LAF-Net: A Deep Residual and Cross-Attention Framework for Day-Ahead Load Forecasting

Zhenlong Jiang, Arezou Ghesmati, Adam Simkowski, Long Zhao
Midcontinent Independent System Operator
Carmel, USA
{zhjiang, aghesmati, asimkowski, lzhao}@misoenergy.org

Cong Feng, Yonghong Chen, Rui Yang
National Laboratory of the Rockies
Golden, USA
{cong.feng, yonghong.chen, rui.yang}@nlr.gov

***Abstract*—Accurate day-ahead load forecasting is essential for reliable power system operations and market efficiency. System operators such as the Midcontinent Independent System Operator (MISO) rely on forecasts from multiple vendors, yet combining them effectively remains a persistent challenge due to vendor-specific biases. This paper presents a novel LSTM–Attention Fusion Network with Error Representation (LAF-Net) that enhances day-ahead hourly load forecasting through deep residual learning and multi-modal cross-attention. The proposed model builds a historical error memory from past vendor performance and dynamically queries it with future hour context to generate adaptive, hour-specific trust weights for each vendor. A bounded residual correction further refines forecasts by mitigating systematic and temporally localized errors. Tested on real MISO LBA data with multi-vendor forecasts, LAF-Net consistently outperforms the best vendor baseline across all 38 LBAs, achieving more than a 40% reduction in system-level mean absolute error (MAE) relative to the best vendor baseline.**

***Index Terms*—Load forecasting, Deep learning, LSTM, Attention mechanism, Multi-vendor forecast fusion**

I. Introduction

Accurate and robust load forecasting is critical for power system operators in daily operations, from unit commitment/dispatch to contingency analysis and emergency preparation. In practice, system operators rely on forecasts from multiple forecasting vendors or forecasts of a set of ensemble members to avoid risk of large bias from a single forecast. However, it poses challenges to operators in obtaining optimal forecasts through selecting and combining multiple forecasts [1].

First, the surging of machine learning introduces a large collection of load forecasting methods. Specifically, deep learning models have been increasingly applied to load forecasting tasks due to their ability to capture complex temporal dependencies and nonlinear patterns in power system data. For example, [2] employed a coarse set-based backpropagation neural network to handle complex short-term load prediction under dynamic and nonlinear conditions, resulting in improved forecasting accuracy. [3] proposed a hybrid Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). model for the Bangladesh power system, which achieved higher precision than conventional models such as LSTM, radial basis function networks, and extreme gradient boosting. [4] introduced a hybrid Transformer-based model combining Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, Sample Entropy, and the Transformer architecture for short-term load forecasting in New York City at 4h, 8h, 12h, and 24h horizons. The proposed method outperformed multiple baselines, achieving a 24-hour Mean Absolute Error (MAE) of 0.95. A broader overview of recent developments in load forecasting methods is available in [5], [6].

To better utilize the various load forecasting methods, research has been done to further improve the forecasting accuracy by forecast fusion. For example, [7] developed a multi-source fusion load forecasting model that dynamically integrates LSTM, bagging, and boosting regressors with adaptive weighting based on an improved error function, enhancing accuracy under fluctuating load conditions. [8] proposed a PCA-based weighted fusion framework that integrates heterogeneous provincial forecasts, dynamically adjusting regional weights by reliability and impact to enhance grid-level short-term load forecasting accuracy and adaptability. [9] developed a multi-source fusion framework for short-term load forecasting that synchronously integrates multi-regional and external feature via a CNN–LSTM encoder-decoder with attention. However, these weighted methods rely on globally or window-static weighting that overlooks hour-level context and vendor-specific error dynamics, may leading to misweighting during ramps and peaks. [10] [11] updates expert weights via cumulative-loss rules such as Hedge or Fixed-Share, with strong regret guarantees. Although each vendor is reduced to a scalar loss stream, so these methods react to errors without modeling structured, time-specific biases or ensuring consistency across the 24-hour forecast horizon.

