Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 8, No.3, July 2026, pp. 02603034-01 ~ 02603034-011
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v8i3.2869

Separable Convolutional Hierarchical Decomposition for Lightweight Residential Load Forecasting in Smart Grids

Satriawan Rasyid Purnama, Henri Tantyoko*, Adi Wibowo, Yesaya Rudolf Susanto Widyanto

Departement of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Jl. Prof. H. Soedarto, S.H., Tembalang, Semarang 50275, Central Java, Indonesia

*henritantyoko@live.undip.ac.id

Abstract. Residential load forecasting is essential for maintaining grid stability and energy management in smart grids. However, achieving accurate real-time forecasting under resource constraints remains challenging because Transformer and LSTM models can be computationally demanding, while lightweight linear models such as DLinear have limited modeling flexibility. This study investigates whether a hierarchical separable convolutional framework can provide accurate and efficient residential load forecasting. To address this, SeparableCLF, a lightweight hierarchical decomposition model using depthwise separable convolution, is proposed and evaluated on hourly OpenEI residential load data from 20 U.S. states (2012) at forecast horizons of 6, 12, 24, 48, and 96 h. Relative to DLinear, SeparableCLF reduced MAPE by 0.93, 1.54, and 0.79 percentage points at 24, 48, and 96 h, respectively while requiring substantially fewer parameters than Transformer and LSTM models.SeparableCLF achieved the lowest MAPE at 12, 24, and 48 h. DLinear achieved the lowest errors at 6 h, whereas LSTM achieved the lowest MAPE at 96 h; at 96 h, SeparableCLF retained the lowest MAE, MSE, and RMSE among the compared models, indicating suitability for real-time forecasting on smart meters and edge-based smart grid devices.

Keywords: Residential load forecasting, smart grid, time-series forecasting, hierarchical decomposition, edge computing

(Received 2025-10-14, Revised 2026-06-04, Accepted 2026-06-04, Available Online by 2026-07-10)

1. Introduction

Residential load forecasting plays a critical role in ensuring grid stability [1], improving energy efficiency [2], and supporting real-time decision-making in smart grids [3]. However, forecasting residential electricity consumption remains challenging due to its volatility, non-stationarity [4], and dependence on complex behavioral and environmental factors [5]. Early approaches relied on statistical

models such as ARIMA [6] and Exponential Smoothing [7], which are effective for linear and stationary data but limited in capturing nonlinear demand patterns. Subsequently, traditional machine learning methods, including SVR, Random Forests, and Gradient Boosting, improved predictive flexibility [8] but often represented time series as independent feature vectors, thereby limiting their ability to model temporal dependencies [9–11].

Recent years have witnessed significant advances in deep learning-based time series forecasting, particularly through Recurrent Neural Networks (RNNs) and Transformer architectures [12]. The Long Short-Term Memory (LSTM) network has been widely adopted for its ability to capture long-term dependencies and nonlinear temporal dynamics in load data [13], with prior studies demonstrating effectiveness in residential load modeling through appliance-level learning [14], multi-layer architectures [15], and frequency-based decomposition [16]. Nevertheless, LSTM-based approaches remain computationally expensive, limiting scalability in real-time and edge applications [17]. In parallel, Transformer-based models, including Informer [18,19], Autoformer [20], and PatchTST [21], further improve long-range dependency modeling through attention mechanisms. However, their attention operations still introduce high memory and computational requirements for long sequences, even in advanced variants [22,23]. Consequently, these models are not well suited for lightweight, low-latency, and resource-constrained forecasting scenarios.

To improve efficiency, recent studies have explored CNN-based and MLP-based alternatives. CNNs offer strong potential due to their local receptive fields, translation invariance, and parallel computation, which naturally capture localized temporal dependencies. Early CNN-based models, such as Temporal Convolutional Networks (TCN), successfully modeled sequential data using dilated and causal convolutions [24]. Subsequent frameworks like MICN and HyDCNN further improved feature extraction through multi-scale convolutional filters and hierarchical decomposition [25]. However, conventional CNN-based models may require larger parameter counts when extended to long-term dependency modeling. In contrast, DLinear demonstrates that simple decomposition-based linear mappings can achieve competitive forecasting performance with substantially lower complexity [26,27]. Despite its efficiency, DLinear remains limited in capturing localized nonlinear patterns because it relies mainly on linear mappings after decomposition.

Despite these advances, existing approaches still exhibit a trade-off between computational efficiency and representational capability. Linear models such as DLinear are lightweight but limited in local nonlinear feature extraction, whereas CNN-based models are expressive but often become computationally heavier when modeling long-horizon dependencies. More importantly, there remains a lack of lightweight and interpretable CNN-based frameworks specifically tailored for residential load forecasting. Existing decomposition-based CNN models such as MICN and HyDCNN mainly emphasize multi-scale feature extraction, while limited work has explored the integration of hierarchical decomposition with computationally efficient convolutional mechanisms such as depthwise separable convolutions in this domain.

