npj Urban Sustainability

https://doi.org/10.1038/s42949-026-00424-y

Article in Press

TAL-Net: a temporal attention LSTM framework for urban electricity and emissions forecasting in the U.S.

Received: 21 July 2025

Accepted: 6 June 2026

Published online: 17 July 2026

Cite this article as: Thomas, A.M., Dey, M., Rana, S.P. TAL-Net: a temporal attention LSTM framework for urban electricity and emissions forecasting in the U.S.. *npj Urban Sustain* (2026). https://doi.org/10.1038/s42949-026-00424-y

Ann Mary Thomas, Maitreyee Dey & Soumya Prakash Rana

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

TAL-Net: A Temporal Attention LSTM Framework for Urban Electricity and Emissions Forecasting in the U.S.

Ann Mary Thomas[1], Maitreyee Dey[1*], Soumya Prakash Rana[2*]

[1*]GENESIS Research Lab, London Metropolitan University, 166-220 Holloway Rd, N7 8DB, London, United Kingdom.
[2*]School of Engineering, University of Greenwich, Central Ave, Gillingham, ME4 4TB, Chatham, United Kingdom.

*Corresponding author(s). E-mail(s): m.dey@londonmet.ac.uk; s.rana@greenwich.ac.uk;
Contributing authors: annmarytttt@gmail.com;

Abstract

Urban sustainability depends critically on accurate forecasting of electricity demand and associated CO_2 emissions, particularly as cities strive to meet climate targets while ensuring energy resilience. However, existing approaches often model these two aspects in isolation, missing key interdependencies that influence urban energy dynamics. This study presents TAL-Net, a novel Temporal Attention LSTM Network designed for the joint forecasting of electricity demand and CO_2 emissions in complex urban energy systems. Leveraging deep learning and domain-informed feature engineering, we evaluate TAL-Net alongside four other state-of-the-art models using high-resolution data from two contrasting U.S. regions - California and Texas - characterized by differing climates, energy portfolios, and urban infrastructure. TAL-Net consistently achieves superior forecasting accuracy, demonstrating lower MAPE, MAE, and RMSE across both regions. Our findings highlight the value of attention mechanisms in capturing temporal patterns and cross-variable dependencies, offering AI-driven insights to guide urban energy planning, carbon mitigation strategies, and grid management in fast-evolving metropolitan contexts.

ARTICLE IN PRESS

Introduction

Urban areas are at the forefront of the climate crisis, driving both energy consumption and greenhouse gas emissions. As cities expand and electrify, particularly through transport and building sectors, there is an urgent need for intelligent tools to monitor, forecast, and manage energy demand and associated carbon emissions sustainably [1–3]. Among the technological enablers, Artificial Intelligence (AI) has emerged as a critical asset, offering sophisticated modeling capabilities that can support real-time decision-making, demand-side management, and climate resilience planning in complex urban environments [4–6].

Carbon dioxide (CO_2) remains the predominant greenhouse gas, primarily emitted from urban transportation, industrial activity, and fossil-fuel-based electricity generation [7–9]. The increasing electrification of cities and population-driven demand growth amplify the importance of accurate, forward-looking tools for managing both electricity usage and emissions [10]. However, regional heterogeneity in climate, infrastructure, and energy portfolios—such as high renewable penetration in California versus fossil fuel dominance in Texas—leads to significant spatial and temporal variability in both electricity demand and emissions intensity [11, 12].

Traditionally, electricity demand and CO_2 emissions have been forecasted separately, missing the critical interdependencies between the two [13, 14]. This siloed approach limits the ability of urban policymakers and grid operators to develop integrated strategies that align energy reliability with decarbonisation goals. Simultaneous forecasting, therefore, is not merely a technical improvement—it is a strategic imperative for urban energy governance, facilitating emissions-aware generation planning, proactive load balancing, and more equitable urban sustainability transitions [15–17].

Previous research addressed demand and emissions forecasting separately primarily because they were traditionally viewed as independent planning objectives-demand forecasting focused on ensuring supply adequacy, while emissions forecasting aimed at regulatory compliance and environmental reporting. However, this separation overlooks the fundamental physical coupling between these variables, as CO_2 emissions are directly influenced by both the level of electricity demand and the fuel mix used to meet that demand at any given hour. By treating them jointly, our approach explicitly models this interdependence, enabling more coherent and actionable forecasts for integrated energy-environmental planning.

Advanced computational models are essential for managing sustainable urban energy systems, particularly as cities aim to decarbonise while meeting growing electricity demands. While numerous studies have applied deep learning—such as Long Short-Term Memory (LSTM) networks—for electricity demand forecasting [18, 19] and regional CO_2 emissions prediction [20–22], these efforts typically treat the two problems independently. This separation limits their effectiveness in urban decision-making, where emissions and electricity use are tightly coupled across sectors like transportation, industry, and residential buildings [23].

To improve forecasting accuracy and capture more complex temporal dynamics, hybrid deep learning architectures have been introduced. Techniques such as CNN-LSTM and RBM-LSTM networks have shown promise in energy modeling by combining spatial and sequential feature learning capabilities [24–26]. Despite these

ARTICLE IN PRESS

ARTICLE IN PRESS

advances, relatively few studies tackle jointforecasting of electricity demand and emissions, and even fewer do so in a way that reflects regional adaptability to differences in climate, infrastructure, and energy policy.

Recent reviews emphasize AI's growing potential to enhance sustainability outcomes through improved demand-side efficiency, renewable integration, and decarbonisation analytics [27, 28]. However, there remains a critical gap in models that offer integrated, regionally flexible forecasting tools suitable for urban energy governance.

Despite progress in AI-driven energy forecasting, most existing studies model either electricity demand or CO_2 emissions in isolation. The simultaneous forecasting of both variables, especially across regions with contrasting geography and energy mixes, such as California and Texas, remains an underexplored area in the literature.

To address the gap in simultaneous forecasting of electricity demand and CO_2 emissions, this study proposes a Temporal Attention LSTM Network (TAL-Net) as a robust and regionally adaptable deep learning solution. Building on prior work on hybrid AI models for power system forecasting and anomaly detection [29–32], TAL-Net captures nonlinear dependencies and time-varying patterns in urban energy dynamics. We systematically evaluate five deep learning architectures, using data from two contrasting U.S. states, California and Texas, to benchmark forecasting accuracy and adaptability.

This study is among the first to jointly model electricity demand and CO_2 emissions using a comprehensive regional dataset, capturing their interdependent temporal and contextual dynamics. It advances the field through sophisticated feature engineering and a systematic evaluation of five deep learning architectures: RNN, LSTM, RBM-MLP, RBM-LSTM, and the proposed TALNet (Temporal Attention LSTM Network). TAL-Net leverages attention mechanisms to effectively model complex temporal dependencies, consistently outperforming baseline models. Collectively, these contributions offer a unified, high-performing framework that supports sustainable regional energy planning through accurate, integrated forecasts.