In this paper, a novel LSTM-Attention–Fusion Network

This work was supported by the U.S. Department of Energy (DOE), The Solar Energy Technology Office, under Award Number DE-EE0011373 . The authors gratefully acknowledge the support provided for research on load and renewable energy forecasting enhancements to improve uncertainty management within the Midcontinent Independent System Operator (MISO) footprint. The views expressed in this paper are solely those of the authors and do not necessarily represent those of MISO

This work was authored in part by the National Laboratory of the Rockies for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Critical Minerals and Energy Innovation. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

979-8-3315-8138-1

with Error Representation (LAF-Net) architecture is developed for next-day hourly load forecasting. The main contributions of this paper are shown as follows. First, cross attention queries a vendor-specific historical error memory with the future hour as context and feeds the gating network to produce per-hour softmax weights over vendors, replacing window-static weighting with hour-level, context-aware trust. Second, LAF-Net summarizes the attended error representation to generate a bounded residual that is added to the gated base forecast, providing anticipatory adjustment of systematic, time-specific biases and improving the consistency of the 24-hour forecast profile. Crucially, we validate LAF-Net on real MISO LBA-level load with multi-vendor forecasts, where it consistently outperforms the best vendor baseline, demonstrating both practical impact and deployment readiness under extreme conditions.

The remainder of this paper is organized as follows. Section II introduces the LSTM–Attention–Fusion Network with Error Representation model. Section III discusses the results. Conclusions are made in Section IV.

II. Methodology

In this section, we present a LAF-Net architecture for next-day hourly load forecasting. The proposed framework is designed to leverage both historical performance patterns of multiple vendors and their future forecasts, integrating them into a unified model that dynamically adjusts trust in each vendor based on temporal context.

A. Problem Formulation and Notation

MISO performs day-ahead load forecasting every afternoon at 2 PM, providing predictions for the 24 hours of next day with a lead time of 10 hours. Let the forecasting horizon be $T_f = 24$ (next day's 24 hours) and the historical window be T_p. We assume there are V vendors providing independent forecasts. The true load at hour t is denoted y_t and the forecast from vendor v is $\hat{y}_t^{(v)}$. From these, we define: (i) historical error tensor $\mathbf{E} \in \mathbb{R}^{T_p \times V} : \mathbf{E}_{\tau,v} = y_{t-\tau} - \hat{y}_{t-\tau}^{(v)}$, (ii) future vendor forecasts $\mathbf{F} \in \mathbb{R}^{T_f \times V} : \mathbf{F}_{h,v} = \hat{y}_{t+h}^{(v)}, \quad h = 1, \ldots, T_f$, and (iii) hour-of-day embeddings for discrete hours $h \in \{0, \ldots, 23\}$, denoted $\mathrm{Emb}(h) \in \mathbb{R}^{d_{\text{emb}}}$. Then the task is to produce $\hat{\mathbf{y}} = (\hat{y}_{t+1}, \ldots, \hat{y}_{t+T_f}) \in \mathbb{R}^{T_f}$ minimizing the mean squared error against the true load.

B. Component Details

The LAF-Net framework is designed with six sequential modules.

a) Hour-of-Day Embedding

Each historical hour τ and future hour h is mapped to an embedding vector:

$$\mathbf{z}_\tau^{\text{hist}} = \mathrm{Emb}(h_\tau^{\text{hist}}), \quad \mathbf{z}_h^{\text{fut}} = \mathrm{Emb}(h_h^{\text{fut}}). \tag{1}$$

These embeddings provide a shared temporal reference frame for aligning historical error patterns with future contexts.

b) Historical Error Memory Encoder.

