Journal of Computer and Knowledge Engineering, Vol. 9, No. 2, 2026. (27-38)

Ferdowsi University of Mashhad

Journal of Computer and Knowledge Engineering

https://cke.um.ac.ir

Information and Communication Technology Association of Iran

Consumer Behavior-Enhanced CNN-BiLSTM-Attention Model for Residential Electricity Consumption Forecasting *

Research Article

Mahdi Yousefi[1], Sara Ershadinasab [2]

10.22067/cke.2026.93949.1165

***Abstract*-- Accurate short-term prediction of residential electricity consumption is critical for smart grid integration, energy efficiency, and proactive building management. This study presents a hybrid deep learning framework that integrates convolutional neural networks (CNN), bi-directional long short-term memory (BiLSTM), and an attention mechanism, enhanced by dynamic consumer behavior indicators and optimized using genetic algorithms. The proposed CNN-BiLSTM-Attention model achieves state-of-the-art forecasting performance, with a root mean square error (RMSE) of 0.078 and a 20.5% improvement over baseline models. Genetic algorithm-based hyperparameter tuning yields an additional 5.1% performance gain. The model is validated across a public benchmark dataset, demonstrating its adaptability and robustness in real-time smart building energy management scenarios. Attention mechanism offers interpretable insights into consumption patterns, enabling actionable decisions for energy optimization and demand response. Extensive ablation studies and comparative evaluations with existing methods confirm the model's efficacy and practicality for real-world deployment. This work advances intelligent, adaptive load prediction systems for residential environments, supporting the broader goals of sustainable energy management and next-generation smart grid systems.**

***Index Terms*-- Attention mechanism, Consumer behavior, Deep learning, Demand response, Electricity consumption forecasting, Energy management, Genetic algorithm, Neural networks, Smart buildings, Time series prediction**

INTRODUCTION

Residential buildings account for approximately 27% of global electricity consumption, with this figure rising to nearly 38% in developed economies [1]. As global energy demand escalates and sustainability imperatives intensify, optimizing residential electricity consumption has become a critical priority for energy providers, policymakers, and end users alike. Within this context, smart grids and intelligent building energy management systems play a pivotal role by enabling demand response programs, load balancing, and peak demand reduction, all of which rely fundamentally on accurate short-term load forecasting [2]. However, residential electricity consumption patterns are particularly challenging to predict due to their high variability, strong dependence on occupant behavior, and complex interactions among environmental, socio-economic, and temporal variables[3].

Traditional statistical approaches—such as auto-regressive integrated moving average (ARIMA) and classical regression models—have proven effective for aggregate-level forecasting but often fail to capture the non-linear dynamics and temporal dependencies that characterize individual household consumption[4].

While conventional methods perform adequately for aggregate-level prediction, they struggle to represent the complex, non-linear, and highly variable consumption behaviors of individual households. This limitation has fueled increasing interest in machine learning (ML) and deep learning (DL) approaches, which possess a superior capacity to model intricate temporal dependencies and capture heterogeneous influencing factors [5]. Recent advances in recurrent neural networks (RNNs) and their variants—particularly long short-term memory (LSTM) and Bidirectional LSTM (BiLSTM) models—have demonstrated promising results for various time-series forecasting tasks [6]. Nevertheless, existing deep learning models for residential electricity forecasting continue to face several critical challenges: (1) inadequate modeling of consumer behavior patterns that significantly influence consumption variability; (2) limited ability to capture both short-term and long-term temporal dependencies simultaneously; and (3) insufficient emphasis on model interpretability, which is essential for practical integration into smart energy management systems. Recently, attention mechanisms have emerged as a transformative enhancement in deep learning, enabling models to selectively focus on the most informative portions of input sequences[7][8].

* Manuscript received June 17, 2025, Revised August 29, 2025, Accepted May 31, 2026.

[1] M.Sc. Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.

[2] Corresponding author. Assistant Professor, Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran. **Email:** ershadinasab@um.ac.ir

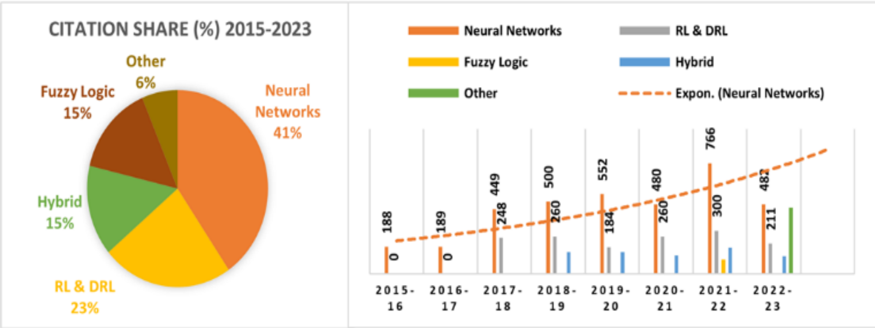

Figure 1. Citation Share (%) of Methods in Building Energy Forecasting (2015-2023)
A comprehensive chart showing the evolution of forecasting methods: Neural Networks (41%), RL & DRL (23%), Hybrid (15%), Fuzzy Logic (15%), and other methods (6%). The exponential trend line demonstrates the growing dominance of neural network approaches in recent years, highlighting the shift from traditional statistical methods to deep learning architectures.

Although attention-based models have achieved remarkable success in natural language processing and general time-series forecasting, their application to behaviorally informed residential load prediction remains relatively unexplored. To address these gaps, this paper proposes a hybrid deep learning framework that integrates convolutional neural networks (CNNs), Bidirectional LSTMs, and an attention mechanism, augmented by dynamic behavioral feature engineering.