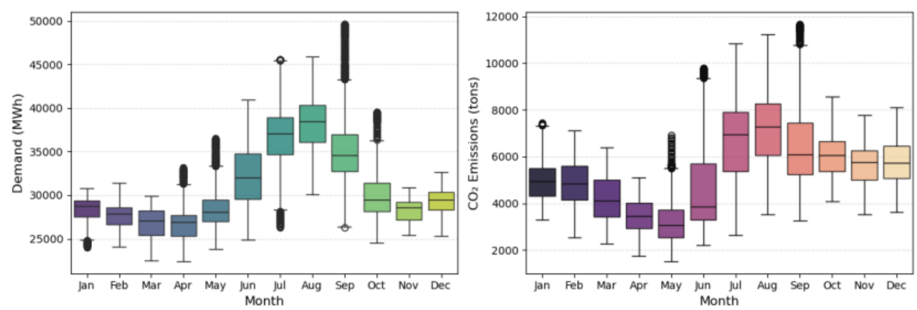
Results

This result section aims to provide an empirical analysis of the data to find the patterns of electricity demand and CO_2 emissions in CAL and TEX, and evaluate the performance of the proposed hybrid TAL-Net (Temporal Attention LSTM Network) model alongside four implemented baseline models, LSTM, RNN, RBM+MLP, and RBM+LSTM. The results are reported for both CAL and TEX, and experimented across different data partitioning schemes to assess the models' robustness and generalisability.

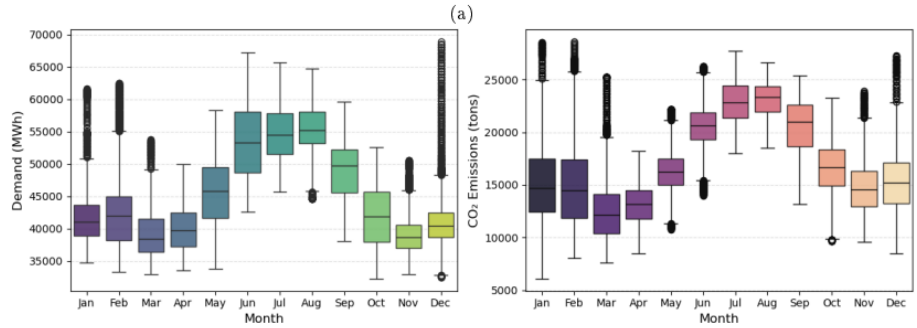
ARTICLE IN PRESS

Empirical Data Analysis

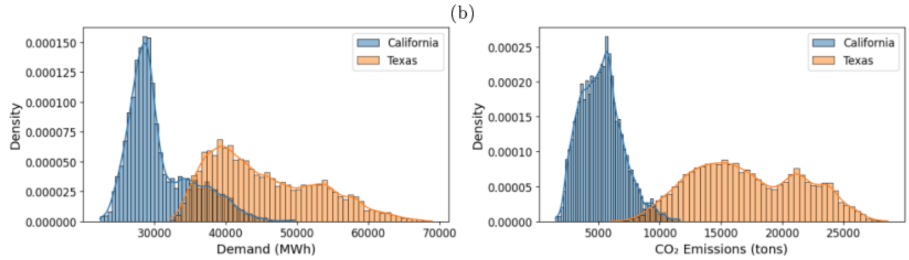
To understand the characteristics of electricity demand and CO_2 emissions in CAL and TEX, a descriptive analysis was conducted. Specifically, Figure 1(a) presents side-by-side boxplots of monthly electricity demand (in MWh, left panel) and CO_2 emissions (in tons, right panel) for California, Figure 1(b) presents the corresponding boxplots for TEX, and Figure 1(c) displays overlaid probability density histograms comparing the distributions of electricity demand (left panel) and CO_2 emissions (right panel) between CAL (blue) and TEX (orange) using the full hourly dataset.



(a)



(b)



(c)

Fig. 1 Distribution analysis of the five-year monthly average electricity demand and CO_2 emissions, highlighting seasonal and regional variability: (a) box plots of electricity demand and CO_2 emissions for CAL, (b) box plots for TEX, and (c) probability density distributions of electricity demand (left) and CO_2 emissions (right) comparing CAL and TEX using hourly data from July 2018 to June 2023.

The boxplots in Figure 1(a) and 1(b) reveal clear seasonal and regional differences for CAL and TEX respectively. CAL exhibits lower electricity demand overall, with values concentrated between approximately 22,000 and 50,000 MWh, and a median of about 29,238 MWh. In contrast, TEX displays higher demand, ranging from 32,000 to nearly 69,000 MWh, and a higher median of approximately 43,695 MWh. Both

ARTICLE IN PRESS

regions show clear seasonal peaks during summer months (July–August), consistent with higher cooling loads. The CO_2 emissions also follow a similar seasonal pattern but with a much higher magnitude in TEX. The median CO_2 emissions in TEX (∼16,715 tons) are over three times higher than in CAL (∼5,182 tons), reflecting the heavier reliance on fossil fuels in TEX's generation mix. The interquartile ranges and presence of outliers also highlight greater variability in TEX, particularly for emissions.

The distribution plots in Figure 1(c) provide further insight into the underlying statistical properties of the target variables. In both regions, electricity demand distributions are approximately unimodal and slightly right-skewed, whereas CO_2 emissions distributions are more concentrated and exhibit sharper peaks. This indicates that while demand and emissions are correlated, their variability and scaling differ across regions. The histograms confirm that CAL's emissions are concentrated at lower values, while TEX's emissions are more dispersed and substantially higher overall.

Table 1 Summary statistics of electricity demand and CO_2 emissions for CAL and TEX.

Variable	**Statistic**	**CAL**	**TEX**
Demand	Mean	30,713	45,380
Demand	Std Dev	4,601	7,549
Demand	Min	22,375	32,185
Demand	25%	27,572	39,215
Demand	Median	29,238	43,695
Demand	75%	33,393	51,198
Demand	Max	49,633	68,943
CO_2 Emissions	Mean	5,254	17,200
CO_2 Emissions	Std Dev	1,729	4,453
CO_2 Emissions	Min	1,508	6,050
CO_2 Emissions	25%	3,947	13,669
CO_2 Emissions	Median	5,182	16,715
CO_2 Emissions	75%	6,290	20,967
CO_2 Emissions	Max	11,665	28,619

ARTICLE IN PRESS

These findings are consistent with the summary statistics presented in Table 1. On average, TEX's electricity demand (mean: 45,380 MWh) exceeds CAL's (mean: 30,713 MWh) by nearly 50%, and the mean CO_2 emissions in TEX (17,200 tons) are more than three times higher than in CAL (5,254 tons). Standard deviations in both demand and emissions are also notably larger in TEX, indicating higher volatility. This empirical analysis underscores the contrasting operational and environmental profiles of the two regions, justifying the need for a flexible forecasting model capable of adapting to such diverse patterns.