We form a time-aware error representation by concatenating vendor errors with their corresponding hour embeddings:

$$\mathbf{X}_\tau^{\text{hist}} = [\mathbf{E}_{\tau,v}; \mathbf{z}_\tau^{\text{hist}}]. \tag{2}$$

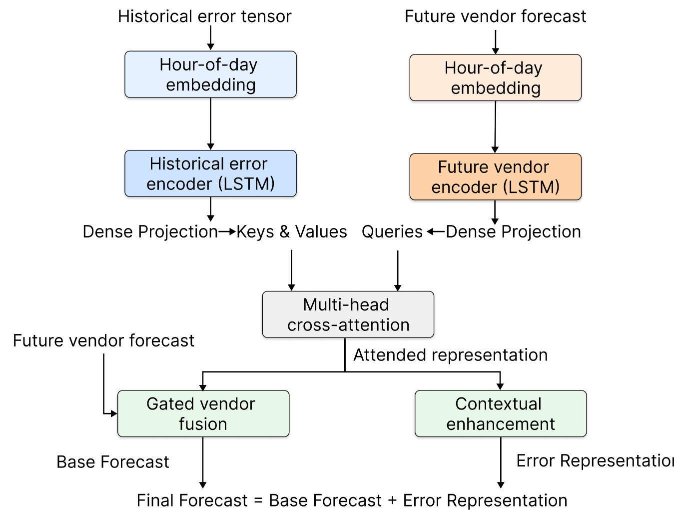

Fig. 1: Model architecture.

An LSTM encodes $\mathbf{X}^{\text{hist}}$ into hidden states $\mathbf{H}^{\text{mem}}$, which are then linearly projected to form attention keys and values:

$$\mathbf{K} = \mathrm{Dense}(\mathbf{H}^{\text{mem}}), \quad \mathbf{V} = \mathrm{Dense}(\mathbf{H}^{\text{mem}}). \tag{3}$$

This memory captures when and how each vendor tends to deviate from the truth.

c) Future Vendor Encoder.

Similarly, we concatenate future vendor forecasts with their time embeddings:

$$\mathbf{X}_h^{\text{fut}} = [\mathbf{F}_h; \mathbf{z}_h^{\text{fut}}],$$

encode them with an LSTM to get $\mathbf{H}^{\text{ven}}$, and then map to attention queries:

$$\mathbf{Q} = \mathrm{Dense}(\mathbf{H}^{\text{ven}}). \tag{4}$$

The queries express the informational needs of each future hour.

d) Multi-Head Cross Attention.

We apply multi-head attention to connect queries $\mathbf{Q}$ with keys and values $(\mathbf{K}, \mathbf{V})$:

$$\mathbf{M} = \mathrm{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}). \tag{5}$$

A residual connection to $\mathbf{H}^{\text{ven}}$ followed by layer normalization yields:

$$\tilde{\mathbf{M}} = \mathrm{LayerNorm}(\mathbf{M} + \mathbf{H}^{\text{ven}}). \tag{6}$$

This step links each future hour to the most relevant historical error patterns.

e) Gated Vendor Fusion.

For each hour h, we concatenate the attended representation, raw forecasts, and time embedding to form the fusion input:

$$\mathbf{u}_h = [\tilde{\mathbf{M}}_h; \mathbf{F}_h; \mathbf{z}_h^{\text{fut}}]. \tag{7}$$

The gating network (an MLP) transforms $\mathbf{u}_h$ into unnormalized gating logits:

$$\mathbf{g}_h = \mathrm{MLP}(\mathbf{u}_h), \tag{8}$$

which are then normalized via the softmax function to obtain vendor-specific trust weights:

$$\mathbf{w}_h = \mathrm{softmax}(\mathbf{g}_h), \quad \sum_{v=1}^{V} w_{h,v} = 1. \tag{9}$$

The base prediction is computed as a weighted combination of vendor forecasts:

$$\hat{y}_{t+h}^{\text{base}} = \sum_{v=1}^{V} w_{h,v} \, F_{h,v}. \tag{10}$$

979-8-3315-8138-1

f) Contextual Enhancement.

