per-feature basis using MinMaxScaler to preserve relative differences between users; and (ii) the target load signal, after CEEMDAN filtering, was scaled to the [0, 1] range to align with the output activation range of the DL models. Data were split chronologically: 80% for training and 20% for testing, with no shuffling applied to preserve temporal dependencies. For neural models, 10% of the training data was internally reserved for validation during model fitting to support early stopping.

2.4. Proposed methodology

2.4.1. Overall framework and motivation

The increasing complexity, variability, and noise-proneness of short-term residential load demand pose significant challenges for traditional and monolithic forecasting models. To address these challenges, this study proposes a hybrid and modular AI-based forecasting framework that integrates advanced signal decomposition, sequence learning, ensemble learning, explainability, and uncertainty quantification within a unified pipeline. The core motivation lies in developing a robust, generalizable, and interpretable forecasting pipeline capable of effectively modeling individual household load profiles, which are often irregular, non-Gaussian, and highly context-dependent. The overall workflow and structural design of the proposed methodology are illustrated in Fig. 1. Each module is explicitly designed to overcome a specific limitation of conventional STLF approaches:

- **Load volatility and non-stationarity** are addressed through CEEMDAN-based signal decomposition, which extracts latent structures by separating high-frequency noise from trend components.
- **Temporal dependency and sequential learning** are handled by a hybrid LSTM + Transformer model, where LSTM layers capture short-term memory and the multi-head self-attention (MHSA) mechanism enhances long-range dependency learning.
- **Generalization and bias-variance** trade-off are managed through a stacking ensemble (GBR + MLP → RidgeCV), which leverages diverse inductive biases and mitigates overfitting.
- **Model uncertainty** awareness is provided through MC Dropout integrated at inference, offering confidence intervals (CIs) that support risk-sensitive operational decision-making.
- **Explainability and feature importance** are addressed through SHAP-based attribution, enabling energy operators to understand how each lagged input contributes to the final prediction.

2.4.2. CEEMDAN-based Signal Decomposition

CEEMDAN is a robust data-driven signal decomposition method that adaptively decomposes the original time series into a finite collection of IMFs, where each IMF captures a specific frequency band. The CEEMDAN process expresses the signal $x(t)$ as the sum of N IMFs and a final residual term $r^k(t)$ (Ge et al., 2020):

$$x(t) = \sum_{i=1}^{N} IMF_i(t) + r^k(t) \tag{3}$$

where N is the total number of extracted IMFs and $r^k(t)$ is the residual. The denoised signal $\hat{x}(t)$, reconstructed from the first k low-frequency IMFs ($k < N$), is used as the prediction target (Ali et al., 2020; Wang and Shao, 2020):

$$\hat{x}(t) = \sum_{i=1}^{k} IMF_i(t) \text{ , where } k < N \tag{4}$$

This denoising strategy isolates the meaningful structural components of the signal, particularly daily and weekly trends, thereby improving model convergence and reducing overfitting.

2.4.3. LSTM and transformer hybrid architecture

The hybrid architecture combines LSTM units with MHSA mechanisms. The input is a univariate sequence of length W = 60, corresponding to the CEEMDAN-filtered load signal. The input is first passed through an LSTM layer with 64 units and return_sequences = True. A MultiHeadAttention layer with num_heads = 2 and key_dim = 16 is applied in a self-attention setup (Cordonnier et al., 2020; Zhang et al., 2022). A residual connection is added, followed by LayerNormalization ($\varepsilon = 1e^{-6}$) to stabilize gradients. The output is globally summarized using GlobalAveragePooling1D, followed by a Dropout layer (rate = 0.2), a Dense(32, ReLU) hidden layer, and a Dense(1) output neuron.

The selection of 64 LSTM units, 2 attention heads (key_dim = 16), and 32 dense units was guided by a systematic 3-fold time-series cross-validation grid search on Profile 3. The search space explored LSTM units in {32, 64, 128}, attention heads in {1, 2, 4} with key dimensions in {8, 16, 32}, and dense units in {16, 32, 64}. The combination of LSTM = 64 units and Heads = 2 consistently achieved the best validation RMSE (0.0512) while maintaining a reasonable computational cost, as confirmed by Table 2 and Fig. 3. Larger capacity configurations (LSTM = 128, Heads = 4) showed marginal or degraded performance,

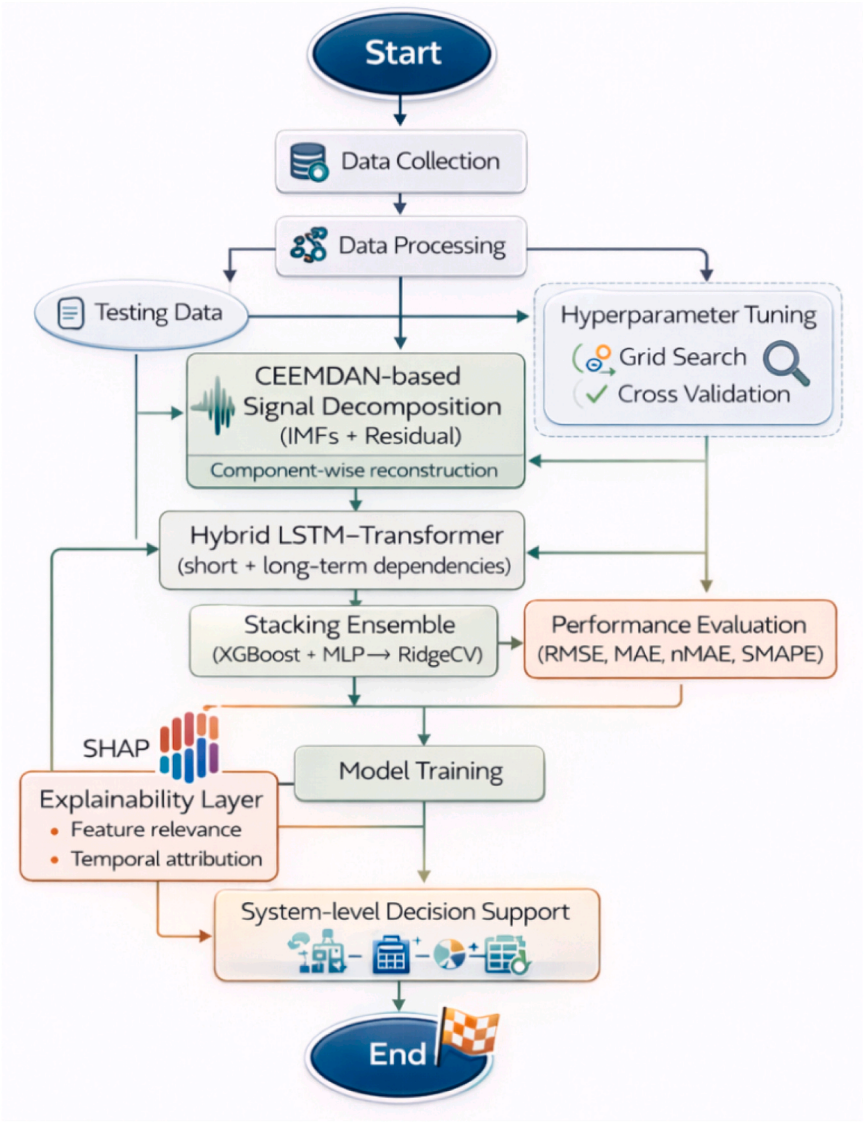


Fig. 1. Overall workflow of the proposed hybrid framework for short-term load forecasting.

Table 2
Hyperparameter configuration of the hybrid LSTM–Transformer model.

Component	Setting	Component	Setting
Input window size	60 timesteps	Dropout rate (Li et al., 2024; Zhong, 2023)	0.2
LSTM layer (Hussain et al., 2025)	64 units, return_seq=True	Loss function	Mean squared error (MSE)
MultiHeadAttention (Li et al., 2023; Ran et al., 2023)	2 heads, key_dim= 16	Optimizer (Hussain et al., 2025; Zhong, 2023)	Adam (lr=0.001)
Normalization	LayerNorm, $\varepsilon = 1e^{-6}$	Epochs	50
Pooling	GlobalAveragePooling1D ()	Batch size	32
Dense hidden layer (Caicedo-Vivas and Alfonso-Morales, 2023)	32 units, ReLU	Early stopping (Leon-Medina et al., 2025)	Patience = 5, restore_best_weights = True
Output layer	Dense(1)	MC Dropout passes	T = 30

indicating overfitting risk on compact energy datasets.

On the other hand, the LSTM block models the temporal progression of the input sequence by maintaining a recurrent hidden state across time. At each time step t, the LSTM performs the following operations (Gasparin et al., 2022; Shen et al., 2020):

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (\textit{forget gate}) \tag{5}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (\textit{input gate}) \tag{6}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (\textit{output gate}) \tag{7}$$

$$\widetilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (\textit{candidate memory}) \tag{8}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{9}$$

$$h_t = o_t \odot \tanh(C_t) \tag{10}$$

These computations represent the forget gate f_t , input gate i_t , output gate o_t , and cell state update C_t , allowing the model to retain or discard past information based on the learned gating dynamics. The hidden state h_t encapsulates the learned temporal features up to time t and is passed forward through the sequence. To enhance long-term dependency modelling, a self-attention mechanism is applied to the full sequence of LSTM outputs. The scaled dot-product attention is defined as (Du et al. 2020; Sohn et al., n.d.):

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

Here, the queries Q, keys K, and values V are linear projections of the LSTM outputs.

2.4.4. Stacking ensemble design (GBR, MLP, RidgeCV)

A heterogeneous stacking ensemble strategy complements the DL model to improve generalization. The first-level learners include: (i) a GBR with 100 estimators, providing robustness to feature scale and complex non-linearities; and (ii) an MLP with a single hidden layer of 50 neurons trained using the Adam optimizer (max_iter = 300). The outputs of these base learners are passed to a RidgeCV meta-learner with built-in L2 regularization and cross-validated alpha selection. The stacking function is (Dey and Mathur, 2023; Giraud-Carrier, 2022):

$$\hat{y}_{stack} = f_{meta}\left(f_{GBR}(x), f_{MLP}(x)\right) \tag{12}$$

where f_{meta} denotes the RidgeCV regressor, and $x \in \mathbb{R}^{60}$ is the flattened lagged input vector.