While per-capita or per-customer normalization could provide additional comparative insights into regional consumption patterns, this study intentionally uses aggregate regional demand values because the primary objective is forecasting total grid-level electricity demand and CO_2 emissions for operational planning purposes. Grid operators and policymakers require absolute demand forecasts (in MWh) to make

ARTICLE IN PRESS

decisions about generation dispatch, capacity planning, and emissions management. The observed differences between CAL and TEX − including TEX's higher absolute demand and emissions − reflect real operational challenges that the forecasting framework must address, regardless of underlying population differences.

Optimal Data Partitioning Selection

In this study, various configurations of training epochs and train–test splits were explored to identify optimal settings for accurate forecasting. All five models were initially trained with epoch values from 1 to 15 to observe performance trends. Across models and target variables, 10 epochs consistently yielded the lowest error metrics, indicating stable convergence and better generalisation; this was adopted as the standard.

Experiments with five train–test split ratios (80–20, 75–25, 70–30, 65–35, and 60–40) were then conducted to evaluate model robustness and identify the optimal split for achieving the most accurate forecasts by balancing sufficient training data with reliable test evaluation. This approach assessed each architecture's sensitivity to training data quantity and helped identify bias–variance trade-offs in prediction.

The performance evaluation of all five compared models, using MAE, RMSE, and MAPE, was computed over five train-test splits (80-20, 75-25, 70-30, 65-35, and 60-40). The results for CAL and TEX are summarised in Tables 2 and 3, respectively.

Table 2 details the performance metrics for CAL. Across all train-test splits, the TAL-Net model consistently achieves the lowest MAE, RMSE, and MAPE values for both electricity demand and CO_2 emissions, with the best results observed under the 80-20 split: MAPE of 0.40% for demand and 3.28% for emissions. The standard LSTM model performs comparably, but slightly worse than the hybrid, while the RNN, RBM+MLP, and RBM+LSTM models show noticeably higher error rates, particularly for CO_2 emissions. The trend suggests that attention mechanisms enhance the LSTM's ability to model the complex temporal patterns present in CAL's energy data.

Similarly, Table 3 presents the metrics for TEX. Again, the TAL-Net outperforms the alternatives, achieving the lowest MAPE values of 0.49% for demand and 4.91% for emissions in the 80-20 split. The standard LSTM ranks second, while RNN, RBM+MLP, and RBM+LSTM continue to exhibit higher errors. The consistent superiority of the hybrid model across both regions highlights its robustness to different data characteristics.

Notably, it has been observed that the 80–20 split generally achieved the best balance of predictive accuracy and model stability across most models and both regions, particularly for the LSTM+Attention model. Wider splits (such as 60–40) tended to slightly degrade performance due to reduced training data, while narrower splits (such as 90–10, which have not been included here) could risk overfitting. By systematically recording the results for each model under each split ratio, it was ensured that the findings are robust and comparable. This approach also highlights the importance of empirical evaluation of partitioning strategies, especially in time-series contexts where temporal dependency must be preserved in the split.

ARTICLE IN PRESS

Table 2 Performance metrics of all models for electricity demand and CO_2 emissions forecasting in CAL under different train–test splits

California							
Model	**Train–Test Split (%)**	**MAE (ED)**	**RMSE (ED)**	**MAPE (ED) (%)**	**MAE** CO_2	**RMSE** CO_2	**MAPE** CO_2 **(%)**
LSTM	80–20	157.98	187.81	0.51	210.03	249.34	3.94
	75–25	213.72	267.01	0.69	254.63	283.96	5.58
	70–30	258.83	304.14	0.87	281.99	324.09	6.28
	65–35	128.28	182.63	0.43	400.34	440.54	8.42
	60–40	127.66	160.90	0.42	378.21	430.75	7.99
RNN	80–20	1191.49	1703.24	3.81	592.82	759.00	10.53
	75–25	1316.87	1813.52	4.12	582.79	730.21	11.00
	70–30	1604.00	2037.28	5.21	508.32	657.08	10.36
	65–35	1380.25	1780.28	4.54	486.21	609.31	10.77
	60–40	1558.59	1982.83	4.98	516.73	686.34	9.78
RBM+MLP	80–20	1605.10	2494.99	4.97	827.21	1113.51	15.90
	75–25	1570.84	2334.83	4.95	827.58	1117.26	15.77
	70–30	1689.90	2487.90	5.33	804.41	1080.12	16.76
	65–35	1715.42	2533.00	5.53	805.30	1077.85	17.01
	60–40	1645.90	2388.70	5.18	882.22	1171.79	17.49
RBM+LSTM	80–20	1571.74	2338.56	4.96	807.43	1122.13	14.88
	75–25	1720.83	2501.35	5.41	782.05	1103.75	14.70
	70–30	1508.07	2206.12	4.84	732.60	1017.46	15.14
	65–35	1668.79	2329.49	5.35	797.96	1062.31	16.05
	60–40	1692.55	2401.91	5.39	884.67	1173.03	17.57
TAL-Net	80–20	127.58	206.07	0.40	163.21	200.51	3.28
	75–25	232.49	301.62	0.79	234.43	278.95	4.79
	70–30	279.27	337.39	0.93	374.04	415.44	8.39
	65–35	380.06	433.64	1.28	411.31	457.32	9.03
	60–40	303.95	390.08	0.96	398.96	474.32	8.33

These experiments validate that careful selection of both training epochs and data splits is crucial for achieving reliable forecasts and avoiding overfitting or under-training, particularly in the context of multivariate energy demand and emissions forecasting.

Performance Assessment Across Compared DL Models

This section presents the performance of all five models, LSTM, RNN, RBM+MLP, RBM+LSTM, and TAL-Net, in forecasting electricity demand and CO_2 emissions across CAL and TEX. Since the 80-20 train-test split consistently produced the best results across all models, the time series plots are shown only for this split for clarity and comparability. The results are reported in terms of actual versus predicted time series plots and error metrics (MAE, RMSE, MAPE). For each model, the performance trends for the testing data are visualized in Figures 2 and 3 for CAL and TEX, respectively.