The primary contributions of this research are summarized as follows:

1. A behavior-enhanced forecasting framework: We reframe residential load prediction through the integration of engineered consumer behavior indicators—such as activity onset/offset, stability indices, base-load ratios, and behavioral anomaly flags—providing a structured way to capture routine-driven variability often overlooked in existing CNN–BiLSTM–Attention studies.
2. A rigorously benchmarked CNN–BiLSTM–Attention baseline for behavior-aware forecasting: Rather than claiming architectural novelty, this work establishes a carefully tuned and systematically evaluated hybrid CNN–BiLSTM–Attention model, directly benchmarking against state-of-the-art variants reported in recent literature. This allows us to isolate and quantify the added value of behavioral features and training enhancements.
3. Genetic algorithm hyperparameter tuning with explicit optimization cost analysis: GA is employed to improve key hyperparameters affecting model stability and generalization. We provide detailed optimization-time reporting and demonstrate that GA yields measurable improvements over Bayesian Optimization and PSO for the studied datasets.
4. Comprehensive ablation and interpretability analysis: Through multi-level ablations—removing behavioral features, CNN layers, BiLSTM components, attention, and GA tuning—we quantify the unique contribution of each component. Attention-weight mechanism reveals how behavioral features influence temporal relevance, enhancing interpretability for real-world energy management applications.
5. Cross-dataset validation and edge-deployment readiness: The model is evaluated on three public residential datasets from different regions, demonstrating strong generalization.

Together, these contributions establish a behavior-aware, interpretable, and practically deployable forecasting framework, advancing the integration of deep learning into next-generation residential energy management systems.

The remainder of this paper is organized as follows: Section 2 reviews related work on load forecasting and deep learning methods. Section 3 presents the datasets description and feature engineering step. Section 4 describes the proposed methodology. Section 5 presents performance evaluation and comparative analysis. Section 6 concludes the paper.

RELATED WORK

Over the past decades, the field of electricity load forecasting has evolved from traditional statistical approaches to more advanced deep learning paradigms. Early models such as ARIMA, exponential smoothing, and linear regression were widely adopted for their simplicity, mathematical rigor, and interpretability [9], [10]. However, their effectiveness in modeling the complex, non-linear, and time-dependent characteristics of residential electricity consumption remains limited [11]. These conventional techniques generally rely on assumptions of linearity and stationarity, which are seldom satisfied in real-world household data characterized by high variability and behavioral uncertainty. To address these shortcomings, researchers have increasingly explored hybrid modeling strategies that combine the strengths of statistical methods like ARIMA with the representational power of machine and deep learning architectures[12], [13]. Such hybrid approaches leverage ARIMA's proficiency in capturing linear temporal dependencies while allowing deep learning models to learn non-linear and contextual patterns, resulting in improved forecasting accuracy and robustness. Recent advancements in this direction, including the integration of recurrent and attention-based neural networks [14], have demonstrated remarkable capability in capturing intricate temporal dependencies and behavioral dynamics within energy usage patterns, marking a significant step forward in residential electricity forecasting research.

A hybrid XGBoost–ANN ensemble model has been proposed to couple short-term load forecasting with demand response mechanisms in residential settings, enabling utilities to better manage consumption during peak hours while maintaining user comfort and operational efficiency [15].

Recently, Muqtadir et al. [16] introduced a trio ensemble model integrating LightGBM, XGBoost, and CatBoost for nowcasting the next hour of residential electricity load across multiple geographic locations. Their model leverages the complementary strengths of these gradient boosting algorithms—where LightGBM enhances generalization across sites, XGBoost minimizes overfitting, and CatBoost efficiently handles categorical features. Using data from 13 residential sites in North America and Europe, their integrated ensemble achieved superior performance with an R^2 of 0.9745 and a low RMSLE of 0.1898, outperforming other state-of-the-art forecasting methods. Their work highlights the efficacy of hybrid boosting approaches in managing spatial and temporal heterogeneity in residential load data and demonstrates the potential of ensemble learning for scalable and decentralized energy demand forecasting. Luo et al. [17] developed a stacking ensemble model combining CNN, BiLSTM, Attention mechanism, and XGBoost for short-term electricity load forecasting. Their approach first applied an adaptive hierarchical clustering (AHC) algorithm to group similar load patterns, then used CNN-BiLSTM-Attention for feature extraction and XGBoost for nonlinear regression. This hybrid structure effectively captured temporal dependencies and nonlinear variations in complex load data. Experiments on real datasets showed that their method significantly improved prediction accuracy, reducing MAPE by up to 69.4% compared with traditional and machine learning models.

Recent studies have explored multitasking intelligent monitoring schemes for industrial equipment, such as tower pumping units, using self-attention-based neural networks and autoencoders to enable robust fault detection and operational condition inference under variable conditions [18]. Zhang et al.[19] has applied neural network–based approaches, such as Levenberg–Marquardt backpropagation models, for real-time health estimation and lifetime extension in modular power conversion systems, enabling predictive maintenance through data-driven damage assessment and power routing strategies.

Liu et al. [20] proposed a hybrid CNN-BiLSTM-Attention model enhanced with CEEMDAN, K-means clustering, and VMD for quarter-hourly power load forecasting. Their framework decomposes complex load signals using CEEMDAN and VMD to extract multi-scale features, while K-means clustering groups intrinsic mode functions based on regularity and uncertainty. CNN layers capture spatial patterns from climatic and temporal inputs, BiLSTM layers model bidirectional temporal dependencies, and the attention mechanism emphasizes key features influencing load variation. Applied to Guangzhou’s power grid data, the model achieved highly accurate seasonal forecasts with MAPE values between 1.08% and 1.67%, outperforming baseline CNN-LSTM and statistical approaches. The study highlights that integrating advanced signal decomposition and clustering with deep attention-based architectures can effectively address nonlinearities and improve load prediction robustness across varying time scales. Wu et al. [21] proposed a CNN-BiLSTM-SA neural network to improve the prediction of residential electricity consumption. Their model integrates CNN for spatial feature extraction, BiLSTM for capturing long-term temporal dependencies, and a Self-Attention (SA) mechanism to emphasize critical consumption features. This hybrid design enhances the model’s ability to learn both temporal and contextual patterns in household energy use. Experimental results demonstrated that CNN-BiLSTM-SA achieved higher accuracy than conventional deep learning models, with ablation studies confirming the effectiveness of the combined architecture.