To correct for systematic shifts, we average $\tilde{\mathbf{M}}$ over the 24 hours:

$$\mathbf{c} = \frac{1}{T_f} \sum_{h=1}^{T_f} \tilde{\mathbf{M}}_h, \tag{11}$$

pass through a two-layer MLP, and bound with tanh:

$$\mathbf{e} = \tanh(\mathbf{W}_2\, \sigma(\mathbf{W}_1 \mathbf{c})) \tag{12}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are weight matrices, and $\sigma(\cdot)$ denotes an elementwise activation function. The final forecast is:

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}^{\text{base}} + \mathbf{e}. \tag{13}$$

This bounded residual ensures stability while enabling day-level adjustments.

III. Results

This section reports the evaluation results of the proposed LAF-Net model using the MISO day-ahead load forecasting dataset. The model is trained and tested for each of the 38 LBAs individually, and its performance is compared with two vendor baselines across LBA-level, regional, and system-wide aggregations. Evaluation metrics include MAE, MAPE, and RMSE. The proposed model contains a total of 90,778 trainable parameters. Given the current data scale, the training process for a single LBA-specific model can be completed within 30 minutes. During inference, due to the relatively compact model size, the model is able to generate the next-day 24-hour forecasts for each LBA within one second.

A. Case Studies

The proposed LAF-Net model is evaluated using real-world data from the Midcontinent Independent System Operator (MISO), which operates a large-scale power system geographically divided into 38 LBAs across three major regions: Central, North, and South. The training data span from May 12, 2024 to March 1, 2025, and the testing period covers March 2 to March 28, 2025. For each LBA, an independent LAF-Net model is trained and validated, resulting in a total of 38 models corresponding to all LBAs. During each day within the study period, MISO performs day-ahead hourly load forecasting at 2 PM, providing next-day 24-hour forecasts from two vendors along with the realized load observations. This setup enables a comprehensive evaluation of the proposed model's ability to dynamically fuse vendor forecasts and improve prediction accuracy across diverse regions and operating conditions.

B. Overall Forecasting Performance Across All LBAs

In Table I, the first column (LBA) is indexed using labels L1–L38. The table reports the mean absolute error (MAE), mean absolute percentage error (MAPE) and Root Mean Square Error (RMSE) across 38 load balancing areas (LBAs), comparing two vendor baselines (V1 and V2) with the proposed LAF-Net model. The results in Table I show that the proposed LAF-Net model consistently outperforms the best vendor across all 38 LBAs in terms of both MAE and MAPE. The improvements are substantial, with error reductions exceeding 30% in several cases such as L4 (35.1% in MAE, 35.3% in MAPE), L12 (31.7%, 32.4%), and L16 (37.5%, 37.8%). These findings highlight the effectiveness of LAF-Net in correcting systematic vendor biases and better capturing temporal load dynamics at the LBA level. For RMSE, LAF-Net also delivers almost universal gains, often with large margins. For example, L4 achieves a 34.1% reduction, L12 a 28.9% reduction, and L16 a 36.0% reduction compared to the best vendor. These results indicate that the model not only lowers average errors but also reduces the variance of extreme deviations, which is crucial for operational reliability. The only exception occurs at LBA L21, where LAF-Net does not yield a positive improvement in RMSE. This is primarily due to an unusual load pattern on the morning of March 5, when the highest load of the day occurred in the morning rather than the typical afternoon peak. As a result, LAF-Net's performance during that period was suboptimal. However, the difference relative to the best vendor is negligible (0.1%), effectively maintaining parity. Overall, these findings demonstrate that LAF-Net systematically improves vendor forecasts across multiple error metrics, with particularly strong benefits for LBAs with higher volatility and challenging demand patterns.