To address this gap, this study proposes SeparableCLF, a lightweight forecasting framework that integrates multi-level hierarchical decomposition with depthwise separable convolution. Depthwise separable convolution, originally popularized in lightweight architectures such as MobileNet [28] and Xception [29], decomposes standard convolution into depthwise and pointwise operations to reduce computational cost while preserving representational capability. The central hypothesis of this study is that combining structured hierarchical decomposition with separable convolution enables more efficient modeling of global trends and local temporal dependencies than linear decomposition alone, thereby improving forecasting accuracy under resource constraints. Unlike existing decomposition-based CNN approaches, SeparableCLF is specifically designed to balance accuracy, interpretability, and parameter efficiency for smart-grid residential load forecasting.

The main contributions of this study are summarized as follows:

- A novel lightweight forecasting framework (SeparableCLF) that integrates hierarchical decomposition with separable convolutional layers for efficient temporal modeling.

- A multi-level hierarchical decomposition strategy using convolution and average pooling to effectively separate trend and seasonal components with experimentally validated depth.
- Conducting comprehensive comparisons with DLinear, LSTM, and Transformer across multiple horizons.

2. Methods

The proposed Separable Convolutional Linear for Forecasting (SeparableCLF) offers a lightweight and efficient framework tailored for time-series forecasting within smart grid environments. It combines hierarchical signal decomposition with separable convolutional forecasting to model temporal interactions using minimal computational resources. Unlike standard Transformer and recurrent models that rely on computationally expensive self-attention or recurrent mechanisms, SeparableCLF employs cascaded one-dimensional convolutional and pooling operations to hierarchically extract trend and seasonal components. These multi-scale features are subsequently processed through depthwise separable convolutions that efficiently capture temporal patterns with a compact parameter set. This architecture delivers high forecasting accuracy with low complexity, making it highly suitable for real-time edge deployment on smart meters. To ensure the proposed methodology is fully self-contained and independently reproducible, all data transformations, tensor shapes, and architectural hyperparameters are explicitly formulated within this section.

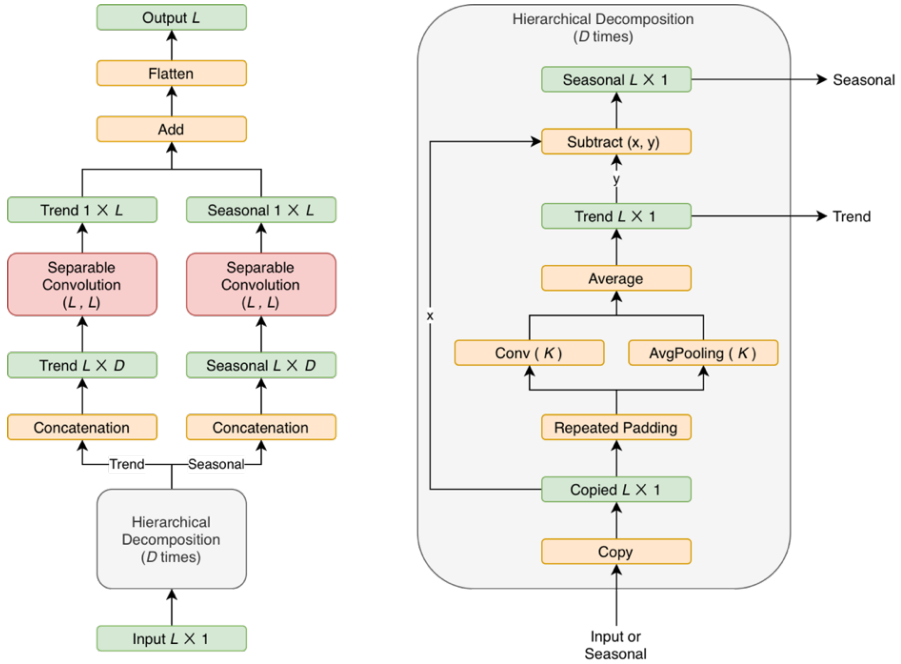

Figure 1. Detailed architecture of the proposed SeparableCLF framework. The left panel illustrates the end-to-end data flow, mapping the input sequence $\mathbf{X} \in \mathbb{R}^{L\times 1}$ to the forecast horizon Y. The right panel zooms into the Hierarchical Decomposition module, detailing the recursive convolutional smoothing across $D = 5$ levels using a dynamically sized kernel $K = \lfloor L/2 \rfloor$ and the subsequent parallel processing via the depthwise Separable Convolutional blocks.