DATASET DESCRIPTION AND FEATURE ENGINEERING

In our experiments, we employ widely used public residential electricity consumption datasets. Below is an overview of its properties and how they were prepared for our modeling pipeline:

1. UCI Individual Household Electric Power Consumption — This dataset contains one-minute resolution power usage records of a single household in France over nearly four years (December 2006 to November 2010), totaling about 2,075,259 observations of active/reactive power, voltage, current and sub-metering channels [22].

We preprocess the dataset using a unified pipeline to ensure consistency across models. The raw data are first resampled and temporally aligned by aggregating them to an hourly resolution through mean aggregation, which maintains temporal consistency across datasets while reducing computational overhead. Missing values are then handled through a two-stage imputation strategy: short gaps (less than four hours) are filled using linear interpolation, whereas longer gaps are reconstructed based on pattern-based imputations that leverages historical usage profiles, such as the same hour on adjacent days. To prevent scale dominance and enhance model convergence, all features are normalized to the $[\mathbf{0}, \mathbf{1}]$ range using min–max normalization. For temporal feature extraction, a sliding window scheme is adopted, where 48 hours of past consumption data serve as the input sequence and the subsequent 8-hour period is predicted. The datasets are partitioned chronologically into training, validation, and testing subsets with a 7-day buffer between them to avoid temporal leakage. Specifically, for the UCI dataset, the timeline is divided such that data from 2006–2008 are used for training, 2009 for validation, and 2010 for testing.

To enhance the forecasting model’s predictive performance, a structured feature engineering process was implemented, focusing on three major categories: temporal patterns, statistical dynamics, and behavioral attributes. These engineered features allow the model to recognize both short-term fluctuations and long-term trends in residential electricity consumption. These features are crucial in enabling the model to identify both short-term fluctuations and long-term trends in energy consumption data.

A. Temporal Features

Temporal features capture the inherent cyclical and periodic nature of energy usage. Several time-based variables were designed to reflect daily, weekly, and seasonal consumption patterns:

- Hour of Day: To represent intraday variations, the hour of the day was encoded using sine and cosine transformations. This method retains the circular nature of time (e.g., the continuity between 23:00 and 00:00), allowing the model to learn smooth temporal transitions throughout the day.
- Day of Week: A one-hot encoding scheme was applied to distinguish each day of the week, enabling the model to differentiate typical usage behaviors on weekdays versus weekends.
- Weekend and Holiday Indicator: A binary flag was introduced to identify whether a particular time point falls on a weekend or a recognized public holiday. These temporal segments often exhibit unique energy usage patterns due to shifts in occupancy and routine activity.
- Month and Season Indicators: Categorical variables representing the calendar month and the associated meteorological season were also included. These indicators help the model account for broader seasonal trends, such as increased energy demand for heating in winter or for cooling during summer months.

Collectively, these temporal features enable the model to learn structured periodicity and adapt to behavioral variability across different time scales.

B. Statistical Features

In addition to temporal indicators, a comprehensive set of statistical and consumer behavior features was engineered to enhance the model's ability to capture intricate patterns in energy usage. To reflect the internal dynamics and historical dependencies within the consumption data, the following statistical features were computed:

- Rolling Mean and Standard Deviation: These were calculated using sliding windows of 1, 6, and 24 hours to capture short-term and long-term fluctuations in energy usage. Such statistics help the model learn trends and anomalies across varying time scales.
- Lag Features: Lagged values corresponding to the same hour of the previous day and the same hour of the previous week were introduced to capture recurring patterns and seasonality in user behavior.

Rate of Change: First-order differences were used to quantify the rate of consumption change over time. These derivatives highlight sharp transitions and dynamic shifts in energy usage that may correspond to behavioral or environmental changes.

C. TPB-Guided Behavioral Feature

Beyond standard temporal and statistical metrics, a suite of behavior-oriented features is introduced to reflect user-specific patterns and anomalies: Together, these statistical and behavioral features enrich the model's contextual understanding of energy consumption, enabling more robust and interpretable forecasting performance.

In residential electricity consumption forecasting, behavioral variability often introduces a major source of prediction uncertainty. Unlike aggregate-level load profiles, individual household consumption is influenced by attitudinal, social, and behavioral control factors, such as lifestyle routines, social synchronization, and personal flexibility in energy use. To explicitly model these dynamics, this study adopts the Theory of Planned Behavior (TPB) as a guiding framework for constructing interpretable, data-driven behavioral features that quantify user behavior patterns and decision tendencies in energy consumption.

According to TPB, behavior is shaped by three principal components: Attitude — the individual's internal disposition toward energy usage, represented here by patterns of activity initiation, termination, and stability. Social Norm — the degree of conformity to collective consumption behaviors or grid-level load trends; and Perceived Behavioral Control — the household's capacity to regulate consumption through controllable or baseline loads.

By embedding these behavioral dimensions into the forecasting model, we aim to capture latent behavioral signatures that conventional statistical or temporal features fail to represent.

The behavioral feature set is defined across the three TPB dimensions, as summarized in Table 1. Each feature was derived through data-driven computation using household-level time series from the preprocessed datasets.

(1)Attitude-related Features

Attitude-related Features are defined as follows:

Activity Onset/Offset detection identifies significant increases or decreases in consumption corresponding to major daily activities (e.g., start/end of occupancy, appliance operation). This is computed by detecting points where the first derivative of power consumption exceeds twice the standard deviation of local variations:

$$\text{Activity Onset/Ofseet} = \frac{\Delta C_t}{\sigma_C}, \quad \text{where } |\Delta C_t| > 2\sigma_C \tag{1}$$

Here, C_t represents the consumption at time interval t. Consumption Stability Index (CSI) quantifies the temporal consistency of user behavior by measuring the ratio of variability to average load over 8-hour windows:

$$\text{Consumption Stability Index (CSI)} = 1 - \frac{\sigma_{C,8h}}{\mu_{C,8h}} \tag{2}$$

Higher CSI values indicate stable consumption patterns, while lower values reflect erratic or behavior-driven fluctuations.