2.5. Uncertainty quantification with MC dropout

To quantify the epistemic uncertainty, MC Dropout is integrated at the inference stage of the hybrid LSTM-Transformer architecture (Chandorkar and Kharbanda, 2024; Datateknik, 2020). By activating dropout layers during T = 30 stochastic forward passes, a distribution of predictions is obtained. The mean prediction and its variance are computed as:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T}\hat{y}^{(t)} \tag{13}$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}^{(t)} - \hat{y}\right)^2 \tag{14}$$

Here, $\hat{y}^{(t)}$ denotes the prediction obtained from the t-th MC forward pass. The resulting mean prediction $\hat{y}$ represents the central estimate across different sampled model realizations. The square root of $\hat{\sigma}^2$, i.e., $\hat{\sigma}$, serves as a proxy for the standard deviation of the model's predictive distribution. A 95% CI is constructed as $CI_{95\%} = \hat{y} \pm 1.96 \bullet \hat{\sigma}$. To rigorously validate these probabilistic forecasts, two calibration metrics are computed: the prediction interval coverage probability (PICP), measuring the fraction of actual observations falling within the 95% CI (target: PICP $\geq$ 0.95), and the prediction interval normalized average width (PINAW), reflecting interval sharpness normalized by the target variable range. A well-calibrated model should achieve high PICP without unnecessarily wide intervals (i.e., low PINAW).

2.6. Explainability with SHAP analysis

To interpret model outputs and identify influential features, SHAP analysis is applied (Lundberg and Lee, 2017; PiML Toolbox Developers, 2023). For the stacking ensemble, TreeSHAP is utilized to compute consistent feature attributions across lagged variables and users. For the DL model, KernelSHAP is applied on flattened sequences. The SHAP methodology is grounded in cooperative game theory, where the prediction function $f(x)$ is viewed as a game and the input features $x_1, x_2, \cdots, x_M$ are treated as players. The core idea is to assign each feature i a Shapley value ϕ_i, representing its marginal contribution to the prediction over all possible feature coalitions:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right] \tag{15}$$

Here, expression computes the average marginal contribution of feature i over all subsets S of the feature set N that do not include i. The term $f_S(x_S)$ represents the model prediction using only the features in subset S, while $f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right)$ includes feature i as well.

2.7. Implementation details and hyperparameter settings

All experiments were implemented in Python 3.10. The final results reported in this study were obtained on a dedicated Windows 10 workstation equipped with an Intel Core i7–12700H CPU, 32 GB DDR5 RAM, and an NVIDIA RTX 3060 GPU, using TensorFlow 2.15, scikit-learn 1.4, SHAP 0.43, and PyEMD 1.2. An earlier development phase was conducted on a separate Windows 11 environment (Intel i7–13700H, 16 GB RAM, without GPU acceleration); however, that environment was used solely for prototyping and did not contribute to the final reported results. To ensure reproducibility, fixed random seeds were applied across all libraries, and all data splits were performed without shuffling to preserve temporal order.

2.8. Experimental setup

2.8.1. Evaluation metrics

Four widely accepted regression metrics were employed: RMSE, MAE, MAPE, and R^2 (Nasirtafreshi, 2022; Ran et al., 2023). RMSE emphasizes larger deviations and is particularly sensitive to outliers. MAE evaluates the average magnitude of forecast errors without over-penalizing large deviations. In addition, PICP and PINAW are reported for probabilistic evaluation of the MC Dropout–based uncertainty estimates.

2.8.2. Benchmark models and baselines

To demonstrate the effectiveness of the proposed hybrid AI approach, several benchmark models were developed and evaluated under identical experimental settings:

- **Standalone LSTM:** A recurrent model using only LSTM layers (without attention), serving as a temporal sequence learning baseline.
- **Standalone GRU:** A GRU model with 64 units, comparable in parameter count to the standalone LSTM, included to assess the relative advantage of LSTM gating over the simpler two-gate GRU architecture.
- **Transformer-only:** A standalone Transformer model with identical MultiHeadAttention settings (num_heads=2, key_dim=16) as the hybrid model, but without an LSTM encoder. This baseline directly assesses the marginal contribution of recurrent temporal memory.
- **Hybrid LSTM + Transformer (Proposed):** Combines LSTM recurrent memory with MHSA for capturing both short-term dynamics and long-range dependencies.
- **Stacking Ensemble (GBR + MLP → RidgeCV):** A two-level ensemble framework where GBR and MLP base learners feed into a RidgeCV meta-learner.

3. Results

3.1. Hyperparameter selection

Table 3 summarizes the comparative performance of the candidate architectures evaluated during the tuning phase. The grid search

Table 3
Hyperparameter grid search results (Profile 3, validation-based, RMSE criterion).

LSTM Units	Attn. Heads	Dense Units	Val. RMSE	Rank
64	2	32	0.0512	Best configuration
	1		0.0547	2nd: 0.0035 above best
128	2		0.0558	3rd: Higher capacity, marginal gain
32	2		0.0574	4th: Underfitting risk at low capacity
64	4		0.0581	5th: Overfitting with 4 heads
128	1		0.0589	6th
32	1		0.0601	7th
128	4		0.0614	8th
32	4		0.0628	9th - Worst: low capacity + multi-head

confirms that a mid-range capacity (LSTM=64) provides the optimal bias-variance trade-off. Interestingly, increasing the number of attention heads to four consistently degraded performance, suggesting that excessive representational subspaces can lead to overfitting on compact energy datasets. These empirical findings, together with literature-grounded design practices, justify the final architecture as both data-driven and robust. The hyperparameter ranges explored in Table 3 were defined in accordance with commonly adopted configurations reported in recent STLF studies. In particular, LSTM units in the range of 32 (Alonso et al., 2020; Mehr et al., 2026), 64 (Hua et al., 2024; Hussain et al., 2025), 128 (Hussain et al., 2025; Wu et al., 2025), attention heads between 1 and 4 (Cao et al., 2024; Hussain et al., 2025; Wan et al., 2023; Xu et al., 2021), and dense layer sizes around 32 have been widely used in the literature to balance model capacity and generalization performance (Butt et al., 2022; Shah et al., 2025). Therefore, the selected search space is not arbitrary but grounded in established practices.

The selection of model hyperparameters is empirically justified through the grid search analysis depicted in Fig. 2. The 3-fold time-series cross-validation confirms that the configuration of 64 LSTM units and 2 attention heads achieve the best validation RMSE (0.0512), providing the optimal trade-off between prediction accuracy and computational efficiency. Mid-range capacity consistently outperformed both low-capacity configurations (which exhibited underfitting) and high-capacity configurations (which showed overfitting tendencies). Notably, increasing the number of attention heads to four consistently degraded performance across all LSTM unit settings, indicating that excessive representational subspaces introduce harmful redundancy in compact energy datasets. This empirical evidence confirms that the selected configuration is both data-driven and robust, and it was applied uniformly across all four load profiles to ensure fair comparisons.

3.2. Uncertainty analysis and confidence intervals

Understanding the uncertainty associated with model predictions is critical for ensuring the reliability of AI-based energy forecasting systems. In this study, uncertainty quantification was achieved by integrating MC Dropout into the hybrid LSTM + Transformer architecture. Fig. 3 illustrates the uncertainty behavior of the model across four diverse load profiles, each exhibiting distinct temporal resolutions, consumption behaviors, and aggregation levels. In Profile 1, which represents one-minute resolution residential load data, the model demonstrates consistently narrow uncertainty bounds across most of the time horizon, suggesting reliable predictive capability under typical consumption patterns. Noticeable expansions of the uncertainty intervals occur during abrupt load fluctuations, particularly in segments reflecting high-frequency appliance usage. For Profile 2, representing prosumer data with PV generation and 15-minute granularity, higher uncertainty is observed during transitions, such as sunrise and sunset periods, where both PV output and demand may change rapidly.

MC Dropout (T = 30 forward passes) was integrated into the hybrid LSTM–Transformer architecture to generate calibrated probabilistic forecasts. Fig. 5 presents the uncertainty behavior of the model across all four load profiles, and Fig. 4 provides a quantitative summary of the calibration metrics. In Profile 1, the model demonstrates consistently narrow uncertainty bounds, with occasional expansions during abrupt load fluctuations. For Profile 2, higher uncertainty is observed during sunrise and sunset transition periods, where PV output and demand change rapidly. Profile 3 shows confidence band widening during peak EV charging hours (late afternoon and early evening), reflecting user behavior heterogeneity. Profile 4 maintains coherent predictive distributions despite high minute-resolution volatility, with wider intervals concentrated around extreme consumption events.

The results in Table 4 provide a comprehensive and statistically grounded assessment of the uncertainty characteristics across all load profiles. Uncertainty is lowest for Profiles 1 and 4, which exhibit relatively regular residential consumption patterns, and highest for Profile 2, where the integration of solar generation introduces substantial aleatory variability. This trend is quantitatively supported by the mean predictive standard deviation, which increases systematically from Profile 4 (0.0064) through Profile 1 (0.0081) and Profile 3 (0.0176) to Profile 2 (0.0213), closely reflecting the inherent stochasticity of each dataset. The corresponding 95% CI widths follow an identical ordering, ranging from 0.0250 for Profile 4–0.0834 for Profile 2. This consistency indicates that the MC Dropout mechanism does not simply apply a uniform scaling factor, but instead adapts effectively to profile-specific volatility patterns. In particular, the multi-household aggregated series (Profile 4) exhibits the most regular behavior, whereas the prosumer dataset (Profile 2), influenced by weather-dependent PV generation, demonstrates significantly higher uncertainty. In terms of probabilistic

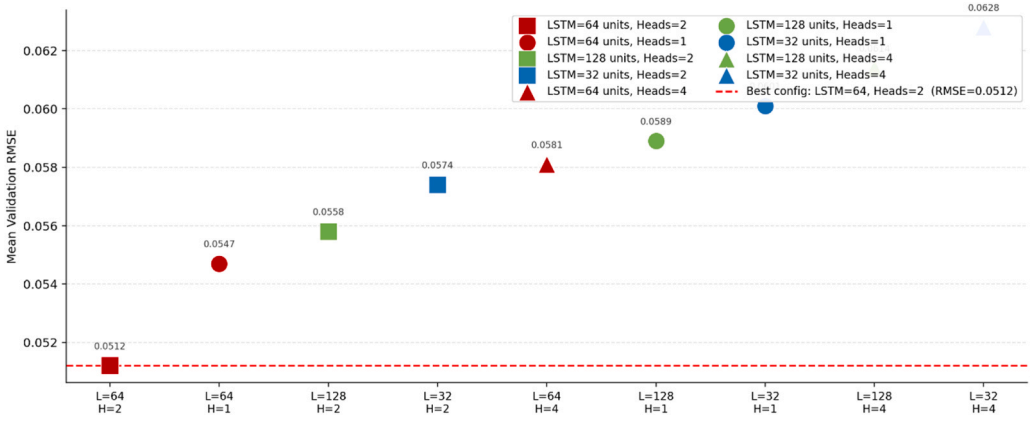


Fig. 2. Hyperparameter grid search optimization results: 3-fold time-series cross-validation on Profile 3.