The proposed SeparableCLF offers a lightweight framework for time-series forecasting. Operating on a strictly univariate historical load signal, the input sequence is formulated as $X \in \mathbb{R}^{L\times 1}$, where L defines the look-back window. To construct the dataset, a sliding window technique is employed: overlapping windows with a stride of 1 are used during training to maximize sample diversity, while non-overlapping windows with a stride equal to the forecast horizon H are strictly enforced during

evaluation to prevent data leakage. The model assumes the univariate data achieves weak stationarity post-normalization and learns an autoregressive mapping function $\mathcal{H}_\theta(\cdot)$ to generate the forecast $\widehat{\mathrm{Y}} \in \mathbb{R}^{H\times 1}$.

The Hierarchical Decomposition module extracts multi-scale trend and seasonal components through recursive convolutional smoothing. Given the initial sequence $\mathrm{X}^{(0)} = \mathrm{X}$, boundary replication padding is applied to preserve edge continuity. At each decomposition level $\mathrm{d} \in \{1, \ldots, \mathrm{D}\}$, the trend component $\mathrm{T}^{(\mathrm{d})}$ is extracted via one-dimensional average pooling using a dynamic kernel size $\mathrm{K} = \lfloor \mathrm{L}/2 \rfloor$, formally defined as $\mathrm{T}_\mathrm{t}^{(\mathrm{d})} = \frac{1}{\mathrm{K}} \sum_{\mathrm{i}=0}^{\mathrm{K}-1} \mathrm{X}_{\mathrm{t}+\mathrm{i}-\lfloor \mathrm{K}/2 \rfloor}^{(\mathrm{d}-1)}$. The seasonal residual is obtained via element-wise subtraction: $\mathrm{S}^{(\mathrm{d})} = \mathrm{X}^{(\mathrm{d}-1)} - \mathrm{T}^{(\mathrm{d})}$, which recursively serves as input for level $\mathrm{d}+1$. The decomposition depth is empirically set to $\mathrm{D} = 5$ to optimally balance fine-grained feature extraction and efficiency. Unlike standard moving averages that filter noise at a single frequency, this recursive design dynamically isolates overlapping high- and low-frequency cyclical variations inherent in univariate residential loads.

Following decomposition, the multi-level features $\mathrm{T}^{(1\ldots\mathrm{D})}$ and $\mathrm{S}^{(1\ldots\mathrm{D})}$ are concatenated and processed by parallel depthwise separable convolutional blocks. To capture adequate receptive fields, the separable kernel is proportional to the sequence length, set as $\mathrm{K}_{\mathrm{sep}} = \lfloor \mathrm{L}/4 \rfloor$, operating with a stride of $\mathrm{S} = 1$ and symmetric zero-padding. This operation decouples spatial and channel-wise filtering, significantly reducing parameters while mapping features to an expanded latent space of $\mathrm{C}_{\mathrm{out}} = 64$ channels followed by a Gaussian Error Linear Unit (GELU) activation. The parallel branches are then fused via element-wise addition to maintain a compact feature representation and flattened into a vector Z_f. The final prediction is generated via a linear projection: $\widehat{\mathrm{Y}} = \mathrm{W}_\mathrm{f}\mathrm{Z}_\mathrm{f} + \mathrm{b}_\mathrm{f}$, where W_f and b_f represent the learnable weight matrix and bias vector, respectively.

A primary advantage of the SeparableCLF architecture is its computational efficiency. By utilizing depthwise separable convolutions, the theoretical complexity is reduced to $\mathcal{O}(\mathrm{L} \cdot \mathrm{C}_{\mathrm{out}})$, achieving a linear scaling property that vastly outperforms the quadratic $\mathcal{O}(\mathrm{L}^2 \cdot \mathrm{C})$ complexity of standard Transformer self-attention mechanisms. However, the model possesses inherent limitations. The reliance on successive average pooling introduces a risk of over-smoothing highly volatile, high-frequency seasonal spikes typical in residential energy use. Furthermore, operating strictly on univariate signals without exogenous contextual features (such as temperature anomalies) limits the model's ability to adapt to severe non-stationarity or sudden concept drifts driven entirely by unobserved external factors.

3. Results and Discussion

The proposed method was evaluated on the publicly available OpenEI dataset, which provides hourly residential load data from various U.S. cities and states. One year of data from 2012 was used, selecting 20 states with at least five households and randomly choosing five from each for analysis. The dataset was split into 80% for training and 20% for testing, with 10% of the training set reserved for validation. All data were normalized between 0 and 1 using training statistics. The model was trained for up to 200 epochs with a batch size of 32, using the Adam optimizer (learning rate 1e-3) and Mean Absolute Error (MAE) loss. Early stopping with 25-epoch patience monitored validation loss and restored the best weights to prevent overfitting. Finally, the outputs were denormalized to their original scale for evaluation.