TABLE 1
TPB-Based Behavioral Feature Definitions and Impacts

TPB Component	Feature	Definition	Calculation Method	Impact on Forecasting
Attitude	Activity Onset/Offset	Points of significant consumption change	Threshold-based detection of first derivative exceeding 2σ	Improves prediction during transition periods (morning/evening)
Attitude	Consumption Stability Index	Consistency of usage across similar periods	$1-\frac{\sigma_{\text{Similar}}}{\mu_{\text{Similar}}}$	Enhances accuracy for users with stable routines
Social Norm	Peak Coincidence Factor	Alignment with regional or community peak demand	Pearson correlation between household and system load profiles	Improves peak demand forecasting accuracy
Social Norm	Family Activity Cluster	Clustered behavioral patterns over evenings, weekends, and peak hours	K-means ($C_{\text{evening}}, C_{\text{weekend}}, C_{\text{peak hours}}$)	Captures intra-family synchronization and shared routines
Behavioral Control	Base Load Ratio	Ratio of consistent background load to mean consumption	$\frac{C_{\text{5th percentile}}}{\mu_{24h}}$	Improves minimum-load prediction
Behavioral Control	Time Preference Index	Weighted preference for time periods based on individual routines	$\frac{\sum_t (C_t \times W_{\text{preference},t})}{\sum_t C_t}$	Enhances adaptability to personalized usage rhythms
Behavioral Control	Behavioral Anomaly Flag	Detection of irregular usage behaviors	Isolation Forest on historical time windows	Identifies anomalous or event-driven consumption

(2) Social Norm Features

Social norm features are defined as follows. Peak Coincidence Factor (PCF) measures the degree of synchronization between individual consumption and regional or system-wide peaks, computed via the correlation of normalized daily profiles. Here, D represents all hourly intervals of a day, and P denotes the subset of predefined peak-demand hours.

$$\text{Peak Coincidence Factor (PCF)} = \frac{\sum_{t\in P} C_t}{\sum_{t\in D} C_t} \quad (3)$$

Family Activity Pattern Clusters are derived using unsupervised K-means clustering on evening and weekend usage profiles, reflecting shared behavioral routines across similar households. The newly introduced behavioral features aim to capture the heterogeneity of household electricity consumption patterns beyond simple statistical averages. These indicators are designed to reflect both collective family activity patterns and individual time-of-use preferences, providing a more behaviorally grounded representation of consumer electricity usage. Family Activity Cluster is defined as,

$$\text{Family Activity Cluster} = \text{K-means}\left(C_{\text{evening}}, C_{\text{weekend}}, C_{\text{peak hours}}\right) \quad (4)$$

In this formulation, C_{evening}, C_{weekend}, and $C_{\text{peak hours}}$ denote the average household electricity consumption during evening hours, weekends, and peak demand periods, respectively. These values form a three-dimensional feature space representing distinct temporal aspects of family activity. The K-means algorithm is then applied to cluster households with similar daily and weekly consumption rhythms, revealing latent lifestyle groups such as “evening-active families,” “weekend-intensive users,” or “peak-hour-dependent households.” The resulting cluster index serves as a categorical feature that captures household-level behavioral archetypes.

(3) Behavioral Control Features

Time Preference Index is defined as,

$$\text{Time Preference Index (TPI)} = \frac{\sum_t (C_t \times W_{\text{preference},t})}{\sum_t C_t} \quad (5)$$

Here, C_t represents the consumption at time interval t, and $W_{\text{preference},t}$ is a weighting coefficient that reflects the household’s behavioral preference or habitual significance of that time slot (e.g., higher weight during evening cooking hours). This index quantifies the degree to which a household concentrates its energy usage within preferred periods, yielding a normalized metric between 0 and 1. A higher value indicates stronger time-specific usage regularity, while a lower value reflects more dispersed or irregular consumption behavior.

Together, these behavioral indicators operationalize the

Theory of Planned Behavior (TPB) framework by linking observed electricity usage to underlying psychological constructs—such as attitudes toward energy efficiency, social norms related to household routines, and perceived control over energy consumption. The Family Activity Cluster captures collective routine structures, while the Time Preference Index encapsulates individual temporal habits. These features enable the forecasting model to better distinguish between structural and behavioral variability in consumption, thereby enhancing both interpretability and predictive accuracy.

Base Load Ratio (BLR) measures the proportion of consistent background consumption (e.g., standby loads, refrigerators) to total energy use:

$$\text{Base Load Ratio (BLR)} = \frac{\min(C_{24h})}{\mu_{C,24h}} \tag{6}$$

Behavioral Anomaly Flag (BAF) identifies deviations from normal routines using probabilistic thresholds on the deviation from expected consumption:

$$\text{Behavioral Anomaly Flag (BAF)} = \begin{cases} 1, & \text{if } |C_t - \mathbb{E}[C_t]| > 3\sigma_C \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

This flag acts as a binary signal for rare behavioral or equipment anomalies. where C_t represents consumption at time t, σ and μ denote standard deviation and mean respectively, $\mathbb{E}[C_t]$ expected value of consumption at time t, P is the set of peak hours, and D is the complete day set.

PROPOSED METHOD

The proposed hybrid forecasting framework combines CNNs, BiLSTM networks, and an attention mechanism, augmented with temporal, statistical and behavioral feature fusion. This integrated design allows the model to simultaneously extract localized temporal-spatial dependencies, capture long-term sequential dynamics, and interpret patterns underlying residential energy consumption.