C. Region- and System-Level Aggregated Performance

Table II presents aggregated forecasting performance at the regional and system levels by summing the load across all 38 LBAs and then computing the corresponding MAE and MAPE. Results are reported for two vendor baselines (V1 and V2) alongside the proposed LAF-Net model. As shown, LAF-Net consistently outperforms the vendor forecasts at both the regional and system scales. In the Central region, MAE and MAPE are reduced by over 33%, while the North and South regions also benefit from notable improvements of 15–20%. At the system-wide level, where the load of all 38 LBAs is aggregated, the LAF-Net model achieves the largest performance gain, reducing MAE by 37.07% and MAPE by 37.38%. These results underscore the scalability of the proposed approach: not only does it deliver consistent accuracy improvements across diverse regions, but it also yields substantial system-level benefits, which are particularly critical for large-scale operational planning.

To further assess the model's generalization capability, we conducted an additional out-of-sample evaluation using an eight-week period from July 6 to August 30, 2025. Across all the LBAs, LAF-Net achieved an average MAE reduction of 26.38% compared to the best-performing vendor forecast. The MAE reduction ranged from a minimum of -1.49% to a maximum of 46.58% among individual LBAs. At the system level, the MAE of the best vendor forecast was 3,120.13, whereas LAF-Net reduced the MAE to 1,101.49, corresponding to a 60.77% reduction relative to the best vendor.

Table III summarizes forecasting performance at the regional and system levels during the daily peak load hour, defined as the hour with the highest actual load within each 24-hour horizon. As shown, LAF-Net achieves substantial accuracy gains at peak load hours across all regions, reducing MAE by 19–42% and MAPE by nearly 20–42%. At the system level, the improvement is most pronounced, with a 41.14%

979-8-3315-8138-1

TABLE I: Integrated LBA-level results across regions: LAF-Net vs. vendor baselines.

LBA	MAE (V1)	MAE (V2)	MAE (LAF)	MAPE (V1)	MAPE (V2)	MAPE (LAF)	RMSE (V1)	RMSE (V2)	RMSE (LAF)
L1	6.70	6.67	5.64	4.96%	4.94%	4.14%	8.76	8.65	7.36
L2	25.77	25.73	24.31	2.69%	2.69%	2.56%	32.49	32.38	29.69
L3	9.16	9.25	8.13	5.98%	6.06%	5.39%	11.80	11.89	10.24
L4	158.64	150.28	97.56	3.14%	2.98%	1.93%	196.47	186.68	122.97
L5	97.77	96.31	91.46	4.41%	4.33%	4.20%	128.16	127.09	116.00
L6	176.89	165.48	137.41	4.18%	3.91%	3.27%	222.95	209.68	171.47
L7	6.37	6.06	4.29	4.02%	3.83%	2.70%	7.91	7.56	5.66
L8	5.96	5.87	5.41	5.22%	5.14%	4.73%	7.41	7.37	6.67
L9	67.41	65.26	52.53	4.56%	4.43%	3.58%	87.28	83.84	67.93
L10	42.69	42.09	39.20	5.10%	5.02%	4.70%	52.63	52.07	49.15
L11	13.81	13.17	12.97	3.97%	3.78%	3.72%	16.88	16.25	16.06
L12	11.51	10.87	7.42	3.40%	3.20%	2.17%	14.07	13.42	9.55
L13	143.26	135.61	108.72	2.88%	2.72%	2.19%	186.55	174.35	139.15
L14	24.44	23.71	23.08	6.91%	6.75%	6.64%	30.93	30.00	28.41
L15	3.12	3.06	2.77	6.36%	6.26%	5.72%	4.20	4.15	3.68
L16	98.04	92.20	57.65	3.23%	3.04%	1.89%	122.17	116.51	74.58
L17	46.76	45.61	36.37	3.26%	3.18%	2.53%	58.66	57.64	45.77
L18	180.89	177.7	140.06	4.45%	4.37%	3.46%	221.53	217.71	179.09
L19	162.25	159.93	131.32	4.62%	4.55%	3.74%	207.00	204.26	168.52
L20	51.16	48.76	36.87	3.46%	3.31%	2.51%	64.21	62.04	47.82
L21	14.12	13.91	13.82	2.89%	2.84%	2.82%	19.23	19.11	19.14
L22	219.32	229.52	182.2	4.74%	4.93%	3.97%	279.71	291.19	232.78
L23	6.51	6.06	4.25	4.23%	3.96%	2.75%	8.08	7.50	5.47
L24	61.67	59.66	59.6	3.22%	3.11%	3.10%	78.90	77.59	77.39
L25	66.52	60.61	56.24	4.78%	4.33%	3.98%	85.34	75.27	71.51
L26	169.79	168.37	138.76	4.17%	4.13%	3.48%	212.98	212.84	179.17
L27	53.12	53.74	47.12	5.52%	5.59%	4.90%	66.38	67.27	60.24
L28	5.86	5.87	5.69	6.05%	6.08%	5.86%	7.30	7.37	7.27
L29	9.58	9.29	7.71	5.15%	5.02%	4.11%	12.33	11.82	10.06
L30	42.9	42.3	35.75	6.76%	6.71%	5.61%	52.75	51.68	44.38
L31	101.78	102.91	83.41	2.62%	2.65%	2.16%	129.02	129.10	103.87
L32	42.68	41.31	37.29	3.20%	3.09%	2.79%	53.66	51.43	45.60
L33	2.28	2.23	2.18	2.21%	2.17%	2.12%	2.92	2.87	2.79
L34	39.23	37.32	26.61	5.78%	5.47%	3.90%	48.07	45.09	34.30
L35	77.04	77.06	62.93	3.84%	3.84%	3.16%	94.53	94.62	78.09
L36	40.94	39.17	34.09	5.79%	5.55%	4.80%	51.52	49.54	43.72
L37	265.50	263.45	260.00	2.82%	2.80%	2.77%	330.29	327.65	322.73
L38	48.77	47.95	40.61	3.69%	3.63%	3.09%	63.83	62.72	52.90