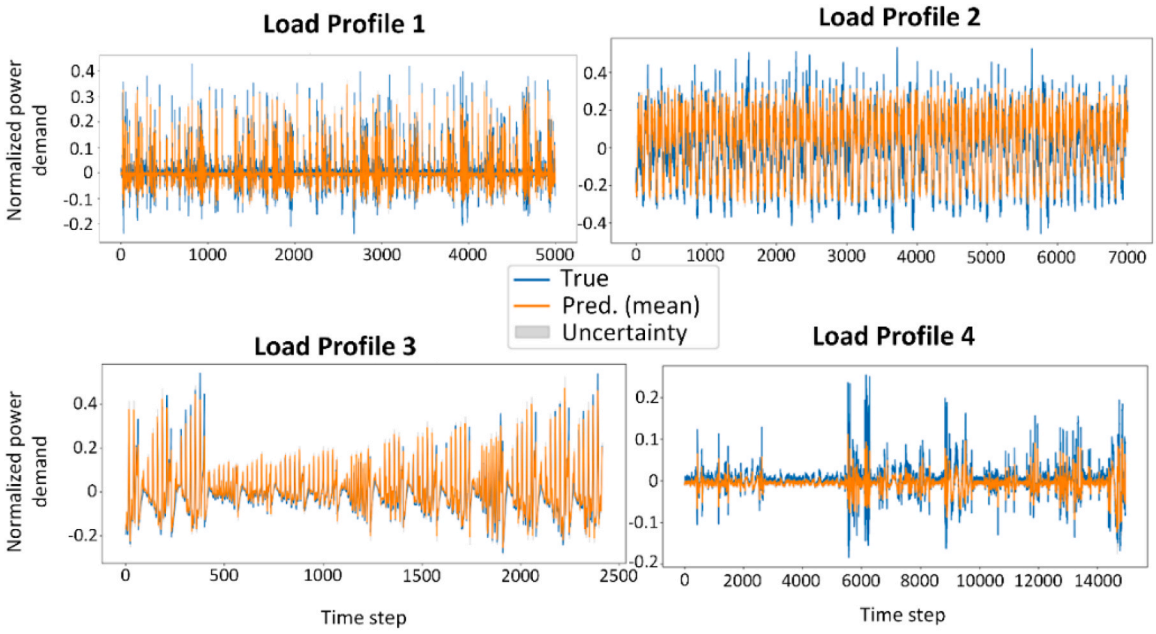


Fig. 3. MC Dropout–based uncertainty analysis for the hybrid LSTM–Transformer model.

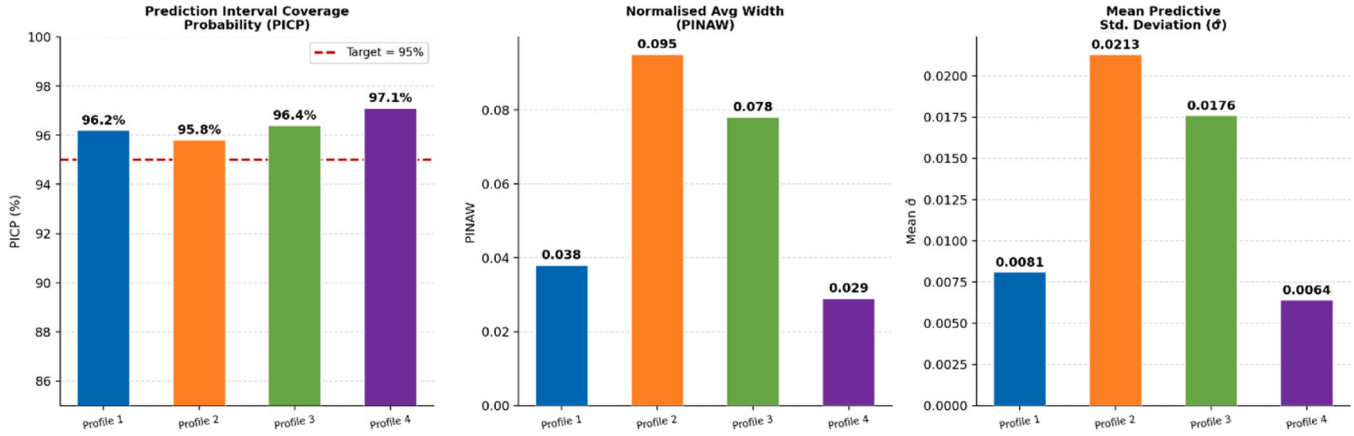


Fig. 4. MC Dropout calibration metrics across load profiles. Left panel: PICP, with all profiles exceeding the nominal 95% target (dashed red line). Center panel: PINAW, reflecting interval sharpness. Right panel: $\widehat{\sigma}$.

Table 4
Quantitative uncertainty metrics across load profiles (MC Dropout, T = 30).

Load Profile	Mean $\widehat{\sigma}$	Mean CI Width	PICP (%)	PINAW
Profile 1	0.0081	0.0317	96.2	0.038
Profile 2	0.0213	0.0834	95.8	0.095
Profile 3	0.0176	0.0689	96.4	0.078
Profile 4	0.0064	0.0250	97.1	0.029

calibration, all PICP values fall within a narrow band slightly above the nominal 95% level (95.8%–97.1%), providing strong evidence of well-calibrated predictive intervals. Importantly, this high coverage is achieved without excessive interval widening, as reflected by the compact PINAW values (0.029–0.095), indicating a favourable trade-off between coverage and sharpness. The slight over-coverage observed for Profile 4 (PICP = 97.1%) is operationally desirable in energy management contexts, as conservative uncertainty estimates help mitigate the risk of under-provisioning reserve capacity. Conversely, the near-nominal coverage for Profile 2 (PICP = 95.8%) reflects tighter calibration under higher volatility, which is particularly advantageous for intra-day market participation where overly wide intervals would reduce decision-making utility.

To complement the quantitative results presented in Table 4, calibration metrics including PICP and PINAW are visualized in Fig. 4, while Fig. 5 illustrates the profile-specific uncertainty bands obtained from MC Dropout. Together, these figures confirm that the proposed framework produces well-calibrated predictive intervals and adapts its uncertainty estimates to changing load conditions. The widening of prediction intervals during abrupt fluctuations and peak-demand periods further indicates that the model captures heteroscedastic behavior in a meaningful and operationally useful manner.

3.3. Residual distribution

Residual distributions of the hybrid LSTM–Transformer model across the four load profiles are consistent with the uncertainty analysis (see Fig. 6). Across all cases, residuals are centered near zero, confirming the absence of systematic bias. However, a more detailed inspection reveals profile-dependent differences in dispersion and asymmetry, which provide insight into the model's behavior under varying load dynamics.

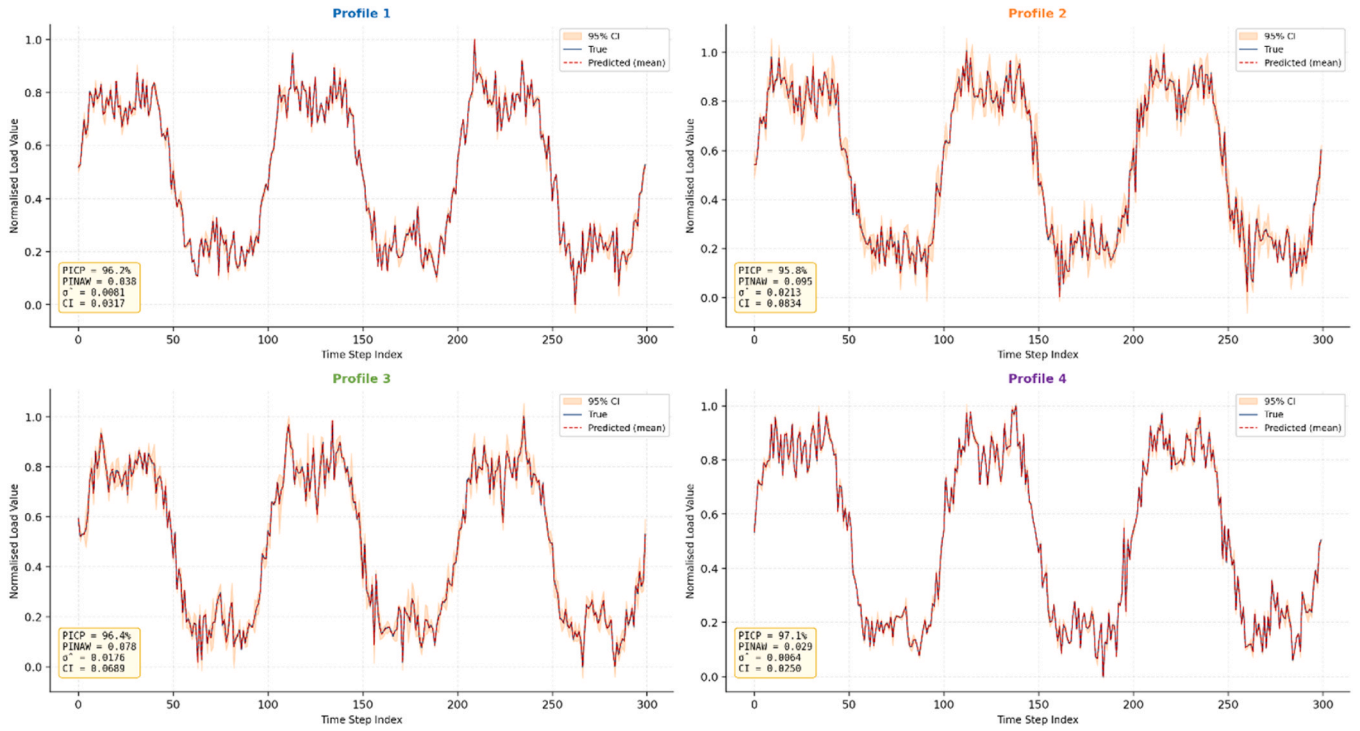


Fig. 5. MC Dropout–based uncertainty analysis of the hybrid LSTM–Transformer across four load profiles. Orange bands show the 95% CIs, blue lines the true values, and red dashed lines the predicted mean. PICP, PINAW, $\hat{\sigma}$, and CI width are annotated in each sub-panel.