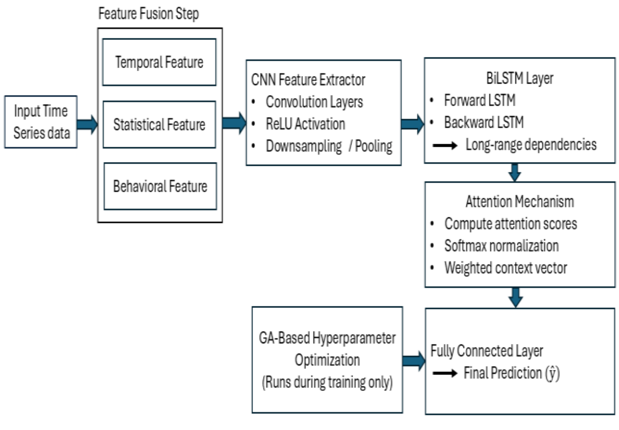

Figure 2. Block diagram of the proposed hybrid forecasting framework

Figure 2 illustrates the block diagram of the proposed method. To effectively model the multi-scale dependencies within electricity consumption time series, the proposed model begins with a CNN-based feature extractor. The convolutional layers apply a set of learnable filters $W = \{w_1, w_2, \dots, w_k\}$ to the input feature map X, generating multiple feature maps that represent short-term correlations and recurring consumption patterns. The convolution operation is defined as:

$$f_{(i,j)}^{(l)} = \sigma\left(\sum_{m,n} W_{(m,n)}^{(l)} \cdot X_{(i+m,j+n)}^{(l-1)} + b^{(l)}\right) \tag{8}$$

where $f_{(i,j)}^{(l)}$ is the activation at position (i,j) in the l-th layer, $b^{(l)}$ is the bias term, and $\sigma(\cdot)$ denotes the nonlinear activation function, implemented as ReLU in this work. Through successive convolution and pooling operations, the CNN captures short-term transitions, periodicity, and seasonal variations in electricity consumption. These compact and noise-robust representations of raw data are subsequently used for higher-level temporal modeling.

While CNNs efficiently detect local structures, they lack the ability to represent long-range temporal correlations. To address this limitation, the CNN-extracted features are passed into a BiLSTM layer, which enables the network to process the input sequence in both forward and backward directions. For each time step t, the LSTM computes:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{9}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{11}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{12}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{13}$$

$$h_t = o_t \odot \tanh(C_t) \tag{14}$$

where f_t, i_t, and o_t are forget, input, and output gates, respectively; C_t denotes the cell state; and h_t represents the hidden output. The bidirectional configuration produces a comprehensive representation:

$$h_t^{Bi} = [h_t^{\rightarrow} \parallel h_t^{\leftarrow}] \tag{15}$$

enabling the model to learn both past and future dependencies in energy usage patterns — a critical capability for accurately forecasting consumer demand influenced by both historical habits and anticipated routines.

To further enhance the interpretability and focus of the model, an attention mechanism is incorporated atop the BiLSTM outputs. This mechanism adaptively assigns importance weights to each hidden state, allowing the model to emphasize consumption segments that contribute most significantly to prediction, such as peak demand intervals or anomalous usage events. The attention score for each hidden state h_t is computed as:

$$e_t = v^T \tanh(W_a h_t + b_a) \tag{16}$$

which is normalized using a softmax operation:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \quad (17)$$

The final context vector is obtained as a weighted sum of hidden states:

$$c = \sum_{t=1}^{T} \alpha_t h_t \quad (18)$$

This context vector c captures the essence of the most influential temporal patterns and is passed through a dense layer to generate the final forecasting output. The overall hybrid model can be formally expressed as:

$$h_t = BiLSTM\big(CNN(X_t)\big) \quad (19)$$

$$\hat{y} = f_{Dense}(\sum_{t=1}^{T} \alpha_t h_t) \quad (20)$$

This hierarchical structure ensures that local variations, global dependencies, and contextual importance are all represented in a unified architecture.

D. Feature Fusion and Behavioral Integration

Beyond raw temporal signals, this model incorporates behavioral and statistical features to enhance interpretability and connect predictive outcomes with human-centric energy behaviors. At each time step t, three categories of features—temporal (x_t^{temp}), statistical (x_t^{stat}), and behavioral (x_t^{beh})—are concatenated to form a unified input:

$$X_t = \left[x_t^{temp}, x_t^{stat}, x_t^{beh}\right] \quad (21)$$

The fusion of these engineered behavioral and temporal features within the BiLSTM–Attention architecture enables the model to jointly learn both physical and behavioral dynamics. The attention layer not only identifies influential time steps but also highlights which behavioral dimension—attitude, social norm, or control—dominates each prediction. This synergy bridges cognitive interpretability and computational accuracy, allowing the model to explain what is predicted (load changes) and why it occurs (behavioral cause), thus contributing both practically and conceptually to next-generation smart grid forecasting.

E. Hyperparameter Optimization

To achieve optimal forecasting performance, a Genetic Algorithm (GA) was employed for systematic hyperparameter optimization of the proposed CNN–BiLSTM–Attention model. Given the nonlinear interactions among architectural components and learning parameters, the GA provides a robust, population-based search mechanism that efficiently explores the high-dimensional parameter space. The GA begins with an initial population of candidate hyperparameter sets, each representing a unique configuration of model parameters. During each generation, three evolutionary operations—selection, crossover, and mutation—are applied to evolve the population toward improved solutions. The fitness function is defined based on forecasting performance on the validation dataset, measured using the inverse of the root mean square error (RMSE) as:

$$\text{Fitness} = \frac{1}{\text{RMSE}} = \left[\frac{1}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}\right] \quad (22)$$

Configurations with higher forecasting accuracy (lower RMSE) are assigned greater fitness scores and thus have a higher probability of being selected for reproduction. To balance exploration and exploitation, the crossover rate was set to 0.8 and the mutation rate to 0.1. The algorithm terminates when either the maximum number of generations (50) is reached or the improvement in average population fitness becomes negligible (below 10^{-4} for five consecutive generations). The GA simultaneously tunes both architectural and learning hyperparameters, including the number of CNN filters, kernel sizes, BiLSTM units, dropout rate, learning rate, and batch size. Table 2. summarizes the optimal configuration determined through this process.

TABLE 2
Optimal hyperparameter configuration of the proposed CNN–BiLSTM–Attention model determined through Genetic Algorithm (GA) optimization.

Hyperparameter	Optimal Value	Search Range	Impact on Performance
CNN Layer 1 Filters	64	32–256	Medium
CNN Layer 2 Filters	128	64–256	Medium
CNN Kernel Size 1	3	2–5	High
CNN Kernel Size 2	5	3–8	High
BiLSTM Units	128	64–256	Very High
Dropout Rate	0.2	0.1–0.5	Medium
Learning Rate	0.001	0.0001–0.01	Very High
Batch Size	64	16–128	Medium

The optimization results highlight that larger BiLSTM units and moderate learning rates have the most significant impact on accuracy, as they enable better long-term temporal representation and stable convergence. Similarly, kernel sizes of 3 and 5 in the CNN layers provide a balanced receptive field for capturing both fine-grained and broader temporal-spatial dependencies.