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 3 Performance metrics of all models for electricity demand and CO_2 emissions forecasting in Texas under different train–test splits

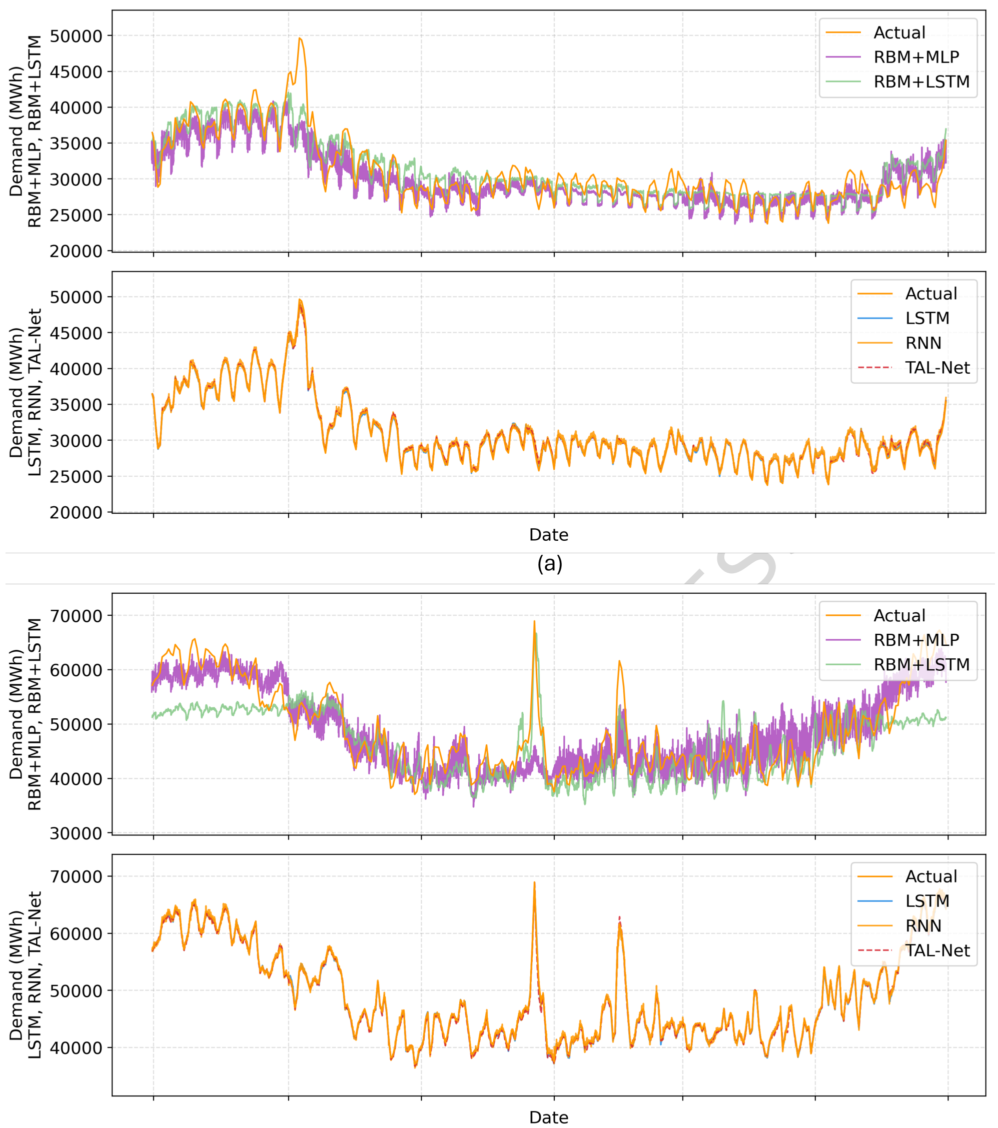
Model	Train–Test Split (%)	MAE (ED)	RMSE (ED)	MAPE (ED) (%)	MAE CO_2	RMSE CO_2	MAPE CO_2 (%)
Texas							
LSTM	80–20	257.16	367.95	0.50	824.10	1043.31	5.96
	75–25	270.54	360.42	0.54	1072.51	1290.23	7.26
	70–30	325.01	426.73	0.66	1145.66	1391.53	7.21
	65–35	454.97	606.66	0.94	1076.15	1327.05	7.15
	60–40	551.74	825.55	1.05	1129.08	1397.20	7.12
RNN	80–20	2581.46	3362.35	5.29	1430.50	1821.62	9.97
	75–25	3267.01	4075.21	6.32	1463.40	1757.34	9.91
	70–30	3304.38	4199.38	6.49	1315.33	1638.75	8.65
	65–35	3303.87	4104.13	6.73	1309.77	1680.49	8.71
	60–40	3300.29	4136.89	6.78	1447.66	1889.57	8.86
RBM+MLP	80–20	3659.13	4875.72	7.15	2422.79	3209.02	17.84
	75–25	6280.16	7479.84	12.15	2052.47	2696.38	13.92
	70–30	5936.98	7132.35	11.69	1855.79	2465.81	12.13
	65–35	5422.93	6525.97	10.93	1741.58	2284.96	11.27
	60–40	4705.56	6022.78	9.28	1924.64	2494.46	12.57
RBM+LSTM	80–20	3291.62	4313.33	6.34	1937.82	2439.56	14.14
	75–25	3824.29	4769.46	7.37	1897.67	2418.63	13.46
	70–30	4041.31	4944.13	8.06	2089.10	2871.06	15.05
	65–35	3867.23	4822.78	7.81	2434.23	3062.52	17.46
	60–40	3700.46	4552.17	7.53	2508.22	3143.74	17.22
TAL-Net	80–20	237.37	310.09	0.49	674.58	869.94	4.91
	75–25	885.03	1527.72	1.58	900.23	1149.90	6.43
	70–30	546.91	717.75	1.09	1384.86	1665.17	9.01
	65–35	411.17	542.81	0.87	751.93	971.10	5.39
	60–40	724.11	1118.53	1.37	860.67	1053.23	5.79

Figures 2 and 3 present combined comparison plots of actual versus predicted values for all five models over the test period (July 2022 to July 2023), for the CAL and TEX regions respectively. Each figure is split into two panels to accommodate the large difference in prediction accuracy between model groups: the upper panel shows RBM+MLP and RBM+LSTM predictions alongside the actual values, while the lower panel shows LSTM, RNN, and TAL-Net (dashed line) on a finer scale. The x-axis represents the hourly date range and the y-axis indicates electricity demand (MWh) or CO_2 emissions (tons). Actual values are shown in orange for demand and green for CO_2 emissions. These plots demonstrate that LSTM-based models, particularly TAL-Net, achieve superior alignment with actual data compared to RNN and RBM-based models, confirming the benefits of advanced recurrent and attention-based architectures.

Figure 2 presents the combined demand forecasting comparison for the CAL region (a) and TEX region (b). In the upper panel of each figure, t he R BM+MLP and RBM+LSTM predictions are shown against the actual demand, revealing substantial