TABLE II: Aggravated region and system-level results across regions: LAF-Net vs. vendor baselines.

Region	MAE (V1)	MAE (V2)	MAE (LAF)	MAPE (V1)	MAPE (V2)	MAPE (LAF)
Central	983.16	954.47	634.13	2.82%	2.74%	1.82%
North	492.00	491.55	395.00	2.95%	2.94%	2.37%
South	495.87	485.68	412.12	2.85%	2.80%	2.38%
System	1637.96	1618.24	1018.35	2.38%	2.35%	1.47%

TABLE III: Aggravated region and system-level results across regions peak load hour: LAF-Net vs. vendor baselines.

Region	MAE (V1)	MAE (V2)	MAE (LAF)	MAPE (V1)	MAPE (V2)	MAPE (LAF)	MBE (V1)	MBE (V2)	MBE (LAF)
Central	959.43	946.22	760.37	2.50	2.47	2.00	-808.93	-805.07	90.42
North	634.69	672.76	376.84	3.49	3.71	2.09	-630.79	-669.44	-202.18
South	578.60	543.14	313.19	2.97	2.78	1.62	-461.10	-429.99	-73.41
SYSTEM	1799.30	1801.87	1059.08	2.39	2.40	1.43	-1699.70	-1786.56	-38.47

reduction in MAE and a 40.36% reduction in MAPE compared to the best vendor forecast. Beyond accuracy, the bias results (MBE) demonstrate a marked correction of systematic underestimation present in vendor forecasts. While both vendors exhibit large negative biases at peak hours (e.g., -1699 MW and -1787 MW at the system level), LAF-Net effectively mitigates this issue, reducing the system-wide MBE to nearly unbiased levels (-38 MW). These results highlight the ability of LAF-Net not only to enhance point accuracy but also to substantially alleviate forecast bias during the most critical operating conditions, thereby improving reliability in system operations and market decisions.