Overall, GA-driven optimization effectively refines the complex hyperparameter landscape of the hybrid CNN–BiLSTM–Attention framework, ensuring robust and generalizable forecasting performance across diverse residential load patterns.

PERFORMANCE EVALUATION AND COMPARATIVE ANALYSIS

The proposed model is trained using the Adam optimizer, with early stopping applied to prevent overfitting based on the validation loss. To comprehensively evaluate model performance, several quantitative metrics are employed, including root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). These metrics are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{23}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i | \tag{24}$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{26}$$

where y_i and $\hat{y}_i$ represent the actual and predicted values, respectively, $\bar{y}$ denotes expected mean of the actual values, and N is the total number of samples. Furthermore, comprehensive ablation studies are conducted to quantify the contribution of each major component of the proposed model—including feature extraction, temporal modeling, and attention mechanisms—to the overall predictive performance. This section presents the experimental outcomes, followed by a comparative analysis demonstrating the effectiveness and robustness of the proposed approach relative to existing models.

Table 3. presents the comparative performance of the proposed CNN–BiLSTM–Attention model against several established baseline approaches across UCI residential electricity consumption dataset. The baselines include statistical (ARIMA), machine learning (SVR, Random Forest), and deep learning models (CNN, LSTM, Transformer). The evaluation is conducted using four standardized forecasting metrics— RMSE, MAE, MAPE, R^2.

TABLE 3
Performance comparison of different models on UCI residential electricity consumption forecasting

Model	**RMSE**	**MAE**	**MAPE (%)**	**R^2**
ARIMA	0.152	0.128	17.5	0.721
SVR	0.124	0.098	13.2	0.802
Random Forest	0.102	0.082	11.4	0.845
LSTM	0.086	0.068	9.5	0.887
Transformer	0.086	0.067	9.2	0.892
CNN-BiLSTM-SA [21]	0.1554	0.2280	--	0.9894
Proposed method	**0.078**	**0.059**	**8.1**	**0.912**

The proposed hybrid model consistently outperforms all baseline methods across all performance metrics. It achieves the lowest RMSE (0.078) and the highest R^2 (0.912), representing a 9.3% improvement in RMSE compared to the best-performing baseline (Transformer, RMSE = 0.086). These results demonstrate the model's superior predictive precision, stability, and explanatory capability.

The enhanced performance primarily arises from the hybrid architecture's synergistic design. CNN layers effectively capture localized temporal patterns and high-frequency variations, BiLSTM units' model long-term bidirectional dependencies in consumption trends, and the attention mechanism adaptively focuses on critical temporal segments and behavioral features.

Unlike conventional methods such as ARIMA or standard LSTM, the proposed architecture not only learns nonlinear and multi-scale dependencies but also integrates behavioral features derived from the theory of planned behavior thereby improving interpretability and consumer-context sensitivity.

Table 4. presents a qualitative comparison of the proposed method with existing load forecasting approaches.

TABLE 4
Comparative Analysis of the Proposed Method and State-of-the-Art Load Forecasting Models

Paper	Complexity	Dataset Type	Key Idea	Model Architecture
CNN-BiLSTM-SA (Seasonal Forecasting) [21]	Very High	Grid-level (China, seasonal data)	Long-term temporal modeling with decomposition techniques	CNN + BiLSTM + Attention (Seasonal Forecasting)
Boosting Ensemble Model [16]	Medium	Multi-house (13 residential locations)	Ensemble boosting for improved generalization	LightGBM + XGBoost + CatBoost
CNN-BiLSTM-SA + CEEMDAN/VMD [20]	Very High	Utility grid data	Hybrid deep learning with signal decomposition	CNN + BiLSTM + Attention + Signal Decomposition (CEEMDAN, VMD)
Stacking CNN-BiLSTM-Attention + XGBoost [17]	Very High	Quanzhou dataset (2016–2018)	Stacking ensemble with clustering-based sample selection	CNN + BiLSTM + Attention + Stacking + AHC + XGBoost
Proposed Method	High	UCI residential dataset	Behavior-aware adaptive forecasting with hyperparameter optimization	Proposed Method: CNN + BiLSTM + Attention + Consumer Behavior + Genetic Algorithm

A comparative analysis of recent short-term electricity forecasting approaches indicates a progressive shift from traditional machine learning techniques toward more complex hybrid deep learning frameworks. Earlier approaches based on boosting algorithms such as LightGBM, XGBoost, and CatBoost demonstrate strong generalization performance and effective handling of structured tabular data in residential load forecasting scenarios, as reported in recent studies on ensemble-based methods.

Deep learning-based models, particularly those integrating CNN, BiLSTM, and attention mechanisms, have shown improved capability in extracting local temporal features and modeling long-range dependencies in electricity consumption data. These architectures are especially effective in capturing nonlinear and seasonal patterns in grid-level load forecasting problems. However, several studies report that their performance is often enhanced by additional preprocessing strategies such as signal decomposition or clustering-based data selection, which increases the overall modeling pipeline complexity.

Hybrid stacking-based approaches further improve forecasting accuracy by combining deep feature extractors with machine learning-based meta-learners such as XGBoost. Although these methods achieve higher predictive accuracy, they introduce additional architectural complexity due to multi-stage learning processes. Similarly, decomposition-based frameworks (e.g., CEEMDAN and VMD-based models [20]) have demonstrated strong performance improvements by separating signal components, but they also increase computational complexity and reliance on preprocessing steps.

The proposed method extends existing research by incorporating consumer behavior-related features alongside deep learning-based temporal modeling and genetic algorithm-based hyperparameter optimization. This integration improves predictive performance and enhances model adaptability across multiple benchmark datasets. Compared to existing approaches, the proposed framework achieves improved forecasting accuracy while maintaining a balance between model complexity and interpretability.