Model performance was assessed using the Mean Absolute Percentage Error (MAPE), defined in (1):

$$MAPE = \frac{100\%}{N} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (1)$$

where y_t and $\hat{y}_t$ represent the actual and predicted load values at time t, respectively, and N denotes the total number of observations. MAPE measures the average percentage deviation between

predicted and actual values, providing an intuitive and scale-independent evaluation of forecasting performance.

For comparative evaluation, three baseline models were implemented alongside the proposed SeparableCLF architecture, namely DLinear, LSTM, and Transformer, representing three distinct paradigms of time-series forecasting. The DLinear model follows the original formulation, which decomposes the input sequence into trend and seasonal components using simple linear mappings with a moving average window size proportional to half the sequence length, providing a strong linear baseline with high efficiency. The LSTM model employs a single recurrent layer with 64 hidden units to capture temporal dependencies across time steps, followed by a dense output layer producing the final sequence prediction, reflecting the memory-based modeling approach commonly used in sequential forecasting. The Transformer model adopts a lightweight encoder-only structure with two stacked self-attention blocks, each consisting of multi-head attention (four heads, feature dimension of 64) and feed-forward sublayers combined with residual connections, layer normalization, and dropout regularization. This configuration enables efficient sequence-to-sequence modeling through global attention mechanisms while maintaining computational tractability. Model performance was evaluated using four complementary metrics: MAPE, MAE, MSE. And RMSE, to provide a more comprehensive assessment of both relative and absolute forecasting errors. All models were trained and evaluated under the same experimental settings to ensure a fair comparison.

Table 1. Multi-Metric Forecasting Performance Comparison Across Prediction Horizons

Horizon	Model	MAPE (%)	MAE	MSE	RMSE
6 h	DLinear	**17.52 ± 9.15**	**2.72 ± 0.54**	**15.46 ± 5.92**	**3.87 ± 0.74**
	LSTM	19.38 ± 10.67	2.87 ± 0.44	15.88 ± 4.98	3.94 ± 0.63
	Transformer	27.31 ± 10.23	4.77 ± 1.36	40.05 ± 20.27	6.14 ± 1.56
	SeparableCLF	19.99 ± 11.52	3.11 ± 0.70	18.54 ± 8.40	4.21 ± 0.92
12 h	DLinear	18.84 ± 9.69	2.96 ± 0.67	17.11 ± 8.55	4.03 ± 0.97
	LSTM	17.42 ± 7.17	2.93 ± 0.69	16.90 ± 8.29	4.00 ± 0.96
	Transformer	20.10 ± 7.80	3.35 ± 1.02	21.24 ± 13.44	4.42 ± 1.33
	SeparableCLF	**16.78 ± 7.84**	**2.73 ± 0.66**	**15.25 ± 8.16**	**3.79 ± 0.95**
24 h	DLinear	23.48 ± 7.34	3.99 ± 1.23	31.65 ± 19.62	5.40 ± 1.63
	LSTM	23.51 ± 7.54	4.03 ± 1.24	32.09 ± 19.80	5.44 ± 1.64
	Transformer	31.43 ± 12.88	4.91 ± 1.55	43.77 ± 29.43	6.33 ± 1.99
	SeparableCLF	**22.55 ± 7.09**	**3.76 ± 1.21**	**29.69 ± 19.22**	**5.21 ± 1.63**
48 h	DLinear	31.06 ± 11.40	5.16 ± 1.44	49.31 ± 25.19	6.81 ± 1.76
	LSTM	31.12 ± 11.25	5.29 ± 1.58	51.62 ± 28.57	6.93 ± 1.94
	Transformer	37.82 ± 15.34	5.80 ± 1.45	58.41 ± 27.03	7.44 ± 1.81
	SeparableCLF	**29.52 ± 11.00**	**4.87 ± 1.38**	**45.22 ± 24.89**	**6.50 ± 1.74**
96 h	DLinear	37.17 ± 15.88	6.20 ± 1.64	68.17 ± 34.33	8.00 ± 2.11
	LSTM	**35.22 ± 15.88**	6.26 ± 1.90	69.98 ± 39.33	8.04 ± 2.37
	Transformer	53.70 ± 25.54	8.38 ± 2.61	118.90 ± 72.03	10.46 ± 3.16
	SeparableCLF	36.38 ± 15.23	**6.04 ± 1.61**	**64.49 ± 31.26**	**7.77 ± 2.05**

Table 1 shows that the relative performance of each model varies across prediction horizons. At 6 h, DLinear achieves the lowest error, indicating that ultra-short-term residential load is still strongly governed by local autocorrelation and can be captured by simple linear decomposition. At 12 h,

SeparableCLF becomes competitive, while at 24 h and 48 h it achieves the best overall performance across the reported metrics. This improvement suggests that hierarchical decomposition helps isolate daily and sub-daily structures, while separable convolution captures localized nonlinear variations without excessive parameter growth. At 96 h, the task becomes more uncertain; although LSTM obtains a slightly lower MAPE, SeparableCLF achieves lower MAE, MSE, and RMSE, indicating better control of absolute errors and large deviations.