ARTICLE IN PRESS

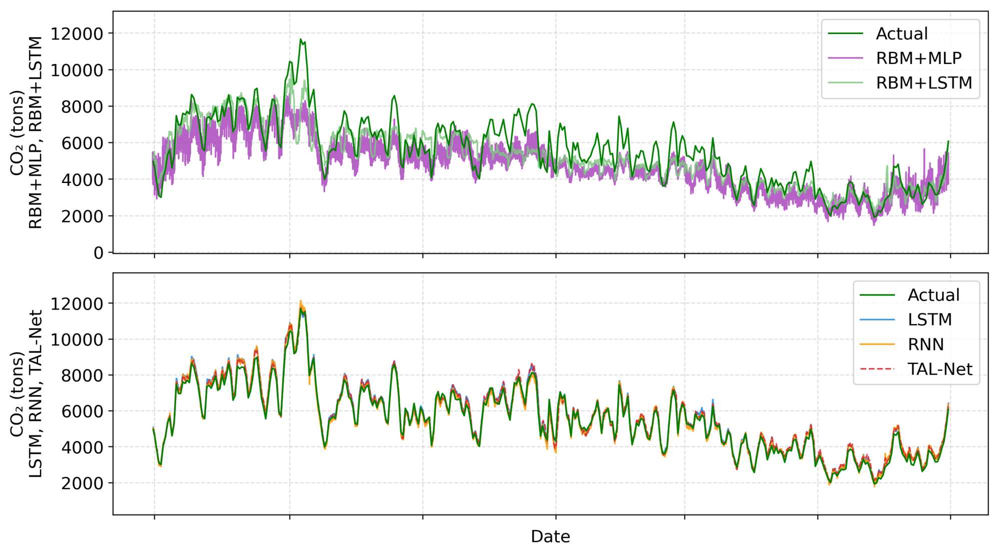
ARTICLE IN PRESS



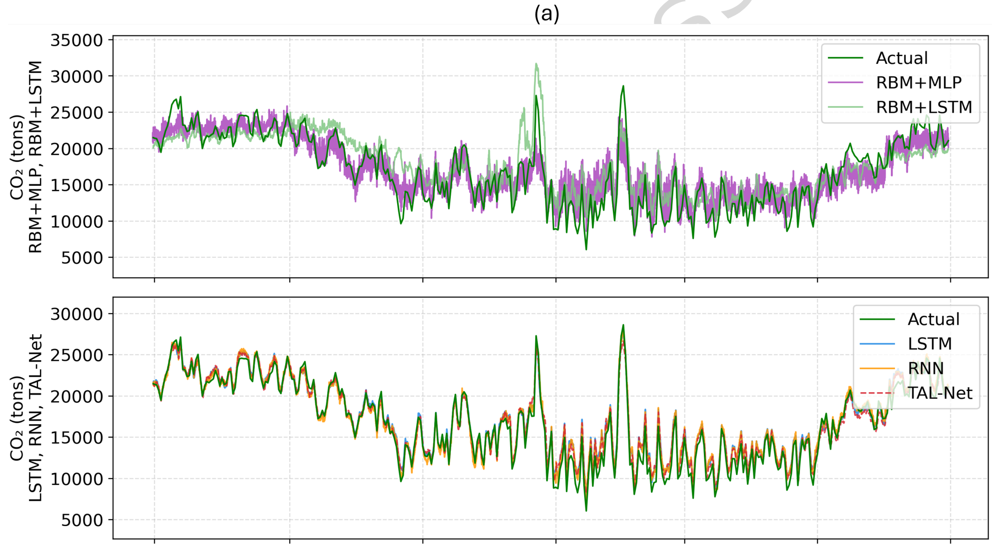
(b)

Fig. 2 Comparison of actual versus predicted electricity demand for all five models over the test period (July 2022–July 2023): (a) CAL region, (b) TEX region. Each figure is split into two panels: the upper panel shows RBM+MLP and RBM+LSTM predictions alongside the actual values (orange line); the lower panel shows LSTM, RNN, and TAL-Net (dashed red line) on a finer scale. TAL-Net and LSTM maintain the closest alignment with actual demand, while RBM-based models exhibit substantially larger deviations.

ARTICLE IN PRESS



(a)



(b)

Fig. 3 Comparison of actual versus predicted CO_2 emissions for all five models over the test period (July 2022–July 2023): (a) CAL region, (b) TEX region. Each figure is split into two panels: the upper panel shows RBM+MLP and RBM+LSTM predictions alongside the actual values (green line); the lower panel shows LSTM, RNN, and TAL-Net (dashed red line) on a finer scale. TAL-Net and LSTM achieve the closest tracking of actual emissions, while RBM-based models show substantially higher prediction errors particularly during high-emission episodes.

ARTICLE IN PRESS

deviations particularly during peak periods. Both models follow the general seasonal trend but exhibit persistent over- and underestimation at key demand events, reflecting the limitations of these architectures in capturing short-term temporal dynamics. The lower panel focuses on LSTM, RNN, and TAL-Net, where the finer y-axis scale makes the differences between these higher-performing models clearly visible. LSTM maintains close alignment with actual values throughout the test period. RNN follows the overall trend but shows greater deviation during peaks, indicating weaker handling of long-term dependencies. TAL-Net (dashed line) achieves the tightest alignment with actual demand in both regions, accurately capturing seasonal, daily, and peak fluctuations.

Figure 3 presents the corresponding CO_2 emissions comparison for the CAL region (a) and TEX region (b). The upper panel confirms that RBM+MLP and RBM+LSTM exhibit considerably higher prediction errors, particularly during high-emission episodes driven by abrupt changes in generation mix or load. In the lower panel, LSTM and TAL-Net both maintain close tracking of actual emissions, with TAL-Net showing the lowest overall deviation. These findings highlight the effectiveness of the attention mechanism in dynamically weighting the most relevant timesteps and improving focus on informative temporal patterns in both demand and emissions forecasting.

Although the RNN predictions visible in the lower panels of Figure 2 appear to visually track the general demand trend, the time series plots can be deceptive at this scale. The RNN captures the overall seasonal pattern but exhibits consistent systematic bias and fails to accurately predict short-term fluctuations and peaks. The quantitative metrics in Tables 2-3 reveal this discrepancy: the RNN's MAPE of 3.81–5.29% indicates prediction errors averaging several percentage points at each timestep, which accumulate substantially over the test period despite the visually similar trend. In contrast, the TAL-Net achieves MAPE values below 0.5%, indicating much tighter alignment with actual values at the hourly level that is difficult to distinguish visually at the full time series scale.

Figure 4 shows the MAPE results for the 80–20 split in both regions and target variables. Each subfigure displays a bar chart comparing the MAPE (%) achieved by all five models, with each model represented by a distinct colour. Figure 4(a) shows the demand forecasting comparison for California, Figure 4(b) shows the CO_2 emissions forecasting comparison for California, Figure 4(c) shows the demand forecasting comparison for Texas, and Figure 4(d) shows the CO_2 emissions forecasting comparison for Texas. The MAPE value for each model is annotated above its corresponding bar. The bar plots highlight the performance gap between the hybrid model and the others, especially in CO_2 emissions forecasting, where it achieves markedly lower percentage errors.

In CAL, the TAL-Net model achieves a MAPE of 0.40% for electricity demand and 3.28% for CO_2 emissions, outperforming the standard LSTM by 0.1–0.6 percentage points. Similar results are observed in TEX, with MAPE values of 0.49% for demand and 4.91% for emissions. These results demonstrate the proposed TAL-Net model's ability to capture long-term dependencies and focus on relevant temporal patterns, consistently delivering the most accurate forecasts across both regions.