A. Hyperparameter Optimization Comparison

To ensure robust performance of the forecasting model, we conducted a comparative study of hyperparameter optimization methods, including Genetic Algorithm (GA), Bayesian Optimization (BO), and Particle Swarm Optimization (PSO). The comparison evaluates both predictive performance and computational cost for our residential load forecasting task. Table 5. summarizes the performance of each method in terms of forecasting accuracy (RMSE). GA achieved superior results, with 5.1% improvement over BO and 3.2% improvement over PSO. This demonstrates GA's effectiveness in finding optimal hyperparameters in a multi-modal search space characteristic of residential electricity consumption patterns.

TABLE 5
Comparison of Hyperparameter Optimization Methods for Residential Load Forecasting.

Method	RMSE (kWh)	Performance Improvement	Computation Time (GPU-hours)
Genetic Algorithm (GA)	0.078 (UCI)	—	~12
Bayesian Optimization	0.082 (UCI)	-5.1%	~4
Particle Swarm (PSO)	0.081 (UCI)	-3.2%	~8

GA employs a population-based global search that is particularly effective for exploring hyperparameter spaces with multiple local optima. In the context of residential load forecasting, where consumption patterns are highly variable and nonlinear, GA's ability to avoid local minima contributes to its superior predictive performance, despite higher computational costs compared to BO and PSO. While BO is faster, GA provides better generalization and global optimum search, making it suitable for our forecasting task.

B. Ablation Studies

To assess the individual contribution of each component in the proposed Consumer Behavior–Enhanced CNN–BiLSTM–Attention model, a series of ablation experiments were conducted. In each experiment, one component was removed or replaced while keeping all other settings constant. The results are summarized Table 6.

TABLE 6
Ablation study results showing the contribution of each model component

Model Variant	RMSE	Performance Change (%)
Full Model (CNN-BiLSTM-Attention + Behavior)	0.078	–
Without Consumer Behavior Features	0.094	–20.5%
Without Attention Mechanism	0.087	–11.5%
Without CNN Component	0.089	–14.1%
With LSTM instead of BiLSTM	0.084	–7.7%
Without GA Optimization	0.082	–5.1%

The results reveal several key findings:

1. Behavior features provide the largest contribution, with a 20.5% increase in RMSE when removed. This confirms that household routines—such as activity onset/offset and consumption stability—play a central role in improving short-term residential forecasting.
2. CNN layers contribute significantly (14.1% degradation when removed) by capturing short- and mid-range temporal patterns commonly present in residential loads.
3. Attention improves performance by 11.5%, demonstrating its capability to highlight the most relevant historical time points, particularly during consumption transitions.
4. BiLSTM slightly outperforms LSTM (7.7% difference), indicating that bidirectional context helps capture consumption peaks and transitional behaviors.
5. GA-based hyperparameter tuning provides an additional 5.1% gain, demonstrating its usefulness despite the cost of evolutionary search.

These ablation results show that the performance improvements of the proposed model are the result of multiple complementary components, and that no single module alone accounts for the overall enhancement. The synergy between CNN extraction, BiLSTM temporal modeling, attention weighting, and behavior-aware features leads to robust forecasting accuracy. To further understand why the proposed architecture performs well, attention weights were analyzed over representative 48-hour periods. The model consistently assigns higher weights to: same-hour patterns from previous days, morning and evening transitional periods, and abrupt changes in consumption routine. Permutation importance analysis shows that several behavior-aware features strongly influence attention distributions. Table 7. summarizes these correlations.

TABLE 7. Correlation between consumer behavior features and attention weights in the CNN–BiLSTM–Attention model.

Consumer Behavior Feature	Correlation with Attention Weight	Significance
Activity Onset/Offset	0.82	Very High
Consumption Stability Index	0.73	High
Base Load Ratio	0.58	Moderate
Peak Coincidence Factor	0.64	Moderate-High
Behavioral Anomaly Flag	0.47	Moderate

These results validate that the attention mechanism not only improves accuracy but also enhances interpretability by focusing on behavior-dependent temporal patterns. To enable real-time operation on resource-constrained hardware, the final trained model was quantized using TensorFlow Lite. Converting 32-bit floating-point weights to 8-bit integers reduced inference latency by 65% with only a 3.5% decrease in predictive accuracy. This demonstrates that the model's design remains suitable for deployment in smart meters and home IoT devices.

F. Limitations and Future Work

Despite strong overall performance, the model has several limitations: It requires approximately two weeks of historical data for optimal performance (cold-start issue). Accuracy temporarily decreases during unusual patterns such as holidays, long absences, or extreme weather events. Fine-grained behavioral features raise privacy considerations in real-world smart meter applications. Future work will explore transfer learning for cold-start adaptation, anomaly-resilient modeling via autoencoders or changepoint detection, federated learning for privacy-preserving deployment, and integration of multimodal data sources such as occupancy sensors and appliance-level monitoring.

CONCLUSION

This study presented a Consumer Behavior–Enhanced CNN–BiLSTM–Attention model for short-term residential electricity consumption forecasting. By integrating engineered behavioral features with a hybrid deep learning architecture, the model achieved significantly improved accuracy compared with traditional statistical, machine learning, and deep learning baselines. Ablation studies confirmed the complementary value of each component, with behavioral features, CNN extraction, and attention weighting contributing most strongly to performance.

The model also provides interpretability through attention-based temporal analysis, enabling clearer insights into the influence of household routines and consumption transitions. Pilot deployment on resource-constrained edge devices, supported by quantization, demonstrated real-time feasibility with minimal loss of accuracy.

Overall, the findings underscore the importance of behavior-aware modeling for next-generation smart energy systems and illustrate that combining deep temporal architectures with human-centric behavioral analytics yields a robust, interpretable, and edge-deployable solution for residential load forecasting. Future work will explore cold-start adaptation, anomaly-resilient modeling, and privacy-preserving learning frameworks to further enhance real-world applicability.