While the multi-metric evaluation demonstrates the forecasting capability of each model, accuracy alone is not sufficient for assessing suitability in real-world smart grid deployment. Since residential load forecasting may need to be performed on edge devices with limited memory and computational capacity, model complexity must also be considered. Figure 2 therefore presents a comparison of the number of trainable parameters across the evaluated models.

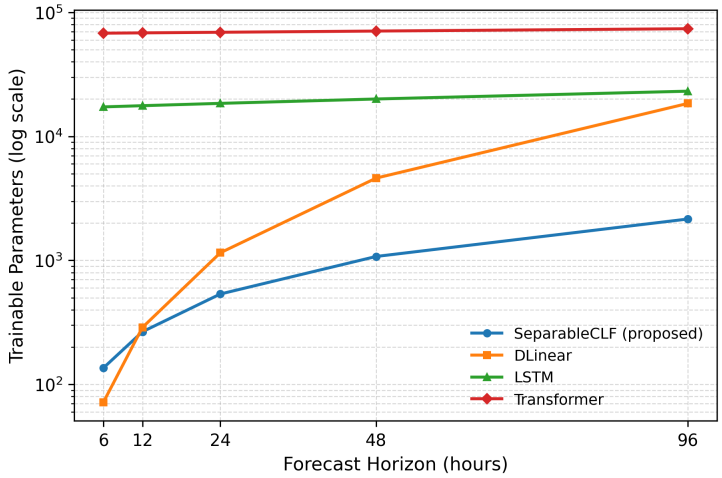

Figure 2. Comparison of Model Complexity (Number of Trainable Parameters)

SeparableCLF achieves the lowest parameter count across all horizons, increasing from only 136 parameters at 6 h to 2,156 parameters at 96 h, compared with 18,432 for DLinear and approximately 23,000 and 74,000 for LSTM and Transformer. This compactness demonstrates a favorable complexity–performance trade-off, as depthwise separable convolutions reduce parameter growth while maintaining competitive accuracy. With less than 5 MB memory usage and approximately 12 ms inference latency, the low parameter count makes SeparableCLF a potential candidate for future resource-constrained deployment, including smart meters, edge concentrators, demand response, and microgrid control.

Figure 3 presents a representative 48-hour forecasting case to examine how each model follows peak, valley, and ramping behavior beyond aggregate error metrics.

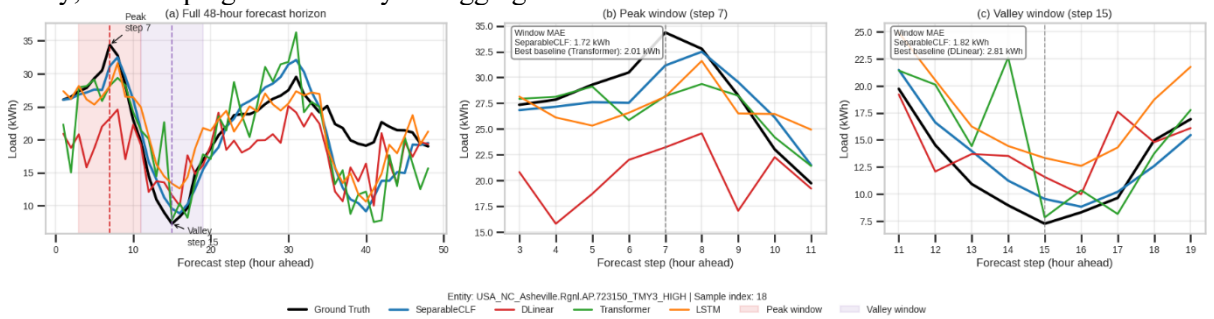

Figure 3. Case-based 48-hour Forecast Comparison with Peak and Valley Windows.

Figure 3 provides a case-based 48-hour comparison with zoomed peak and valley regions. The results show that SeparableCLF better preserves local peak–valley trajectories and ramping behavior, whereas DLinear tends to over-smooth local changes and LSTM/Transformer show larger phase or amplitude deviations. This qualitative evidence supports the aggregate metrics by showing that the proposed model improves not only average error but also local trajectory consistency.