ARTICLE IN PRESS

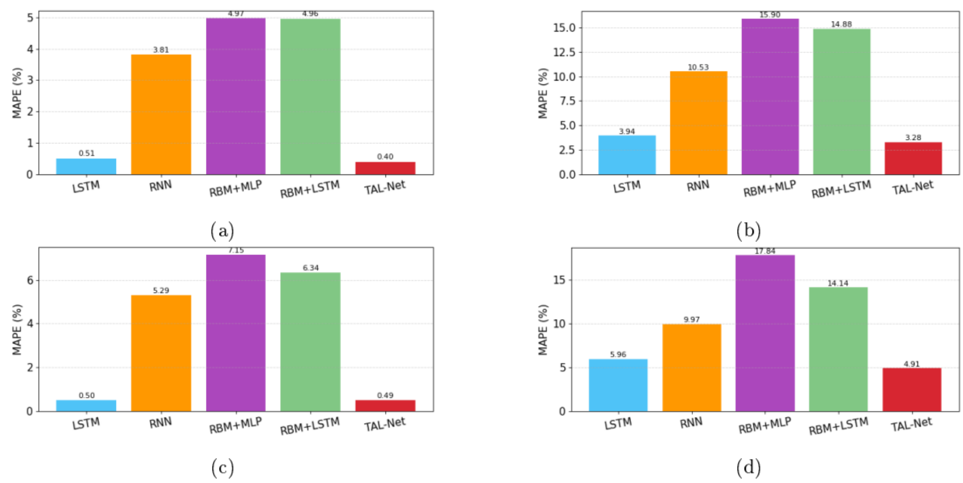


Fig. 4 Comparative analysis of MAPE (%) for electricity demand and CO_2 emissions forecasts in California and Texas using an 80-20 split: (a) California Demand Forecasting MAPE Comparison, (b) California CO_2 Emissions Forecasting MAPE Comparison, (c) Texas Demand Forecasting MAPE Comparison, and (d) Texas CO_2 Emissions Forecasting MAPE Comparison.

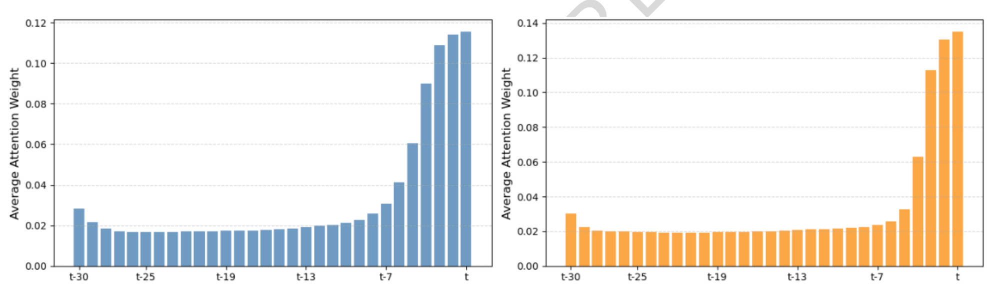


Fig. 5 Average attention weights learned by TAL-Net across the 30-hour input window for CAL (left) and TEX (right).

Figure 5 shows higher weights indicate greater importance assigned to that timestep for next-hour prediction.The figure displays two bar charts showing the average attention weight (y-axis) assigned to each of the 30 input timesteps (x-axis, labelled from $t-30$ to t) for CAL (left, blue bars) and TEX (right, orange bars). Both regions exhibit highest attention on the most recent timesteps (t *to* $t-5$), confirming that near-term patterns are most informative for forecasting. TEX shows a sharper concentration on immediate history, while CAL maintains slightly more distributed attention across the window. The non-zero baseline weights for earlier timesteps indicate the model retains capacity to capture longer-range temporal dependencies when relevant.

To provide interpretability of the TAL-Net model, we visualised the learned attention weights across the 30-hour input window (Figure 5). The attention mechanism

ARTICLE IN PRESS

ARTICLE IN PRESS

assigns substantially higher weights to the most recent timesteps, with the highest weight ($\sim 0.12 - 0.14$) at timestep t (the hour immediately preceding the prediction). This confirms that near-term demand and emissions patterns are most predictive of the next hour's values. Interestingly, both regions also show elevated weights around $t - 30$, suggesting the model captures daily periodicity effects. The consistent attention patterns across CAL and TEX demonstrate that TAL-Net learns generalisable temporal importance structures despite the regions' differing energy profiles.

Discussion

This study evaluated multiple deep learning architectures for jointly forecasting electricity demand and CO_2 emissions in two contrasting U.S. regions, CAL and TEX. Among the five models compared, the proposed TAL-Net architecture consistently achieved the lowest MAE, RMSE, and MAPE values across both regions and train–test splits, demonstrating clear superiority in predictive performance.

While the absolute MAPE improvement of approximately 0.1–1% may appear modest, it is important to contextualize this in the domain of electricity demand forecasting, where even small percentage reductions translate to substantial operational savings - a 0.1% improvement in demand forecasting for CAL's 30,000+ MWh average hourly demand represents roughly 30 MWh per hour, which accumulates to significant cost and emissions implications over time. Furthermore, the attention mechanism provides an interpretable means of identifying which historical timesteps most influence predictions, offering insights that a standard LSTM cannot provide. The consistency of TAL-Net's superior performance across both regions and all train–test splits demonstrates its robustness, whereas LSTM's performance varied more substantially under different data partitioning conditions.

TAL-Net's enhanced performance, which combines LSTM's capacity to model long-range temporal dependencies with a dynamic attention mechanism that identifies and prioritises the most relevant time steps. This enables the model to adapt effectively to regional variations in energy mix, climate, and demand profiles. Notably, it maintains low error rates even with relatively limited training data, underscoring its robustness and applicability to real-world urban energy systems.

Our results show that TAL-Net not only improves demand forecasting—particularly in capturing seasonal peaks and demand shocks—but also significantly reduces errors in CO_2 emissions predictions. This is especially relevant in contexts where emissions variability is driven by complex, fuel-specific generation patterns and consumption dynamics.

These findings have meaningful implications for urban sustainability. For grid operators, city planners, and policymakers, accurate joint forecasts of electricity demand and emissions offer a data-driven foundation for optimising generation planning, enhancing grid resilience, and achieving decarbonisation targets. By integrating forecasting into urban energy governance, AI models like TAL-Net can facilitate the strategic deployment of renewables, emissions-aware demand management, and more equitable, sustainable urban energy transitions.

ARTICLE IN PRESS

It is important to consider whether the spatial scale of these forecasts aligns with the scale at which policy decisions are implemented. The CAL region aggregates multiple balancing authorities spanning the state, while TEX corresponds to the ERCOT balancing authority. Forecasts at this regional scale are most directly useful for bulk power system operators and state-level energy agencies making generation dispatch and capacity planning decisions. Translating these forecasts to city or district-level policy actions - such as municipal carbon budgeting or neighbourhood demand response - would require further spatial disaggregation, which represents an avenue for future work. Additionally, both regions are embedded within larger interconnections: CAL is part of the Western Interconnection and exchanges significant power with neighbouring regions, while TEX (ERCOT) operates with limited interstate ties. The current model does not explicitly account for inter-regional power flows, which may affect both demand balance and emissions intensity. Incorporating interchange data in future work could improve forecast accuracy and policy relevance.