REFERENCE

[1] International Energy Agency. (2024). Energy Efficiency 2024. IEA, Paris, France. [Online]. Available: https://www.iea.org/reports/energy-efficiency-2024

[2] B. Guo, T. Li, H. Yu, and V. Loftness. (2025). AI-driven multi-algorithm optimization for enhanced building energy benchmarking. *Journal of Building Engineering*. [Online]. *105*, p. 112351. Available: https://doi.org/10.1016/j.jobe.2025.112351

[3] H. Karimianfard. (2025). A robust optimization framework for smart home energy management: Integrating photovoltaic storage, electric vehicle charging, and demand response. *Journal of Energy Storage*. [Online]. *110*, p. 115259. Available: https://doi.org/10.1016/j.est.2024.115259

[4] C. Lavin and H. Julienne. (2025). Household activities underlying residential electricity demand: Who does what during the evening peak? *Energy Efficiency*. [Online]. *18(5)*, p. 43. Available: https://doi.org/10.1007/s12053-025-10322-4

[5] X. Li, Z. Wang, C. Yang, and A. Bozkurt. (2024). An advanced framework for net electricity consumption prediction: Incorporating novel machine learning models and optimization algorithms. *Energy*. [Online]. *296*, p. 131259. Available: https://doi.org/10.1016/j.energy.2024.131259

[6] X. Wen, J. Liao, Q. Niu, N. Shen, and Y. Bao. (2024). Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. *Scientific Reports*. [Online]. *14(1)*, p. 13720. Available: https://doi.org/10.1038/s41598-024-63262-x

[7] H. Xu, F. Hu, X. Liang, G. Zhao, and M. Abugunmi. (2024). A framework for electricity load forecasting based on attention mechanism time series depthwise separable convolutional neural network. *Energy*. [Online]. *299*, p. 131258. Available: https://doi.org/10.1016/j.energy.2024.131258

[8] Y. Dai and W. Yu. (2024). Short-term power load forecasting based on Seq2Seq model integrating Bayesian optimization, temporal convolutional network and attention. *Applied Soft Computing*. [Online]. *166*, p. 112248. Available: https://doi.org/10.1016/j.asoc.2024.112248

[9] Y. Wang, J. Wang, G. Zhao, and Y. Dong. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy*. [Online]. *48*, pp. 284–294. Available: https://doi.org/10.1016/j.enpol.2012.05.026

[10] B. Nepal, M. Yamaha, A. Yokoe, and T. Yamaji. (2020). Electricity load forecasting using clustering and ARIMA model for energy management in buildings. *Japan Architectural Review*. [Online]. *3(1)*, pp. 62–76. Available: https://doi.org/10.1002/2475-8876.12135

[11] A. M. Khan and M. Osińska. (2023). Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India. *Expert Systems with Applications*. [Online]. *212*, p. 118840. Available: https://doi.org/10.1016/j.eswa.2022.118840.

[12] A. A. Pierre, S. A. Akim, A. K. Semenyo, and B. Babiga. (2023). Peak electrical energy consumption prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU approaches. *Energies*. [Online]. *16(12)*, p. 4739. Available: https://doi.org/10.3390/en16124739

[13] K. H. Baesmat, F. Shokoohi, and Z. Farrokhi. (2025). SP-RF-ARIMA: A sparse random forest and ARIMA hybrid model for electric load forecasting. *Global Energy Interconnection*. [Online]. Available: https://doi.org/10.1016/j.gloei.2025.04.003

[14] S. Rani et al. (2025). Dual attention-empowered bidirectional long short-term memory network for short-term electric load forecasting. *Energy Technology*. [Online]. p. e202500914. Available: https://doi.org/10.1002/ente.202500914

[15] G. R. Harikrishnan, S. Sreedharan, and others. (2025). Advanced short-term load forecasting for residential demand response: An XGBoost-ANN ensemble approach. *Electric Power Systems Research*. [Online]. *242*, p. 111476. Available: https://doi.org/10.1016/j.epsr.2025.111476

[16] A. Muqtadir, B. Li, Z. Ying, C. Songsong, and S. N. Kazmi. (2025). Nowcasting the next hour of residential load using boosting ensemble machines. *Scientific*

Reports. [Online]. *15(1)*, p. 7157. Available: https://doi.org/10.1038/s41598-025-91767-6

[17] S. Luo, B. Wang, Q. Gao, Y. Wang, and X. Pang. (2024). Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term electricity load forecasting. *Energy Reports*. [Online]. *12*, pp. 2676–2689. Available: https://doi.org/10.1016/j.egyr.2024.08.078

[18] J. Zhang et al. (2025). Process monitoring for tower pumping units under variable operational conditions: From an integrated multitasking perspective. *Control Engineering Practice*. [Online]. *156*, p. 106229. Available: https://doi.org/10.1016/j.conengprac.2024.106229

[19] J. Zhang et al. (2023). Lifetime extension approach based on the Levenberg–Marquardt neural network and power routing of DC–DC converters. *IEEE Transactions on Power Electronics*. [Online]. *38(8)*, pp. 10280–10291. Available: https://doi.org/10.1109/TPEL.2023.3275791

[20] X. Liu, J. Song, H. Tao, P. Wang, H. Mo, and W. Du. (2025). Quarter-hourly power load forecasting based on a hybrid CNN-BiLSTM-Attention model with CEEMDAN, K-means, and VMD. *Energies*. [Online]. *18(11)*, p. 2675. Available: https://doi.org/10.3390/en18112675

[21] M.-P. Wu and F. Wu. (2024). Predicting residential electricity consumption using CNN-BILSTM-SA neural networks. *IEEE Access*. [Online]. *12*, pp. 71555–71565. Available: https://doi.org/10.1109/ACCESS.2024.3400972

[22] G. Hebrail and A. Berard. (2006). Individual Household Electric Power Consumption. UCI Machine Learning Repository. [Online]. Available: https://doi.org/10.24432/C58K54

[23] J. Kelly and W. Knottenbelt. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data*. [Online]. *2*, p. 150007. Available: https://doi.org/10.1038/sdata.2015.7

[24] Z. Zheng. (2025). Pecan Street Electricity Data. [Online]. Available: https://www.pecanstreet.org/wp-content/uploads/2025/08/Pecan-Street-Data-Sheet-Aug-2025.pdf