D. *Detailed Analysis of Selected LBA*

To conduct a more in-depth evaluation, LBA L18, which represents the highest-load area within the system, is selected for detailed analysis. As illustrated in Figure 2, presents the forecasting results for LBA L18 from March 2 to March 8, 2025, comparing actual load profiles with forecasts from LAF-Net and vendor baselines. The proposed model demonstrates strong alignment with the observed values on March 2 to 4, where both morning load increases were captured almost perfectly. On March 5, the forecasts deviated more noticeably, with LAF-Net underestimating the mid-day peak; however, its errors remained smaller than those of the vendors. From March 5 and 6, the model provided consistently closer tracking of both daily peaks and valleys, outperforming vendors particu-

979-8-3315-8138-1

larly in high-load conditions where baseline forecasts tended to overpredict.

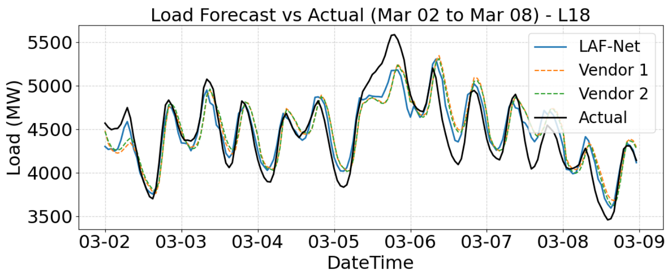

Fig. 2: Time series plot for L18.

Figure 3 presents the mean absolute error (MAE) by hour of day for the L18 LBA, comparing LAF-Net against two vendor baselines. It is observed from figure that LAF-Net achieves substantially lower error than both vendors during most hours of the day, with particularly pronounced gains around 06:00, where its error is more than 100 MW smaller than the vendors. At 08:00, however, the model underperforms slightly relative to the vendors, and similar but smaller deviations occur around 01:00 and 03:00. Beyond these isolated cases, LAF-Net consistently reduces forecasting error, especially during midday and evening demand peaks, highlighting its robustness in capturing daily load dynamics while mitigating vendor biases.

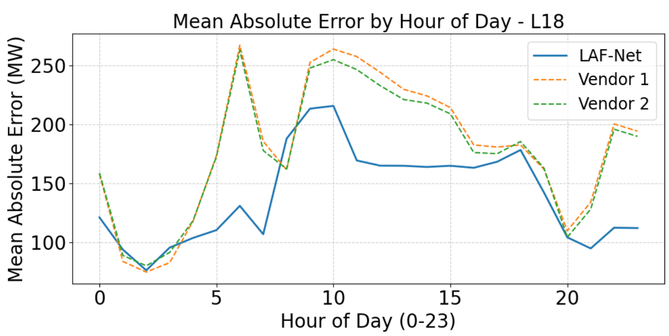

Fig. 3: Mean absolute error by hour of day.

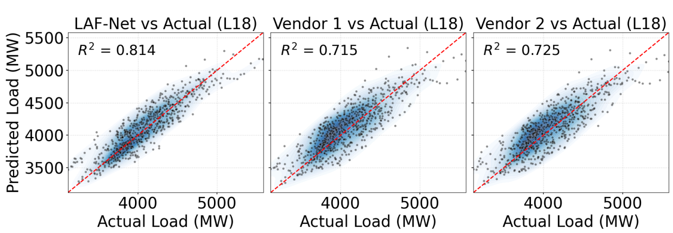

Fig. 4: Predicted vs actual.

The Figure 4 compares the predicted versus actual load for the L18 region using LAF-Net and two vendor forecasts. Among the three models, LAF-Net shows the tightest clustering of points around the red diagonal line, achieving the highest coefficient of determination ($R^2 = 0.814$) compared to Vendor 1 ($R^2 = 0.715$) and Vendor 2 ($R^2 = 0.725$). This indicates that LAF-Net provides a substantially better fit and lower prediction error overall. Moreover, it is clearly visible that when the actual load exceeds 5000 MW, LAF-Net's predictions remain close to the ideal line, while both vendors exhibit larger deviations, suggesting that LAF-Net captures high-load conditions more accurately.