To further examine whether the observed performance differences are statistically reliable, a balanced paired statistical analysis was conducted. Instead of pooling all sample-level prediction windows directly, the errors were first averaged for each horizon–entity–model combination. This procedure prevents shorter horizons from dominating the statistical test due to their larger number of prediction windows. The Wilcoxon signed-rank test was then applied to the paired horizon–entity errors, resulting in 100 paired observations for each comparison, obtained from five forecasting horizons and twenty entities. The improvement value is defined as the baseline error minus the SeparableCLF error; therefore, a positive value indicates that SeparableCLF achieves a lower error.

Table 2. Statistical Significance of Improvements (SeparableCLF vs. Baselines)

Comparison	Metric	Mean Improvement	95% Confidence Interval	p-value (Wilcoxon)	Significance
vs. DLinear	MAE	0.108	[0.033, 0.182]	2.57×10^{-4}	Yes
	RMSE	0.644	[0.092, 1.196]	9.18×10^{-4}	Yes
vs. LSTM	MAE	0.143	[0.066, 0.217]	8.35×10^{-6}	Yes
	RMSE	0.181	[0.075, 0.290]	4.61×10^{-5}	Yes
vs. Transformer	MAE	0.362	[-0.360, 1.125]	1.87×10^{-2}	Yes
	RMSE	0.181	[0.064, 0.303]	4.28×10^{-5}	Yes

The Wilcoxon signed-rank test in Table 2 confirms that the improvements of SeparableCLF over the baselines are statistically significant under the balanced horizon–entity evaluation. The positive mean improvements indicate that SeparableCLF generally reduces absolute and relative errors compared with DLinear, LSTM, and Transformer. However, the modest gain over LSTM in some percentage-based metrics suggests that the advantage should be interpreted as balanced multi-horizon robustness rather than uniform dominance at every horizon.

Table 3 summarizes the MAPE (Mean ± Standard Deviation) for the full model versus these variants across horizons of 6 to 96 hours.

Table 3. Ablation Study Results (Mean ± Std. Dev. MAPE %)

Horizon	Full	w/o Hierarchical	Concat Fusion	Sequential
6 h	**19.99 ± 11.52**	20.42 ± 10.74	20.84 ± 11.67	59.92 ± 11.98
12 h	**16.78 ± 7.84**	17.28 ± 7.83	16.85 ± 7.42	42.86 ± 13.50
24 h	**22.55 ± 7.09**	23.19 ± 6.85	24.81 ± 7.35	46.42 ± 10.19
48 h	**29.52 ± 11.00**	29.83 ± 11.01	29.57 ± 11.35	42.36 ± 16.74
96 h	36.38 ± 15.23	**35.24 ± 14.27**	35.34 ± 14.78	46.10 ± 19.71

The ablation results in Table 3 show that the full SeparableCLF design provides the most balanced performance across the main operational horizons. Removing hierarchical decomposition weakens short- and medium-term accuracy, confirming the importance of multi-level trend–seasonal separation. Alternative fusion or sequential processing variants either increase complexity or degrade stability,

indicating that parallel component processing with compact additive fusion is better aligned with the lightweight forecasting objective.

Additional 48-hour horizon ablation experiments were conducted to justify the main architectural choices. The decomposition-depth analysis showed that depth 4 achieved the lowest MAPE of 29.52%, while depth 5 remained competitive with a MAPE of 29.74% and stable overall errors. The kernel-size analysis also indicated that using a wider temporal receptive field was beneficial, with the decomposition kernel 4 and separable-convolution kernel 8 producing the strongest result, achieving MAE, MAPE, MSE, and RMSE values of 3.5950, 30.4514%, 26.2814, and 4.9494, respectively. The current parameter setting was therefore selected because it preserves competitive accuracy while avoiding excessive decomposition and maintaining the lightweight design objective of SeparableCLF. These findings support the selected hierarchical decomposition and depthwise separable convolution design as a practical trade-off between accuracy, stability, and model complexity.

To further investigate the robustness of SeparableCLF across different residential profiles, an entity-level error analysis was conducted. For each entity, the mean MAPE was calculated across the corresponding test samples, and the resulting distribution is visualized in Fig. 4. This analysis helps identify whether the model errors are broadly distributed across all entities or concentrated in a small number of difficult load profiles.