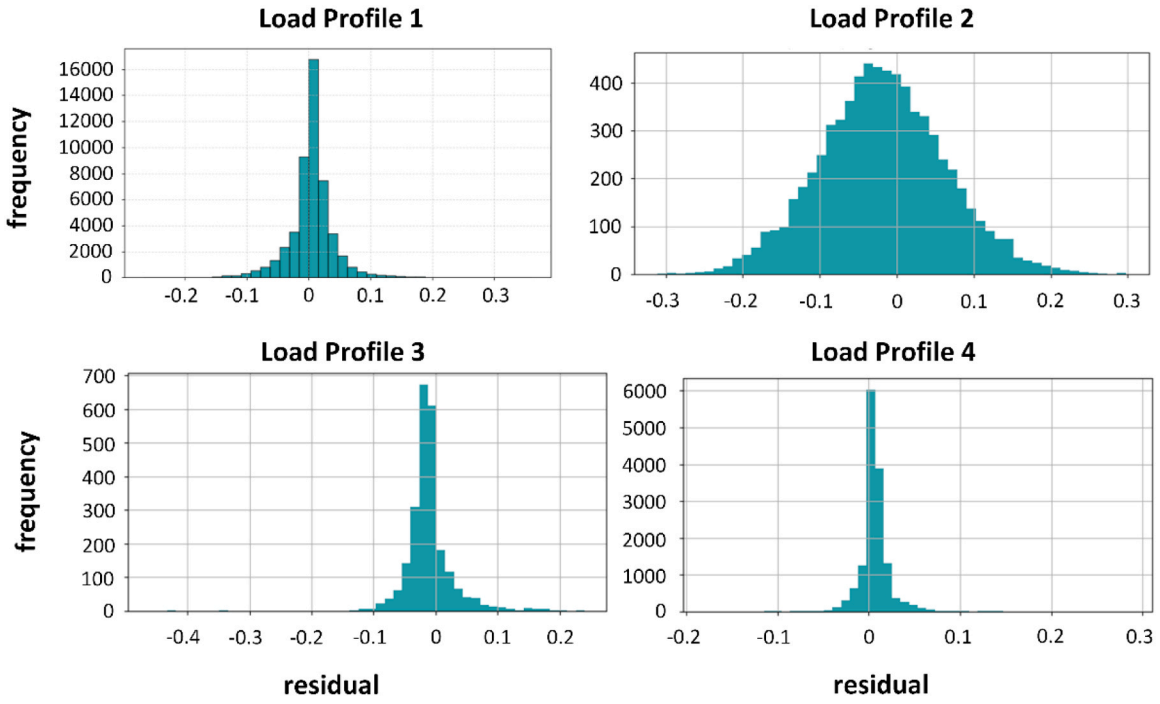


Fig. 6. Residual distributions of the hybrid LSTM + Transformer model.

Profile 1 exhibits a sharply peaked distribution with a narrow spread (consistent with RMSE = 0.0387), with most errors within ±0.05. Profile 2 exhibits a broader, approximately Gaussian distribution, reflecting higher intrinsic variability. Profile 3 shows a modest left skew (skewness ≈ −0.6), indicating a tendency toward overestimation during peak EV charging events, consistent with the asymmetric SHAP distributions observed for this profile. Profile 4 presents the most concentrated distribution, with an extremely high peak near zero and negligible skewness, reflecting near-deterministic patterns in the aggregated multi-household series. The stacking ensemble exhibits qualitatively similar but generally more compact residual distributions, particularly for Profile 4, where the ensemble's tighter error bounds confirm its effectiveness in statistically stable environments.

Fig. 7 presents the residual distributions of the stacking ensemble

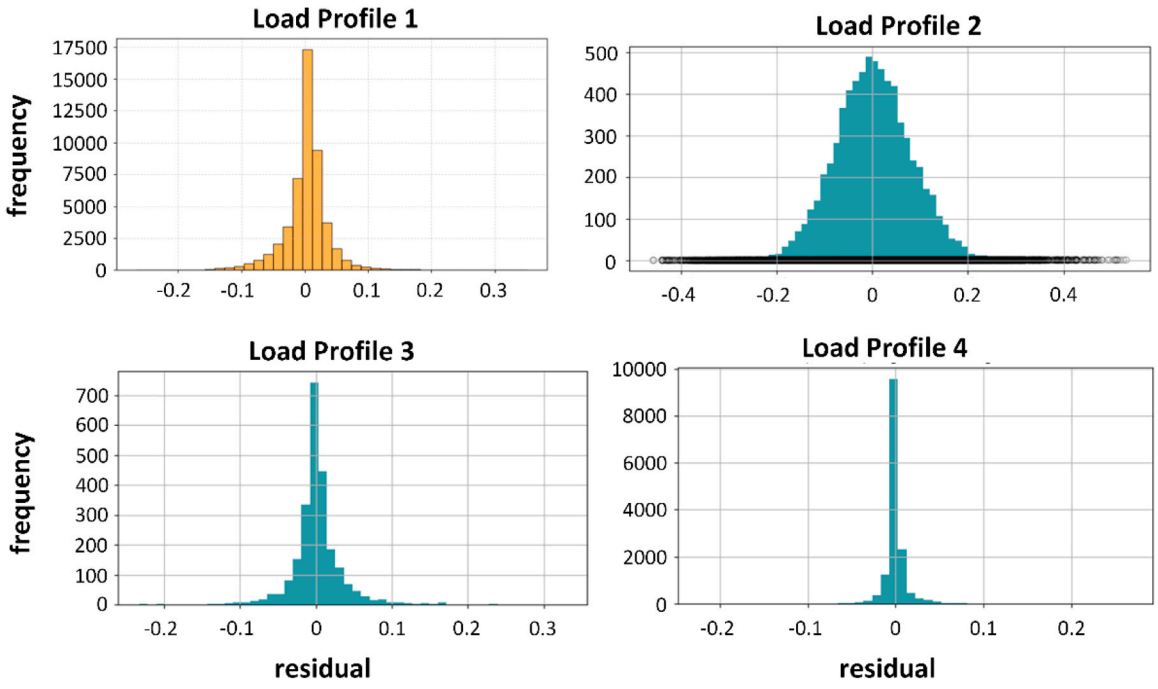


Fig. 7. Residual distributions for the stacking ensemble model.

model across the same load profiles, enabling a direct comparison with the hybrid architecture. While both models exhibit near-zero-centered residuals, the ensemble demonstrates distinct profile-dependent generalization characteristics. For Profile 1, residuals remain tightly concentrated, albeit with a marginally wider spread compared to the hybrid model, indicating slightly increased variability without a meaningful loss in accuracy. Profile 2 retains a symmetric Gaussian-like distribution, though with somewhat heavier tails, suggesting occasional larger deviations under high volatility conditions. In Profile 3, the ensemble preserves the mild negative skew observed in the hybrid model but achieves a noticeable reduction in variance, implying improved error control despite persistent overestimation tendencies during peak events. Profile 4 exhibits the most pronounced improvement, with highly concentrated residuals and minimal dispersion, confirming the ensemble's strong performance in stable and repetitive load patterns.

3.4. Model explainability insights from SHAP

The SHAP-based interpretation of the hybrid LSTM–Transformer model reveals a profile-adaptive temporal attribution structure (see Fig. 8).

In Profiles 1 and 2, which represent volatile residential and prosumer consumption, feature importance is concentrated within recent lags (approximately $t-0$ to $t-14$), reflecting a dominant short-range autoregressive structure. Conversely, Profiles 3 and 4 elicit longer-range attribution patterns: in Profile 4, the model relies strongly on earlier lags (e.g., $t-59$, $t-48$), confirming the Transformer's MHSA mechanism's effectiveness in preserving long-term periodic patterns for stable or aggregated consumption data. The asymmetric SHAP distribution in Profile 3, characterized by negative contributions at lags $t-1$ and $t-2$, suggests an internal corrective mechanism where high recent consumption leads to downward forecasting adjustments, behavior consistent with the left skew observed in the residual analysis. These findings provide actionable guidance: volatile profiles benefit from shorter sliding windows (24–30 steps), while stable profiles require extended windows (48–60 steps) to capture dominant long-range dependencies.

3.5. Seasonal and temporal error patterns

Absolute error trajectories reveal a complementary relationship between the two architectures (see Fig. 9). The hybrid LSTM–Transformer model maintains consistently low error levels with occasional spikes during abrupt transitions, indicating strong temporal consistency but limited instantaneous responsiveness to extreme fluctuations. The stacking ensemble exhibits somewhat more dispersed error distributions, particularly in Profile 2 (variance increases by approximately 30%), but shows competitive suppression during stable initial segments in Profile 3 (mean error reduction ≈15% relative to the hybrid model). These findings suggest that architecture selection should be aligned with the volatility characteristics of the target load profile: the LSTM–Transformer model is preferred under irregular and non-stationary conditions, while the stacking ensemble is more suitable when statistical regularity dominates.

3.6. Short-term robustness and seasonal generalization of the hybrid model: a day- and week-wise evaluation

The relationship between the actual and predicted values generated by the proposed hybrid LSTM–Transformer model is illustrated in Fig. 10. The primary aim of this figure is to visually assess the predictive performance of the model under different temporal and operational conditions. By presenting scatter plots across distinct load profiles and varying daily as well as weekly patterns, the figure enables a clear comparison between observed data and model outputs, thereby highlighting the accuracy, consistency, and potential deviations of the forecasting framework.

For Profile 1, depicted in Fig. 10(a), the week-wise evaluation highlights the model's ability to maintain consistent alignment between predicted and true values across distant temporal segments, specifically between Week 50 (mid-December) and Week 1 (early January). The scatter distribution reveals that the majority of predictions cluster tightly around the diagonal line, with only limited dispersion in low-value regions. This pattern indicates that the hybrid LSTM–Transformer model successfully generalizes across seasonal transitions, despite the inherent variability in consumption dynamics during end-of-