To assess the relative contributions of feature engineering versus model architecture, an ablation study was conducted comparing TAL-Net and LSTM under three feature configurations: Full (27 features), Basic (10 temporal/seasonal features only), and No Energy Mix (22 features excluding generation mix variables). Table 4 presents the results.

The ablation analysis reveals that feature engineering contributes substantially more to predictive performance than architectural differences between LSTM and TAL-Net. When using only basic temporal features, both models exhibit degraded performance (MAPE 5–9% for demand, 15–18% for CO_2), demonstrating that the engineered lag, rolling, and energy mix features are essential for accurate forecasting. Notably, removing energy mix features has a disproportionately large impact on CO_2 emissions forecasting, increasing MAPE from 3.41% to 16.98% in CAL, confirming that generation fuel composition is critical for emissions prediction.

With the full feature set, TAL-Net and LSTM achieve comparable demand forecasting accuracy, but TAL-Net shows consistent advantages in CO_2 emissions prediction (3.41% vs 4.48% in CAL). Furthermore, TAL-Net demonstrates greater robustness under feature-limited conditions, suggesting that the attention mechanism helps extract more information from constrained inputs. These findings indicate that while feature engineering is the primary driver of performance, the attention mechanism provides additional value, particularly for the more complex emissions forecasting task.

In Table 4 shows, "Full (27)" includes all engineered features; "Basic (10)" contains only temporal and seasonal features; "No Energy Mix (22)" excludes generation fuel composition features. Bold values indicate best performance for each region-target combination.

To justify the selection of the 30-hour input window, sensitivity experiments were conducted with window sizes of 6, 12, 24, 30, 48, and 72 hours. Table 5 and Figure 6 present the results. Figure 6 plots the MAPE (%) on the y-axis against input window size (in hours) on the x-axis for both electricity demand forecasting (left panel) and CO_2 emissions forecasting (right panel). CAL results are shown as blue circles and TEX results as orange squares, with connecting lines illustrating the performance trend across window sizes of 6, 12, 24, 30, 48, and 72 hours.

ARTICLE IN PRESS

Table 4 Ablation study comparing TAL-Net and LSTM performance under different feature configurations.

Region	Model	Features	MAPE Demand (%)	MAPE CO_2 (%)
CAL	TAL-Net	Full (27)	**1.29**	**3.41**
CAL	LSTM	Full (27)	1	4.48
CAL	TAL-Net	Basic (10)	5.25	17.23
CAL	LSTM	Basic (10)	5.71	18.44
CAL	TAL-Net	No Energy Mix (22)	4.5	16.98
TEX	TAL-Net	Full (27)	0.53	4.75
TEX	LSTM	Full (27)	0.44	4.73
TEX	TAL-Net	Basic (10)	9.54	15.42
TEX	LSTM	Basic (10)	8.83	15.83
TEX	TAL-Net	No Energy Mix (22)	4.97	10.28

Table 5 TAL-Net performance sensitivity to input window size. MAPE (%) for electricity demand and CO_2 emissions forecasting across different lookback windows. Bold values indicate best performance.

Window (hours)	CAL Demand (%)	CAL CO_2 (%)	TEX Demand (%)	TEX CO_2 (%)
6	0.61	4.17	**0.35**	5.34
12	0.55	3.19	0.38	**4.88**
24	0.78	**2.46**	0.51	5.27
30	**0.41**	2.80	0.53	5.83
48	0.42	3.99	0.78	5.45
72	0.46	4.30	0.73	5.10

IN PRESS

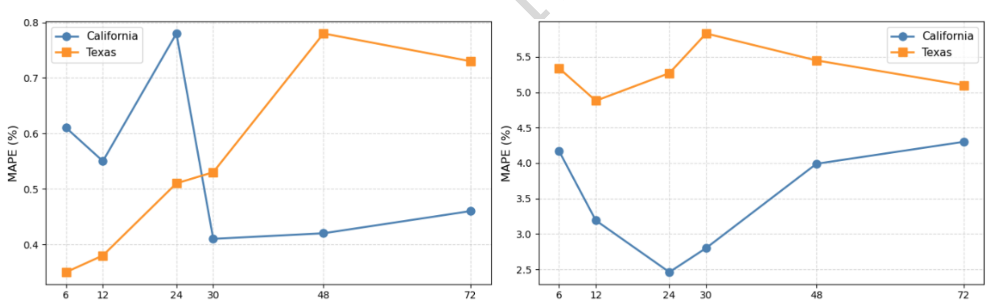


Fig. 6 Sensitivity analysis of TAL-Net forecasting performance to input window size for (a) electricity demand and (b) CO_2 emissions. CAL (blue circles) and TEX (orange squares) show different optimal windows, with the 30-hour selection providing balanced performance across both regions.

The results reveal region-specific optimal windows: CAL achieves best demand forecasting at 30 hours (MAPE 0.41%) and best CO_2 forecasting at 24 hours (MAPE 2.46%), while TEX performs optimally with shorter windows of 6–12 hours for demand (MAPE 0.35–0.38%). These differences likely reflect the distinct temporal dynamics of

ARTICLE IN PRESS

each region's grid, CAL's higher renewable penetration introduces greater variability that benefits from longer historical context, whereas TEX's fossil-fuel-dominated grid exhibits more consistent short-term patterns. The 30-hour window was selected as a balanced configuration that captures at least one full daily cycle, aligns with day-ahead market operations, and provides near-optimal performance across both regions and forecasting targets.

Despite these strengths, the study has some limitations. First, this study uses regional-level data aggregated at the balancing authority scale (CAL for CAL and TEX for TEX) rather than city-level data. However, the methodology is directly applicable to urban sustainability planning, as urban areas are the primary drivers of electricity demand within these regions. The framework could be adapted to metropolitan-level data when available. Second, the models were trained on historical hourly data without incorporating exogenous variables such as weather forecasts or economic indicators. Third, although validated on two distinct regions, broader testing across diverse geographic and operational contexts is needed to confirm generalizability.

Future work will aim to integrate real-time data streams, incorporate renewable generation forecasts and relevant economic factors, and apply interpretability techniques to better understand model decisions. Additionally, the current framework focuses on one-step-ahead (next-hour) forecasting; extending to longer horizons (e.g., day-ahead or week-ahead) would require architectural modifications such as sequence-to-sequence models or iterative multi-step prediction with uncertainty quantification. These advances will provide actionable insights for stakeholders and promote transparent, data-driven decision-making in energy management.

Methods

This section outlines the methodology adopted in this study, detailing the data sources, pre-processing steps, selection of regions, implementation of machine learning models, and performance evaluation metrics.