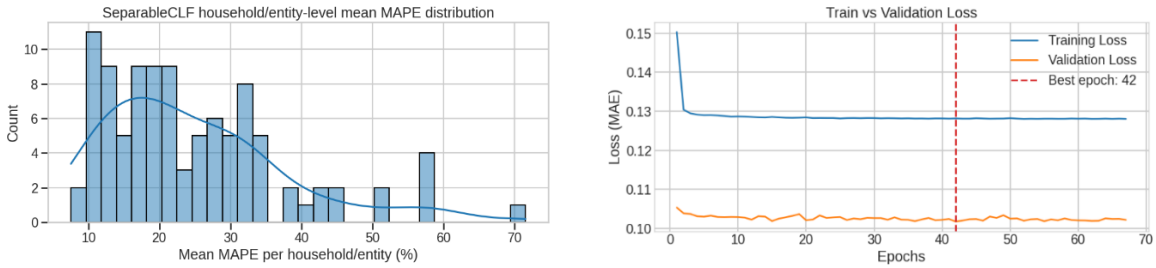

Figure 4. Robustness and Training Stability Analysis: (a) Mean MAPE per Household Distribution and (b) Training Versus Validation Loss.

Figure 4 summarizes the robustness and training stability of SeparableCLF. Most households fall within a moderate MAPE range, although a right-skewed tail indicates several difficult profiles at longer horizons. The training and validation curves converge without substantial divergence, suggesting no clear overfitting or underfitting, while early stopping helps retain the best validation performance.

4. Conclusion

This study introduced SeparableCLF, a lightweight residential load forecasting framework that combines hierarchical decomposition with depthwise separable convolution. Unlike DLinear's purely linear mappings, SeparableCLF introduces deothwise-separable convolution to model local nonlinear patterns while retaining a compact parameterization. The model achieved competitive and stable performance across multiple horizons, particularly at the 24-hour and 48-hour horizons, while maintaining a substantially smaller parameter count than the LSTM and Transformer baselines.

From an engineering perspective, deployment readiness therefore remains to be established on representative smart-meter or edge hardware. The framework can be integrated into smart meters, home energy management systems, and local microgrid controllers to enable low-latency forecasting and decentralized decision-making. In addition, the lightweight design supports edge-based deployment scenarios where computational resources, memory capacity, and power consumption are limited.

This study has several limitations. The evaluation relies solely on the OpenEI 2012 dataset, which may limit generalizability across different countries, climates, building types, and household behaviors. Furthermore, the current framework is strictly univariate and does not incorporate exogenous variables such as weather, calendar effects, or occupancy information.

Future work will therefore focus on validating SeparableCLF across multiple datasets and deployment contexts to improve robustness and generalization capability. Additional research will investigate the integration of exogenous features into the hierarchical decomposition framework to enhance forecasting accuracy under dynamic operating conditions. The lightweight forecasting architecture will also be extended to other smart grid applications, including photovoltaic power generation forecasting, distributed energy resource management, and broader edge-based energy analytics tasks.

Declaration of AI and AI assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT for language editing and manuscript refinement. The authors reviewed and edited the content as needed and take full responsibility for the final manuscript.

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.

Acknowledgements

The author would like to express sincere gratitude to Diponegoro University for the non-state budget (non-APBN) funding support in 2025, managed by the Institute for Research and Community Service (LPPM UNDIP). This financial support has significantly contributed to the smooth implementation and success of this research.

References

[1] Muqtadir A, Li B, Ying Z, Songsong C, Kazmi SN. Nowcasting the next hour of residential load using boosting ensemble machines. Sci Rep 2025;15:7157. https://doi.org/10.1038/s41598-025-91767-6.

[2] La Tona G, Luna M, Di Piazza MC. Day-ahead forecasting of residential electric power consumption for energy management using Long Short-Term Memory encoder–decoder model. Math Comput Simul 2024;224:63–75. https://doi.org/https://doi.org/10.1016/j.matcom.2023.06.017.

[3] G R H, Sreedharan S, C NB. Advanced short-term load forecasting for residential demand response: An XGBoost-ANN ensemble approach. Electric Power Systems Research 2025;242:111476. https://doi.org/https://doi.org/10.1016/j.epsr.2025.111476.

[4] Kaur S, Bala A, Parashar A. A multi-step electricity prediction model for residential buildings based on ensemble Empirical Mode Decomposition technique. Science and Technology for Energy Transition (STET) 2024;79. https://doi.org/10.2516/stet/2024001.

[5] Saini VK, Singh R, Mahto DK, Kumar R, Mathur A. Learning Approach for Energy Consumption Forecasting in Residential Microgrid. 2022 IEEE Kansas Power and Energy Conference, KPEC 2022, 2022. https://doi.org/10.1109/KPEC54747.2022.9814744.

[6] Gupta A, Kumar A. Mid Term Daily Load Forecasting using ARIMA, Wavelet-ARIMA and Machine Learning. 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 2020, p. 1–5. https://doi.org/10.1109/EEEIC/ICPSEurope49358.2020.9160563.