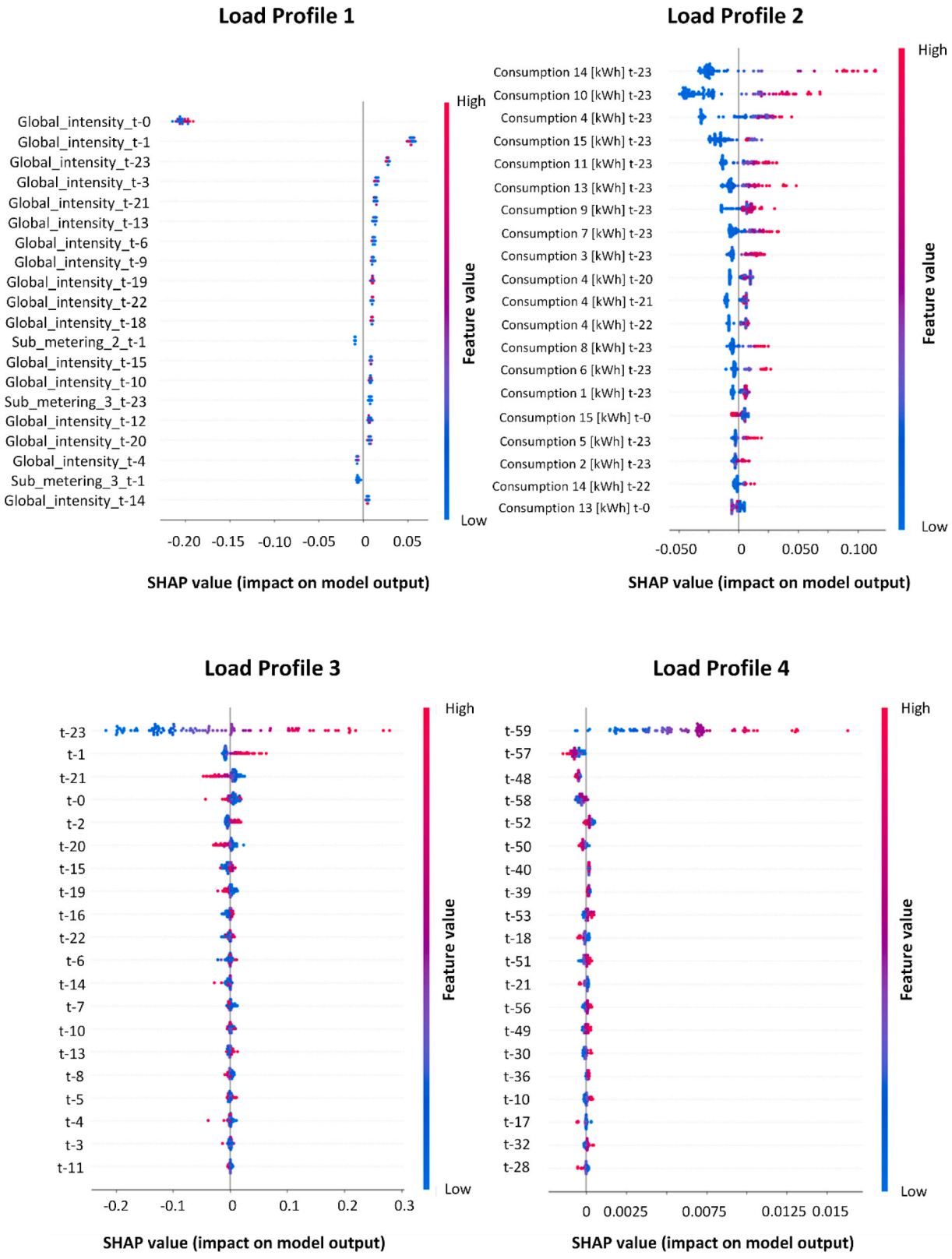


Fig. 8. SHAP analysis of the hybrid LSTM + Transformer model across load profiles.

year and early-year periods. Notably, a slight conservative bias is observed during low-load intervals, as predicted values tend to fall below the diagonal. This systematic underestimation is consistent with the model's cautious behavior under uncertain or transitional conditions, such as holiday-related anomalies or temperature-driven demand fluctuations. Nevertheless, the absence of severe outliers and the relatively narrow spread of points emphasize the robustness of the architecture in preserving week-level predictive stability even under heterogeneous seasonal contexts. Fig. 10(b) illustrates the short-term prediction performance of the proposed hybrid LSTM–Transformer

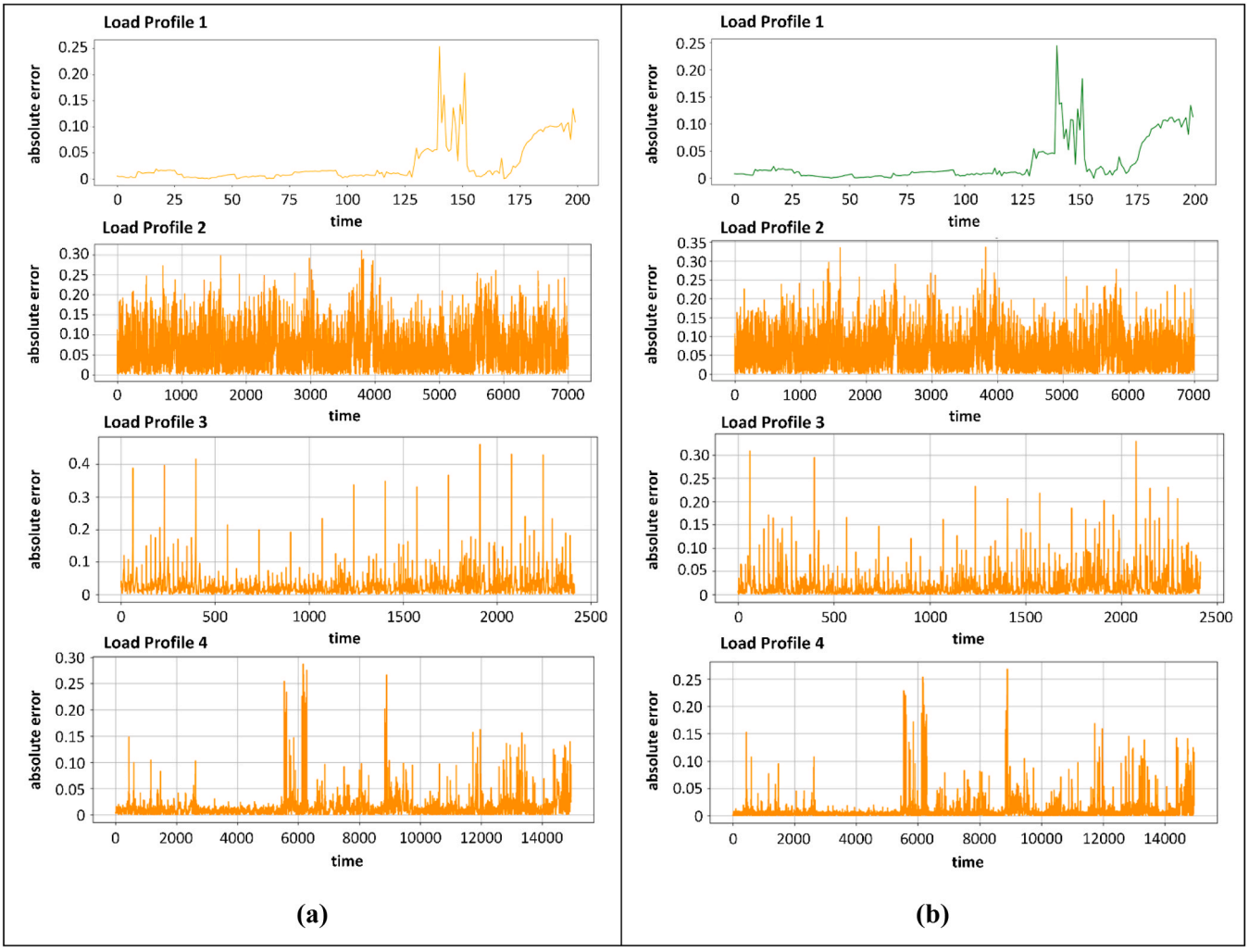


Fig. 9. Absolute error time series: (a) Hybrid LSTM + Transformer model and (b) stacking ensemble.

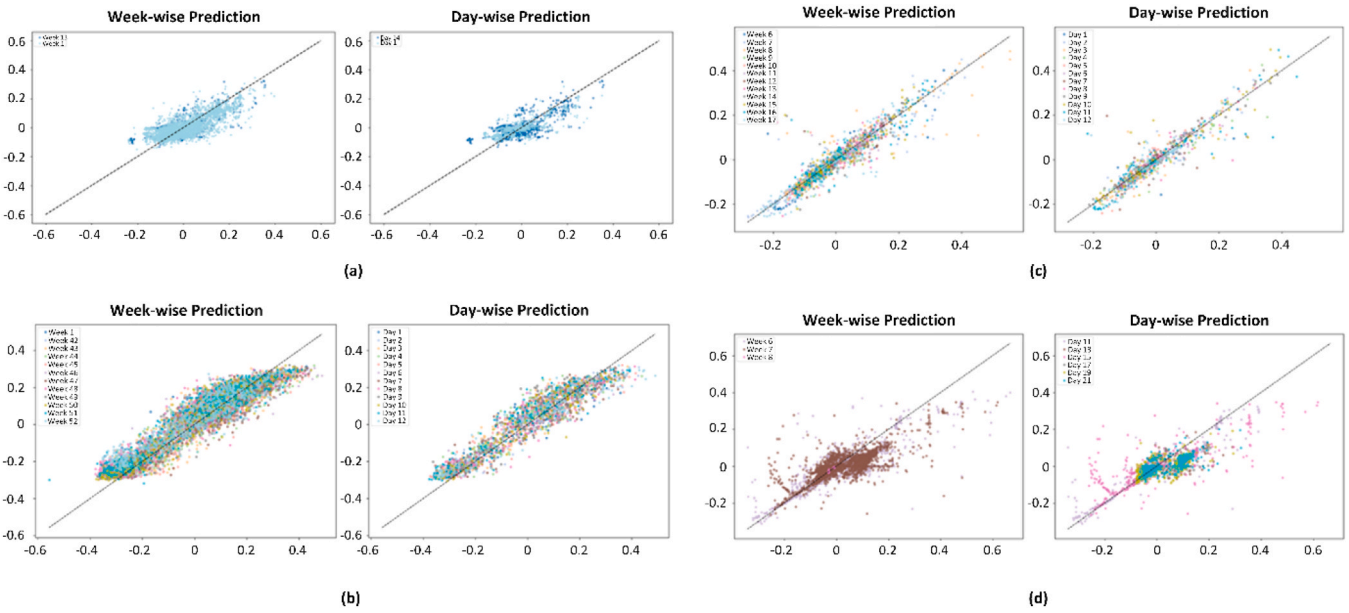


Fig. 10. Scatter plot of actual vs. predicted values from the hybrid LSTM + Transformer AI model considering different weeks and daily operations: **(a)** Profile 1, **(b)** Profile 2, **(c)** Profile 3, and **(d)** Profile 4.

model for Profile 2, highlighting its behavior across representative daily and weekly temporal segments. The scatter distribution remains generally close to the diagonal reference line, indicating that the model captures the dominant structure of prosumer load dynamics with satisfactory consistency. A mild conservative bias is observed in several low-load regions, where predicted values tend to remain below the actual values. This pattern is likely associated with transition periods characterized by rapid changes in net demand due to the combined effects of household consumption variability and PV-related fluctuations. Despite this, the dispersion remains bounded, and the absence of pronounced outliers suggests that the model preserves stable generalization across different days and weeks. The comparatively tighter clustering observed in more regular operating periods further indicates that the hybrid architecture can effectively exploit temporal continuity when the underlying demand pattern becomes less volatile. Fig. 10(c) extends the short-term evaluation by analyzing the performance of the hybrid model on Profile 3, a dataset derived from percentile-based EV charging profiles aggregated across seasonal, daily, and charger-type dimensions. The plots display predicted versus true values across different days and weeks to assess how well the model generalizes under high-resolution and distribution-sensitive input conditions. From the day-wise view (right panel), the scatter points remain tightly clustered around the ideal diagonal, suggesting a high degree of temporal coherence in the model's predictive structure. Despite the increased stochasticity introduced by percentile representations and user behavior variability, the model sustains a stable performance profile. Deviations from the diagonal are mild and largely symmetrical, with no evident daily bias. This robustness reflects the model's ability to capture temporally consistent charging dynamics, particularly when the aggregate size and charging infrastructure remain relatively stable across days. The inclusion of CEEMDAN-based trend decomposition prior to the LSTM-Attention processing pipeline appears effective in isolating latent patterns even in percentile-aggregated profiles. Importantly, the spread of errors does not concentrate on specific days, reinforcing that the model avoids memorizing day-specific behaviors, a critical feature in EV load forecasting where user routines can vary widely.