IV. Conclusion and Future Work

This study proposed the LSTM–Attention Fusion Network with Error Representation (LAF-Net) to enhance day-ahead load forecasting by dynamically integrating multi-vendor predictions. Evaluated on real-world MISO LBA-level data, LAF-Net consistently outperformed the best vendor baselines across all 38 balancing areas, achieving an average MAE reduction of 14.9% and a system-level peak-hour error reduction exceeding 40%. We plan to conduct a comprehensive ablation analysis and report these results in our future work, as well as include evaluations on additional seasons to further assess cross-season generalization.

Despite these promising results, several limitations remain. The current model depends on supervised learning with historical vendor data and may face performance degradation under distributional shifts, such as unseen weather regimes or evolving demand patterns. Moreover, LAF-Net provides deterministic forecasts, which limits its applicability in uncertainty-aware operational planning. Future studies will explore probabilistic extensions, online adaptive learning, and joint modeling of renewable generation and load within the same attention-fusion framework to enhance robustness, interpretability, and real-time usability in large-scale power system operations.

References

[1] D. Yang and J. Kleissl, "Summarizing ensemble nwp forecasts for grid operators: Consistency, elicitability, and economic value," *International Journal of Forecasting*, vol. 39, no. 4, pp. 1640–1654, 2023.

[2] Z. Xiao, S.-J. Ye, B. Zhong, and C.-X. Sun, "Bp neural network with rough set for short term load forecasting," *Expert Systems with Applications*, vol. 36, no. 1, pp. 273–279, 2009.

[3] S. H. Rafi, S. R. Deeba, E. Hossain *et al.*, "A short-term load forecasting method using integrated cnn and lstm network," *IEEE access*, vol. 9, pp. 32 436–32 448, 2021.

[4] Z. Zhao, C. Xia, L. Chi, X. Chang, W. Li, T. Yang, and A. Y. Zomaya, "Short-term load forecasting based on the transformer model," *information*, vol. 12, no. 12, p. 516, 2021.

[5] R. Chandrasekaran and S. K. Paramasivan, "Advances in deep learning techniques for short-term energy load forecasting applications: A review," *Archives of Computational Methods in Engineering*, vol. 32, no. 2, pp. 663–692, 2025.

[6] W. Waheed and Q. Xu, "Data driven long short-term load prediction: Lstm-rnn, xg-boost and conventional models in comparative analysis," *Computational Intelligence*, vol. 41, no. 3, p. e70084, 2025.

[7] X. Hu, X. Tang, Q. Zhang, Z. Chen, and Y. Zhang, "Combined machine learning forecasting method for short-term power load based on the dynamic weight adjustment," *energy reports*, vol. 9, pp. 866–873, 2023.

[8] H. Ye, X. Teng, B. Song, K. Zou, M. Zhu, and G. He, "Multi-source data fusion-based grid-level load forecasting," *Applied Sciences*, vol. 15, no. 9, p. 4820, 2025.

[9] Y. Gou, C. Guo, and R. Qin, "Ultra short term power load forecasting based on the fusion of seq2seq bilstm and multi head attention mechanism," *Plos one*, vol. 19, no. 3, p. e0299632, 2024.

[10] F. Ziel, "Smoothed bernstein online aggregation for short-term load forecasting in ieee dataport competition on day-ahead electricity demand forecasting: Post-covid paradigm," *IEEE Open Access Journal of Power and Energy*, vol. 9, pp. 202–212, 2022.

[11] G. Lambert, B. Hamrouche, and J. De Vilmarest, "Frugal day-ahead forecasting of multiple local electricity loads by aggregating adaptive models," *Scientific Reports*, vol. 13, no. 1, p. 15784, 2023.

979-8-3315-8138-1