[7] Beydoun H, Khan A, Arefifar SA. Comparative Analysis of Time Series and Artificial Intelligence Algorithms for Short Term Load Forecasting. 2021 IEEE Canadian Conference on

Electrical and Computer Engineering (CCECE), 2021, p. 1–7. https://doi.org/10.1109/CCECE53047.2021.9569166.
[8] Inala KP, Gaddam S, Etti S, Kashetty P, Karangula J, Anaparthi N. Machine Learning Algorithms for Load Forecasting in Smart Grid. In: Panda G, Basu M, Siano P, Affijulla S, editors. Proceedings of Third International Symposium on Sustainable Energy and Technological Advancements, Singapore: Springer Nature Singapore; 2024, p. 487–99.
[9] Wang L, Lin Y, Song T, Chen Y, Li K, Ran J. Short-term residential electricity consumption forecast considering the cumulative effect of temperature, dual decomposition technology and integrated deep learning. Energy Informatics 2025;8:94. https://doi.org/10.1186/s42162-025-00552-2.
[10] Tantyoko H, Nurjanah D, Rusmawati Y. Utilizing Sequential Pattern Mining and Complex Network Analysis for Enhanced Earthquake Prediction. Advance Sustainable Science, Engineering and Technology 2024;6. https://doi.org/10.26877/asset.v6i4.1003.
[11] Lim B, Zohren S. Time-series forecasting with deep learning: a survey. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2021;379:20200209. https://doi.org/10.1098/rsta.2020.0209.
[12] Waheed W, Xu Q, Aurangzeb M, Iqbal S, Dar SH, Elbarbary ZMS. Empowering data-driven load forecasting by leveraging long short-term memory recurrent neural networks. Heliyon 2024;10. https://doi.org/10.1016/j.heliyon.2024.e40934.
[13] Bartouli M, Hagui I, Msolli A, Helali A, Hassen F. Smart Grid Load Forecasting Models Using Recurrent Neural Network and Long Short-Term Memory. Jordan Journal of Electrical Engineering 2025;11:35–51. https://doi.org/10.5455/jjee.204-1703066445.
[14] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. IEEE Trans Smart Grid 2019;10:841–51. https://doi.org/10.1109/TSG.2017.2753802.
[15] Abbasimehr H, Shabani M, Yousefi M. An optimized model using LSTM network for demand forecasting. Comput Ind Eng 2020;143:106435. https://doi.org/https://doi.org/10.1016/j.cie.2020.106435.
[16] Zhang Q, Ma Y, Li G, Ma J, Ding J. Short-Term Load Forecasting Based on Frequency Domain Decomposition and Deep Learning. Math Probl Eng 2020;2020:7240320. https://doi.org/https://doi.org/10.1155/2020/7240320.
[17] Ribes S, Trancoso P, Sourdis I, Bouganis C-S. Mapping Multiple LSTM models on FPGAs. 2020 International Conference on Field-Programmable Technology (ICFPT), 2020, p. 1–9. https://doi.org/10.1109/ICFPT51103.2020.00010.
[18] Taleblou R. Comparing the performance of different deep learning architectures for time series forecasting. Journal of Mathematics and Modeling in Finance 2025;5:63–87. https://doi.org/10.22054/jmmf.2025.83410.1157.
[19] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. CoRR 2020;abs/2012.07436.
[20] Wu H, Xu J, Wang J, Long M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. CoRR 2021;abs/2106.13008.
[21] Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. International Conference on Learning Representations 2023.
[22] Demiss BA, Elsaigh WA. Application of novel hybrid deep learning architectures combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN): construction duration estimates prediction considering preconstruction uncertainties. Engineering Research Express 2024;6:032102. https://doi.org/10.1088/2631-8695/ad6ca7.
[23] Huang Z, Wang P, Wang J, Miao H, Xu J, Zhang P. Improving Transformer Based End-to-End Code-Switching Speech Recognition Using Language Identification. Applied Sciences 2021;11. https://doi.org/10.3390/app11199106.

[24] Bai S, Kolter JZ, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. CoRR 2018;abs/1803.01271.
[25] Li Y, Li K, Chen C, Zhou X, Zeng Z, Li K. Modeling Temporal Patterns with Dilated Convolutions for Time-Series Forecasting. ACM Trans Knowl Discov Data 2021;16. https://doi.org/10.1145/3453724.
[26] Tantyoko H, Purnama SR, Vianita E. Multi-Horizon Short-Term Residential Load Forecasting Using Decomposition-Based Linear Neural Network. Advance Sustainable Science, Engineering and Technology 2025;7. https://doi.org/10.26877/asset.v7i3.2033.
[27] Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, AAAI Press; 2023. https://doi.org/10.1609/aaai.v37i9.26317.
[28] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. CoRR 2017;abs/1704.04861.
[29] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 1800–7. https://doi.org/10.1109/CVPR.2017.195.