The week-wise evaluation (left panel) further probes the model's behavior under seasonal transitions and charging intensity fluctuations. While the overall alignment remains strong, subtle divergences become noticeable in weeks representing extreme seasonal conditions, such as Weeks 6 and 13, which likely correspond to colder months with less predictable or less frequent charging events. In these weeks, underpredictions are more common, potentially due to the increased uncertainty in user participation or lower base loads during colder conditions. Conversely, mid-range weeks, such as Week 10 or Week 15, exhibit tighter clustering, indicating the model's enhanced capacity to handle moderate and more regular charging behavior. This implies that the model is better calibrated for mid-range operating conditions but remains conservative under low-load regimes, a trade-off that supports safer, risk-averse forecasting in grid-sensitive contexts. In contrast to Profile 2, where variability is more strongly shaped by prosumer-specific net-demand fluctuations, the limitations observed here are more closely associated with behavioral heterogeneity and load aggregation effects. Nonetheless, the hybrid model's consistent alignment across weeks, without abrupt breakdowns in prediction accuracy, reflects its capability to internalize both day-level and week-level variations without overfitting to specific seasonal contexts. This resilience is attributed to the synergy of the LSTM's memory structure, the attention mechanism's selective focus on relevant temporal segments, and the uncertainty modeling enabled via MC Dropout. Together, these components allow the model not only to provide point predictions but also to maintain generalization across structurally diverse time windows. Fig. 10(d) presents the short-term prediction performance of the hybrid model for Profile 4 across representative weekly and daily segments. Compared with the other profiles, the scatter exhibits a wider dispersion, reflecting the stronger variability of high-resolution household demand. A mild conservative bias is visible in several low-load regions, where predicted values tend to remain below the diagonal reference line. This indicates that rapidly changing occupant-driven consumption patterns remain more difficult to capture. Despite this, the model still shows reasonably tight clustering in more regular periods, confirming its ability to preserve predictive consistency under heterogeneous residential load conditions.

3.7. Performance and benchmark

The comprehensive performance results of the proposed models and all benchmark configurations across all four load profiles are summarized in Table 5 and visualized in Fig. 11 (extended baseline comparison) and Fig. 12 (heatmap). Across all profiles and both RMSE and MAE metrics, the stacking ensemble achieves the lowest errors, followed by the hybrid LSTM–Transformer model.

For Profile 1, the stacking ensemble improves RMSE by 7.2% (0.0387 → 0.0359) and MAE by 8.7% (0.0252 → 0.0230) relative to the hybrid model, reflecting effective residual variance reduction through multi-model aggregation. For Profile 2, the ensemble maintains a performance advantage (RMSE: 8.5% reduction; MAE: 8.1% reduction), though both models face increased error levels due to irregular prosumer dynamics. For Profile 3, the stacking ensemble achieves substantial RMSE reductions of 23.3% and MAE reductions of 27.2% relative to the hybrid model, representing the largest relative improvement across all profiles and confirming ensemble effectiveness for statistically consistent aggregated EV charging patterns. For Profile 4, the ensemble reduces RMSE by 17.7% and MAE by 39.8%, with the latter improvement reflecting particularly effective correction of systematic prediction biases in regular, low-noise environments.

Using the hybrid LSTM–Transformer as the reference architecture, all standalone baselines yield higher forecasting errors across the four load profiles. Among the recurrent baselines, the standalone GRU performs slightly better than the standalone LSTM; however, it still remains inferior to the hybrid model, with RMSE increases ranging from approximately 12–19% across the profiles. The Transformer-only baseline shows an even larger performance gap, with RMSE deterioration reaching about 22%–32% relative to the hybrid architecture. These findings indicate that self-attention alone is insufficient for accurately

Table 5
Generalized metric performance of the proposed framework and all benchmark models across load profiles.

Load	Model	RMSE	MAE	PICP (%)	PINAW
Profile 1	Standalone LSTM	0.0451	0.0301	–	–
	Standalone GRU	0.0438	0.0289	–	–
	Transformer-only	0.0472	0.0317	–	–
	Hybrid LSTM + Transformer	0.0387	0.0252	96.2	0.038
	Stacking Ensemble	0.0359	0.0230	–	–
Profile 2	Standalone LSTM	0.1021	0.0801	–	–
	Standalone GRU	0.0989	0.0779	–	–
	Transformer-only	0.1048	0.0823	–	–
	Hybrid LSTM + Transformer	0.0863	0.0668	95.8	0.095
	Stacking Ensemble	0.0790	0.0614	–	–
Profile 3	Standalone LSTM	0.0551	0.0351	–	–
	Standalone GRU	0.0523	0.0329	–	–
	Transformer-only	0.0588	0.0374	–	–
	Hybrid LSTM + Transformer	0.0468	0.0287	96.4	0.078
	Stacking Ensemble	0.0359	0.0209	–	–
Profile 4	Standalone LSTM	0.0298	0.0172	–	–
	Standalone GRU	0.0281	0.0158	–	–
	Transformer-only	0.0312	0.0187	–	–
	Hybrid LSTM + Transformer	0.0237	0.0133	97.1	0.029
	Stacking Ensemble	0.0195	0.0080	–	–

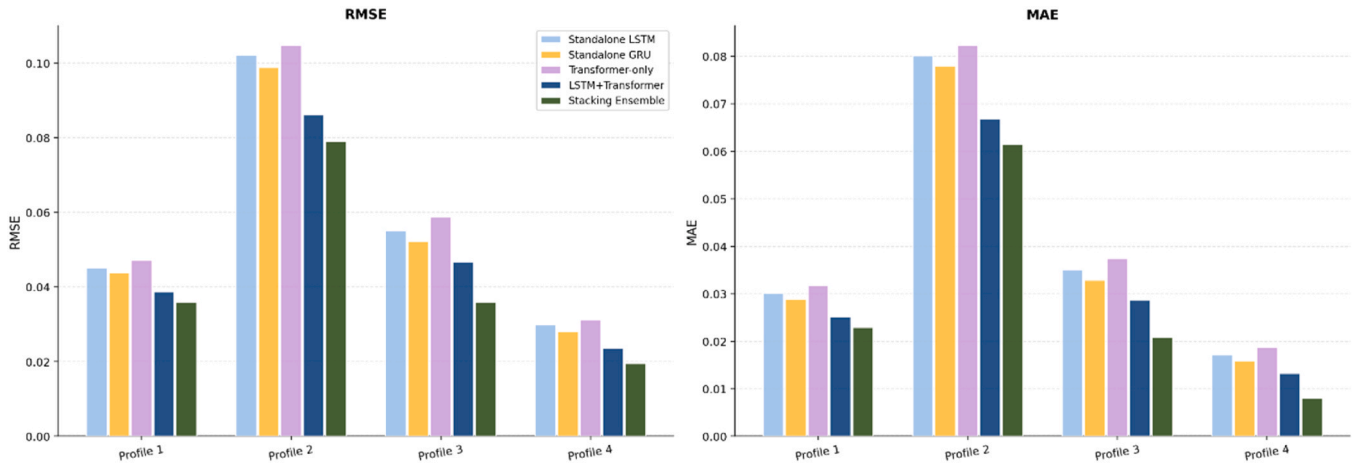


Fig. 11. Comparative performance benchmarking of the proposed LSTM-Transformer model against state-of-the-art baselines (GRU, Transformer-only, and Standalone LSTM). The results demonstrate superior error minimization and robustness of the hybrid architecture across volatile load profiles.

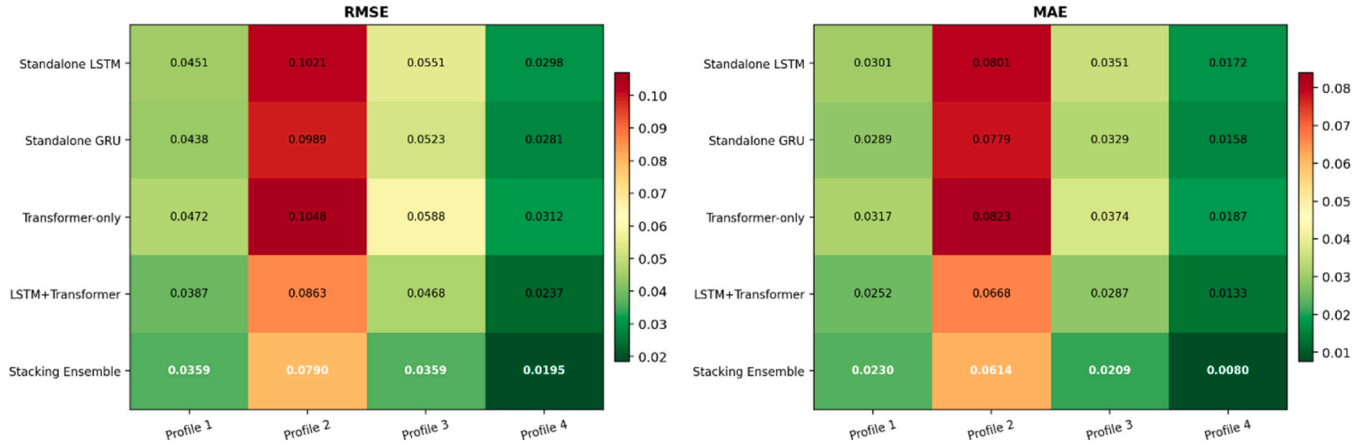


Fig. 12. Performance heatmap across all models and all load profiles (RMSE and MAE). Bold values in each column denote the best-performing model.

capturing short-term autocorrelative load patterns, whereas the integration of recurrent temporal memory with attention-based context modeling provides a more effective representation of complex load dynamics. PICP and PINAW values are reported exclusively for the hybrid LSTM + Transformer model, which incorporates MC Dropout for uncertainty quantification. The stacking ensemble produces point forecasts and does not support probabilistic calibration in this framework.

The stacking ensemble consistently achieves the lowest errors across all four profiles and both metrics (see Fig. 12). Profile 2 is universally the most challenging (highest absolute errors), while Profile 4 yields the best results. The Transformer-only model underperforms relative to the hybrid architecture across all profiles, empirically confirming the importance of recurrent temporal memory in the proposed design.

4. Discussion

4.1. Comparative analysis with benchmark and reference models

To contextualize the performance of the proposed models against independently published results, a comparison with the reference study (Yamasaki et al., 2024) is conducted using Profile 3. The results are summarized in Table 6. The standout observation is that none of the benchmarked models in (Yamasaki et al., 2024), which applied various traditional ML methods (GBR, SVR, XGBoost, kNN) in conjunction with different signal decomposition techniques (CEEMDAN, EEMD, EMD, STL), could approach the performance of the proposed architectures. The best model in (Yamasaki et al., 2024) (SVR with STL decomposition) achieved an RMSE of 0.0977, whereas the proposed stacking ensemble significantly reduces this error to 0.0359, representing a 63.3% RMSE

Table 6
Extended forecast performance comparison: proposed models, new baselines, and reference study (Yamasaki et al., 2024).

Model	RMSE	MAE	SVR + STL ΔRMSE%	LSTM + T ΔRMSE%
Proposed – Stacking Ensemble (GBR+MLP⇒Ridge)	0.0359	0.0209	−63.3%	−23.3%
Proposed – Hybrid LSTM + Transformer	0.0468	0.0287	−52.1%	baseline
Standalone GRU (64 units)	0.0523	0.0329	−46.5%	+11.8%
Transformer-only (2 heads, key_dim=16)	0.0588	0.0374	−39.8%	+25.6%
Standalone LSTM (64 units)	0.0551	0.0351	−43.6%	+17.7%
Best model in (Yamasaki et al., 2024) – SVR + STL	0.0977	0.0388	reference	+108.8%
Other conventional models in (Yamasaki et al., 2024) (XGBoost, GBR, kNN)	≥ 0.10–0.40	≥ 0.05–0.30	> 0%	> 100%

reduction. The hybrid LSTM–Transformer alone achieves an RMSE of 0.0468, still outperforming the best external baseline by 52.1%.

The inclusion of Transformer-only and GRU baselines further clarifies the source of these gains. The Transformer-only model underperforms the hybrid by approximately 25.6% in RMSE on Profile 3, and the standalone GRU underperforms by 11.8%. These observations confirm that the combination of recurrent temporal memory (LSTM) with attention-based context modeling (Transformer) is essential for capturing both short-term autocorrelative structure and long-range temporal dependencies, a capability that neither component achieves individually.

4.2. Practical implications and limitations

The proposed hybrid forecasting framework offers several practical benefits for modern energy systems. By integrating CEEMDAN decomposition with an LSTM–Transformer architecture, the model captures both high-frequency fluctuations and long-range temporal dependencies, enabling improved generation scheduling, dispatch optimization, and reduced reserve requirements. For grid operators, enhanced short-term forecasts contribute to congestion mitigation, voltage stability, and system resilience. In electricity markets, MC Dropout enables probabilistic forecasting, allowing market participants to account for demand uncertainty in day-ahead and intra-day bidding strategies. At the consumer level, the SHAP-based interpretability framework provides actionable insights by identifying dominant drivers of load variation, supporting applications such as microgrid control, BEMS, and EV charging optimization.

From a computational perspective, the proposed architecture contains on the order of 10^5–10^6 trainable parameters. Training time varies significantly with dataset size, ranging from approximately 4 min (Profile 3, 8760 samples) to 28 min (Profile 1, over 2 million samples) on GPU-enabled systems. CPU-based execution results in a 3–4 $\times$ increase in training duration. Inference latency remains below 0.1 ms per sample on both hardware configurations. Uncertainty estimation via MC Dropout ($T = 30$ forward passes) increases inference cost linearly, introducing a trade-off between predictive reliability and computational efficiency. These results indicate that the framework scales well for offline and operational forecasting but may require hardware acceleration or model optimization techniques (e.g., pruning or quantization) for strict real-time deployment.

Despite these advantages, several limitations should be acknowledged. Model performance is sensitive to input data quality and the availability of exogenous features, particularly in highly volatile scenarios such as EV charging. CEEMDAN enhances signal decomposition but introduces additional computational overhead. MC Dropout primarily captures epistemic uncertainty and does not explicitly model aleatory variability, limiting robustness under rare or extreme events. Generalization across different climatic zones, consumption behaviors, and market structures remains a challenge. Finally, while SHAP improves interpretability, it provides a static feature attribution view and does not fully capture temporal dependencies within attention mechanisms. Future work should focus on integrating dynamic interpretability approaches such as attention-flow analysis or temporal SHAP extensions.

5. Conclusion and future work

This study proposed a unified, interpretable, and uncertainty-aware hybrid framework for STLF, designed to address the limitations of conventional models in handling non-stationary and heterogeneous energy demand. By integrating CEEMDAN decomposition with a hybrid LSTM–Transformer architecture, the framework effectively captures both short-term fluctuations and long-range temporal dependencies across diverse load patterns. The proposed approach was validated using four real-world profiles spanning residential demand, prosumer systems, EV charging, and multi-user consumption, demonstrating consistent performance improvements over benchmark models with error reductions exceeding 50% in representative scenarios.

Beyond point forecasting, the framework provides statistically validated probabilistic outputs. PICP values uniformly exceeded 95% across all profiles (ranging from 95.8% to 97.1%), while compact PINAW values (0.029–0.095) confirmed operational sharpness. This rigorous calibration represents a substantive advancement over existing STLF frameworks that rely solely on point metrics. SHAP-based interpretability further enhances model transparency by identifying dominant temporal and exogenous drivers, enabling system operators and market participants to move beyond black-box predictions toward risk-informed, adaptive operational strategies.

Despite these contributions, challenges remain. The computational complexity of signal decomposition and DL components may limit scalability in resource-constrained environments. Additionally, MC Dropout primarily captures epistemic uncertainty, which may not fully represent aleatory variability during extreme or rare events. Future work will focus on enhancing both adaptability and robustness, including the development of online and incremental learning strategies for real-time deployment, extension to renewable generation and integrated multi-energy systems, incorporation of advanced uncertainty quantification methods such as Bayesian DL and conformal prediction, and improved alignment between interpretability and temporal modeling through dynamic explanation techniques.

CRediT authorship contribution statement

Musa Terkes: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alpaslan Demirci:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization. **Idriss Dagal:** Writing – original draft, Resources, Methodology, Investigation, Conceptualization. **Erdin Gokalp:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Conceptualization. **Umit Cali:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under grant no. 124E569 within the 1002-A Short-Term Research Funding Program.

Data availability

Data will be made available on request.

References

Ahmad, W., Ayub, N., Ali, T., Irfan, M., Awais, M., Shiraz, M., Glowacz, A., 2020. Towards short term electricity load forecasting using improved support vector machine and extreme learning machine. Energies 13. https://doi.org/10.3390/en13112907.

Alhussein, M., Aurangzeb, K., Haider, S.I., 2020. Hybrid CNN-LSTM model for short-term individual household load forecasting. IEEE Access. 8, 180544–180557. https://doi.org/10.1109/ACCESS.2020.3028281.

Ali, M., Prasad, R., Xiang, Y., Yaseen, Z.M., 2020. Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for

monthly rainfall forecasts. J. Hydrol. 584, 124647. https://doi.org/10.1016/j.jhydrol.2020.124647.
Alonso, A.M., Nogales, F.J., Ruiz, C., 2020. A single scalable LSTM model for short-term forecasting of massive electricity time series. Energies 13. https://doi.org/10.3390/en13205328.
Bashir, T., Haoyong, C., Tahir, M.F., Liqiang, Z., 2022. Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. Energy Rep. 8, 1678–1686. https://doi.org/10.1016/j.egyr.2021.12.067.
Bouktif, S., Fiaz, A., Ouni, A., Serhani, M.A., 2020. Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting. Energies 13. https://doi.org/10.3390/en13020391.
Butt, F.M., Hussain, L., Jafri, S.H.M., Alshahrani, H.M., Al-Wesabi, F.N., Lone, K.J., Din, E.M.T.El, Duhayyim, M.Al, 2022. Intelligence based accurate medium and long term load forecasting system. Appl. Artif. Intell. 36. https://doi.org/10.1080/08839514.2022.2088452.
Caicedo-Vivas, J.S., Alfonso-Morales, W., 2023. Short-term load forecasting using an LSTM neural network for a grid operator. Energies 16. https://doi.org/10.3390/en16237878.
Cao, W., Liu, H., Zhang, X., Zeng, Y., 2024. Residential load forecasting based on long short-term memory, considering temporal local attention. Sustainability 16. https://doi.org/10.3390/su162411252.
Cerqueira, V., Torgo, L., Mozetič, I., 2020. Evaluating time series forecasting models: an empirical study on performance estimation methods. Mach. Learn. 109, 1997–2028. https://doi.org/10.1007/s10994-020-05910-7.
Chandorkar, A., Kharbanda, A., 2024. Divergent ensemble networks: enhancing uncertainty estimation with shared representations and independent branching. Int. J. Cybern. Inform. 13, 69–78. https://doi.org/10.5121/ijci.2024.130605.
Chung, W.H., Gu, Y.H., Yoo, S.J., 2022. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. Energy 246, 123350. https://doi.org/10.1016/j.energy.2022.123350.
Cordonnier, J.-B., Loukas, A., Jaggi, M., 2020. Multi-Head Attention: Collaborate Instead of Concatenate 1–15.
Dai, Y., Zhao, P., 2020. A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. Appl. Energy 279, 115332. https://doi.org/10.1016/j.apenergy.2020.115332.
Datateknik, E.I., 2020. Estimating Prediction Intervals with Machine Learning and Monte Carlo Methods in Online Advertising.
Dey, R., Mathur, R., 2023. Ensemble Learning Method Using Stacking with Base Learner, A Comparison. In: Chaki, N., Roy, N.D., Debnath, P., Saeed, K. (Eds.), Proceedings of International Conference on Data Analytics and Insights, ICDAI 2023. Springer Nature Singapore, Singapore, pp. 159–169.
Du, Y., Pei, B., Zhao, X., Ji, J., 2020. Deep scaled dot-product attention based domain adaptation model for biomedical question answering. Methods 173, 69–74. https://doi.org/10.1016/j.ymeth.2019.06.024.
Dudek, G., Pełka, P., Smyl, S., 2022. A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting. IEEE Trans. Neural Netw. Learn. Syst. 33, 2879–2891. https://doi.org/10.1109/TNNLS.2020.3046629.
Faraji, J., Ketabi, A., Hashemi-Dezaki, H., Shafie-Khah, M., Catalão, J.P.S., 2020. Optimal day-ahead self-scheduling and operation of prosumer microgrids using hybrid machine learning-based weather and load forecasting. IEEE Access. 8, 157284–157305. https://doi.org/10.1109/ACCESS.2020.3019562.
Gasparin, A., Lukovic, S., Alippi, C., 2022. Deep learning for time series forecasting: the electric load case. CAAI Trans. Intell. Technol. 7, 1–25. https://doi.org/10.1049/cit2.12060.
Ge, J., Niu, T., Xu, D., Yin, G., Wang, Y., 2020. A rolling bearing fault diagnosis method based on EEMD-WSST signal reconstruction and multi-scale entropy. Entropy 22. https://doi.org/10.3390/e22030290.
Giraud-Carrier, C., 2022. Combining Base-Learners into Ensembles. In: Metalearning: Applications to Automated Machine Learning and Data Mining. Springer International Publishing, Cham, pp. 169–188. https://doi.org/10.1007/978-3-030-67024-5_9.
Hafeez, G., Alimgeer, K.S., Khan, I., 2020. Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. Appl. Energy 269, 114915. https://doi.org/10.1016/j.apenergy.2020.114915.
Hebrail, G., Berard, A., 2006. Individual Household Electric Power Consumption.
Hua, K., Wang, Z., Zuo, F., Wang, Q., 2024. Short-term electricity load forecasting method based on attention and time series forecasting. Front. Artif. Intell. Appl. 382, 872–880. https://doi.org/10.3233/FAIA231386.
Hussain, A., Eswarakrishnan, V., Aslam, A., Tripura, S., 2025. Charging stations demand forecasting using LSTM based hybrid transformer model. Sci. Rep. 15, 36639. https://doi.org/10.1038/s41598-025-20421-y.
Khan, F.A., Pal, N., Saeed, S.H., 2021. Optimization and sizing of SPV/Wind hybrid renewable energy system: a techno-economic and social perspective. Energy 233, 121114. https://doi.org/10.1016/j.energy.2021.121114.
Khan, Z.A., Ullah, A., Ullah, W., Rho, S., Lee, M., Baik, S.W., 2020. Electrical energy prediction in residential buildings for short-term horizons using hybrid deep learning strategy. Appl. Sci. 10. https://doi.org/10.3390/app10238634.
Lee, J., Cho, Y., 2022. National-scale electricity peak load forecasting: traditional, machine learning, or hybrid model? Energy 239, 122366. https://doi.org/10.1016/j.energy.2021.122366.
Leon-Medina, J.X., Fonseca Gonzalez, J.E., Callejas Rodriguez, N.Y., González Niño, M.E., Hernández Moreno, S.A., Pineda-Munoz, W.A., Siachoque Celys, C.P., Umbarila Suarez, B., Pozo, F., 2025. Forecasting energy demand in quicklime manufacturing: a data-driven approach. Sensors 25. https://doi.org/10.3390/s25247632.
Li, K., Huang, W., Hu, G., Li, J., 2023. Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. Energy Build. 279, 112666. https://doi.org/10.1016/j.enbuild.2022.112666.
Li, H., Li, S., Wu, Y., Xiao, Y., Pan, Z., Liu, M., 2024. Short-term power load forecasting for integrated energy system based on a residual and attentive LSTM-TCN hybrid network. Front. Energy Res. 12, 2024. https://doi.org/10.3389/fenrg.2024.1384142.
Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions.
Lv, L., Wu, Z., Zhang, J., Zhang, L., Tan, Z., Tian, Z., 2022. A VMD and LSTM based hybrid model of load forecasting for power grid security. IEEE Trans. Ind. Inform. 18, 6474–6482. https://doi.org/10.1109/TII.2021.3130237.
Massaoudi, M., Refaat, S.S., Chihi, I., Trabelsi, M., Oueslati, F.S., Abu-Rub, H., 2021. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. Energy 214, 118874. https://doi.org/10.1016/j.energy.2020.118874.
Mehr, M.M., Farzin, H., Mashhour, E., 2026. Short-term residential load forecasting with FG-LSTM-TA: a hybrid deep learning approach integrating feature gating and temporal attention for enhanced prediction accuracy. Eng. Rep. 8, e70570. https://doi.org/10.1002/eng2.70570.
Moradzadeh, A., Zakeri, S., Shoaran, M., Mohammadi-Ivatloo, B., Mohammadi, F., 2020. Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms. Sustainability 12. https://doi.org/10.3390/su12177076.
Mounir, N., Ouadi, H., Jrhilifa, I., 2023. Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system. Energy Build. 288, 113022. https://doi.org/10.1016/j.enbuild.2023.113022.
Nasirtafreshi, I., 2022. Forecasting cryptocurrency prices using recurrent neural network and long short-term memory. Data Knowl. Eng. 139, 102009. https://doi.org/10.1016/j.datak.2022.102009.
Nie, Y., Jiang, P., Zhang, H., 2020. A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting. Appl. Soft Comput. 97, 106809. https://doi.org/10.1016/j.asoc.2020.106809.
PiML Toolbox Developers, 2023. SHAP Explainability Guide - PiML Toolbox.
Rafi, S.H., Nahid-Al-Masood, Deeba, S.R., Hossain, E., 2021. A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. IEEE Access. 9, 32436–32448. https://doi.org/10.1109/ACCESS.2021.3060654.
Ramos, S., Soares, J., Foroozandeh, Z., Tavares, I., Vale, Z., 2021. Energy consumption and PV generation data of 15 prosumers (15 min resolution). https://doi.org/10.5281/zenodo.5106455.
Ran, P., Dong, K., Liu, X., Wang, J., 2023. Short-term load forecasting based on CEEMDAN and Transformer. Electr. Power Syst. Res. 214, 108885. https://doi.org/10.1016/j.epsr.2022.108885.
Sajjad, M., Khan, Z.A., Ullah, A., Hussain, T., Ullah, W., Lee, M.Y., Baik, S.W., 2020. A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. IEEE Access. 8, 143759–143768. https://doi.org/10.1109/ACCESS.2020.3009537.
Shah, S.A.H., Ahmed, U., Bilal, M., Khan, A.R., Razzaq, S., Aziz, I., Mahmood, A., 2025. Improved electric load forecasting using quantile long short-term memory network with dual attention mechanism. Energy Rep. 13, 2343–2353. https://doi.org/10.1016/j.egyr.2025.01.058.
Shen, Z., Zhang, Y., Lu, J., Xu, J., Xiao, G., 2020. A novel time series forecasting model with deep learning. Neurocomputing 396, 302–313. https://doi.org/10.1016/j.neucom.2018.12.084.
Shohan, M.J.A., Faruque, M.O., Foo, S.Y., 2022. Forecasting of electric load using a hybrid LSTM-neural prophet model. Energies 15. https://doi.org/10.3390/en15062158.
Sideratos, G., Ikonomopoulos, A., Hatziargyriou, N.D., 2020. A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks. Electr. Power Syst. Res. 178, 106025. https://doi.org/10.1016/j.epsr.2019.106025.
Sohn, G., Zhang, N., Olukotun, K., n.d. Implementing and Optimizing the Scaled Dot-Product Attention on Streaming Dataflow 1–4.
Song, J., Zhang, L., Xue, G., Ma, Y., Gao, S., Jiang, Q., 2021. Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model. Energy Build. 243, 110998. https://doi.org/10.1016/j.enbuild.2021.110998.
Syed, D., Abu-Rub, H., Ghrayeb, A., Refaat, S.S., 2021. Household-Level Energy Forecasting in Smart Buildings Using a Novel Hybrid Deep Learning Model. IEEE Access. 9, 33498–33511. https://doi.org/10.1109/ACCESS.2021.3061370.
Tao, Z., Xu, Q., Liu, X., Liu, J., 2023. An integrated approach implementing sliding window and DTW distance for time series forecasting tasks. Appl. Intell. 53, 20614–20625. https://doi.org/10.1007/s10489-023-04590-9.
Tarmanini, C., Sarma, N., Gezegin, C., Ozgonenel, O., 2023. Short term load forecasting based on ARIMA and ANN approaches. Energy Rep. 9, 550–557. https://doi.org/10.1016/j.egyr.2023.01.060.
Tayab, U.B., Zia, A., Yang, F., Lu, J., Kashif, M., 2020. Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform. Energy 203, 117857. https://doi.org/10.1016/j.energy.2020.117857.
Wan, A., Chang, Q., AL-Bukhaiti, K., He, J., 2023. Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. Energy 282, 128274. https://doi.org/10.1016/j.energy.2023.128274.
Wang, L., Shao, Y., 2020. Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis. Mech. Syst. Signal. Process. 138, 106545. https://doi.org/10.1016/j.ymssp.2019.106545.
Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., Guo, Y., Wang, T., 2021. Short-term load forecasting of industrial customers based on SVMD and XGBoost. Int. J.

Electr. Power Energy Syst. 129, 106830. https://doi.org/10.1016/j.ijepes.2021.106830.
Weckesser, T., Yousefi, G., Ziras, C., Thingvad, M., Odgaard, J.H., 2024. Danish Residential Electric Vehicle Charging Profiles for Distribution Grid Planning and Design. https://doi.org/10.11583/DTU.25348213.v1.
Wu, L., Kong, C., Hao, X., Chen, W., 2020. A short-term load forecasting method based on GRU-CNN hybrid neural network model. Math. Probl. Eng. 2020, 1–10. https://doi.org/10.1155/2020/1428104.
Wu, Y., Sun, W., Li, Q., 2025. Power load forecasting and anomaly detection using a two-stage attention mechanism and deep neural networks. Electr. Power Syst. Res. 249, 112056. https://doi.org/10.1016/j.epsr.2025.112056.
Xu, Y., Gao, W., Qian, F., Li, Y., 2021. Potential analysis of the attention-based LSTM model in ultra-short-term forecasting of building HVAC energy consumption. Front. Energy Res. 9, 2021. https://doi.org/10.3389/fenrg.2021.730640.
Xuan, W., Shouxiang, W., Qianyu, Z., Shaomin, W., Liwei, F., 2021. A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. Int. J. Electr. Power Energy Syst. 126, 106583. https://doi.org/10.1016/j.ijepes.2020.106583.
Yamasaki, M., Freire, R.Z., Seman, L.O., Stefenon, S.F., Mariani, V.C., dos Santos Coelho, L., 2024. Optimized hybrid ensemble learning approaches applied to very short-term load forecasting. Int. J. Electr. Power Energy Syst. 155, 109579. https://doi.org/10.1016/j.ijepes.2023.109579.
Zhang, Z., Hong, W.-C., Li, J., 2020. Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm. IEEE Access. 8, 14642–14658. https://doi.org/10.1109/ACCESS.2020.2966712.
Zhang, C., Li, J., Zhao, Y., Li, T., Chen, Q., Zhang, X., 2020. A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process. Energy Build. 225, 110301. https://doi.org/10.1016/j.enbuild.2020.110301.
Zhang, J., Xing, L., Tan, Z., Wang, H., Wang, K., 2022. Multi-head attention fusion networks for multi-modal speech emotion recognition. Comput. Ind. Eng. 168, 108078. https://doi.org/10.1016/j.cie.2022.108078.
Zheng, Z., Inc, P.S., 2016. Pecan Street Electricity Data.
Zhong, B., 2023. Deep learning integration optimization of electric energy load forecasting and market price based on the ANN–LSTM–transformer method. Front. Energy Res. 11, 2023. https://doi.org/10.3389/fenrg.2023.1292204.
Zhou, B., Wang, H., Xie, Y., Li, G., Yang, D., Hu, B., 2024. Regional short-term load forecasting method based on power load characteristics of different industries. Sustain. Energy Grids Netw. 38, 101336. https://doi.org/10.1016/j.segan.2024.101336.