Data Description

The five-year hourly operational dataset from July 1, 2018, to June 30, 2023, published by the United States (US) Energy Information Administration (EIA) has been used in this study [33]. This dataset was also utilized by Biswas et al. [34] to quantify the effects of solar power adoption on CO_2 emissions reduction across U.S. regions; however, their study employed a distributed lag statistical model to estimate causal effects of solar generation on emissions, whereas our work focuses on developing a deep learning-based forecasting framework for jointly predicting future electricity demand and CO_2 emissions. This data provides regional information about electricity grid operations within the US and records power imports and exports between regions. It also includes hourly records of electricity demand, CO_2 emissions, and net electricity generation from different energy sources such as coal, natural gas, nuclear, hydroelectric, and solar. It should be noted that CAL and TEX refer to EIA-930 grid regions, not the U.S. states of California and Texas in their entirety. CAL encompasses a group of

ARTICLE IN PRESS

ARTICLE IN PRESS

balancing authorities within the broader California grid area, while TEX corresponds to a single balancing authority (ERCOT).

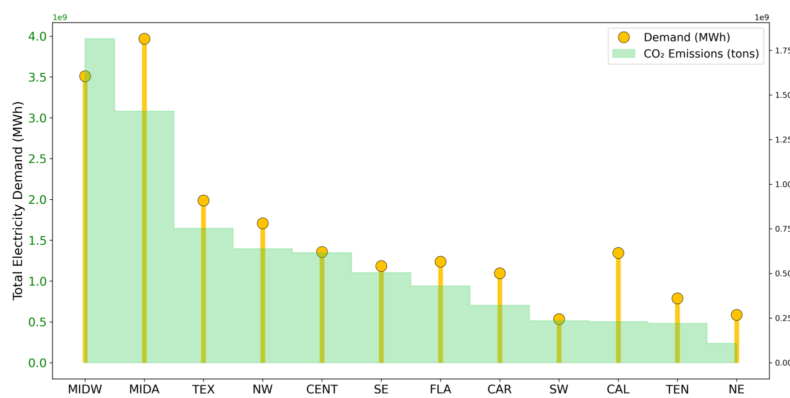


Fig. 7 Electricity demand (yellow lollipop markers, left axis) and total CO_2 emissions (green shaded area, right axis) by region, with red markers indicating the peak emission values for each region. California region (CAL) and Texas region (TEX) are highlighted for their contrasting profiles - California with high demand but relatively low emissions due to renewable penetration, and Texas with both high demand and high emissions due to fossil fuel reliance. All values represent aggregated totals for the full study period (1 July 2018–30 June 2023), with electricity demand in MWh and CO_2 emissions in metric tons.

For this research, the scope was narrowed to focus on two specific regions: the CAL region and the TEX region, which were selected due to their markedly different energy and emissions profiles. The CAL region is characterised by a higher share of renewable energy in its generation mix and a temperate climate, resulting in relatively lower emissions compared to its electricity demand. The TEX region (corresponding to the ERCOT balancing authority) relies heavily on fossil fuels for power generation and is subject to more extreme weather conditions, leading to both high electricity demand and significant CO_2 emissions. While the MIDW region exhibits the highest absolute CO_2 emissions across all U.S. regions as visible in Figure 7, TEX was selected over MIDW for two reasons. First, TEX corresponds to a single, well-defined balancing authority (ERCOT) that operates as a largely isolated grid with minimal interstate interconnections, making it a cleaner and more tractable case study for regional forecasting. In contrast, MIDW aggregates multiple balancing authorities with complex interchange patterns. Second, the scientific motivation is to evaluate model adaptability across regions with *contrasting* energy profiles: CAL (high renewables, low emissions intensity) and TEX (fossil-dominated, high emissions intensity) represent the two ends of this spectrum. Future work will extend the framework to MIDW and other regions to further validate generalisability. This contrast provides a meaningful and robust basis to test the ability of forecasting models to adapt to diverse regional characteristics, as illustrated in Figure 7, which compares electricity demand and CO_2 emissions across all U.S. regions and highlights the outlier positions of CAL and TEX for their demand and emissions patterns. In Figure 7, the x-axis lists U.S. regions sorted by total CO_2 emissions (descending), with yellow lollipop markers indicating

ARTICLE IN PRESS

total electricity demand (left y-axis, in MWh) and a green shaded area representing total CO_2 emissions (right y-axis, in tons). Red markers highlight the peak emission value for each region. California (CAL) is notable for its relatively high demand but low emissions owing to renewable energy penetration, while Texas (TEX) exhibits both high demand and high emissions due to fossil fuel reliance.

Data Pre-Processing

The original dataset contained approximately 525,060 rows and 45 columns, representing region-level hourly operational data for the entire United States power grid. It provided comprehensive records of electricity demand, CO_2 emissions, and detailed power generation information across all regions over five years.

For this analysis, the focus is on California and Texas due to their contrasting energy mixes and emissions profiles, as shown in 8. Figure 8 (a) presents a map of the United States highlighting the geographic locations of California (orange) and Texas (blue), the two study regions. Figures 8(b) and 8(c) display grouped bar charts of annual average electricity generation (in kWh) by fuel source - coal, natural gas, nuclear, hydro, and solar - for each year from 2018 to 2023, for California and Texas respectively. Figures 8(d) and 8(e) present doughnut charts showing the proportional contributions of different fuel sources to total CO_2 emissions in each region, with segments for coal (blue), natural gas (orange), and other sources (purple). After filtering for these two states, the working dataset includes 87,561 rows and retains all 45 columns covering the entire five-year period.

To better understand the regional differences in energy sources and CO_2 emissions, Figures 8(b) and 8(c) show the annual average electricity generation by source and the corresponding contributions to CO_2 emissions for California and Texas, respectively. In this figure, the x-axis represents the years from 2018 to 2023, and the y-axis shows the average annual electricity generation in kilowatt-hours, with different colours. California's bars show larger proportions of solar and hydro compared to Texas, which is dominated by natural gas and coal. Figures 8(d) and 8(e) present doughnut charts of CO_2 emissions where the segments correspond to coal (blue), natural gas (orange), and other sources (purple). In California, natural gas contributes the majority of emissions, while in Texas, emissions are almost equally split between coal and natural gas.

To prepare the data for modelling, the following pre-processing and feature engineering steps were performed. At the beginning, temporal features such as month, day of the week, weekend indicator, day of the year, and week of the year were extracted from the timestamp to capture periodic patterns in demand and emissions. Seasonal indicators for winter, spring, summer, and autumn were derived and one-hot encoded to represent the seasonal effects on energy consumption and emissions.

Missing hourly records within each region were identified and addressed using forward and backwards filling techniques, ensuring the time series remained continuous. Lagged features were then computed to account for temporal dependencies, including electricity demand and CO_2 emissions from the previous hour, the same hour on the previous day, and the same hour one week prior. Rolling mean features were computed over 3-hour and 24-hour windows to capture short- and medium-term trends in both demand and emissions.

ARTICLE IN PRESS

ARTICLE IN PRESS

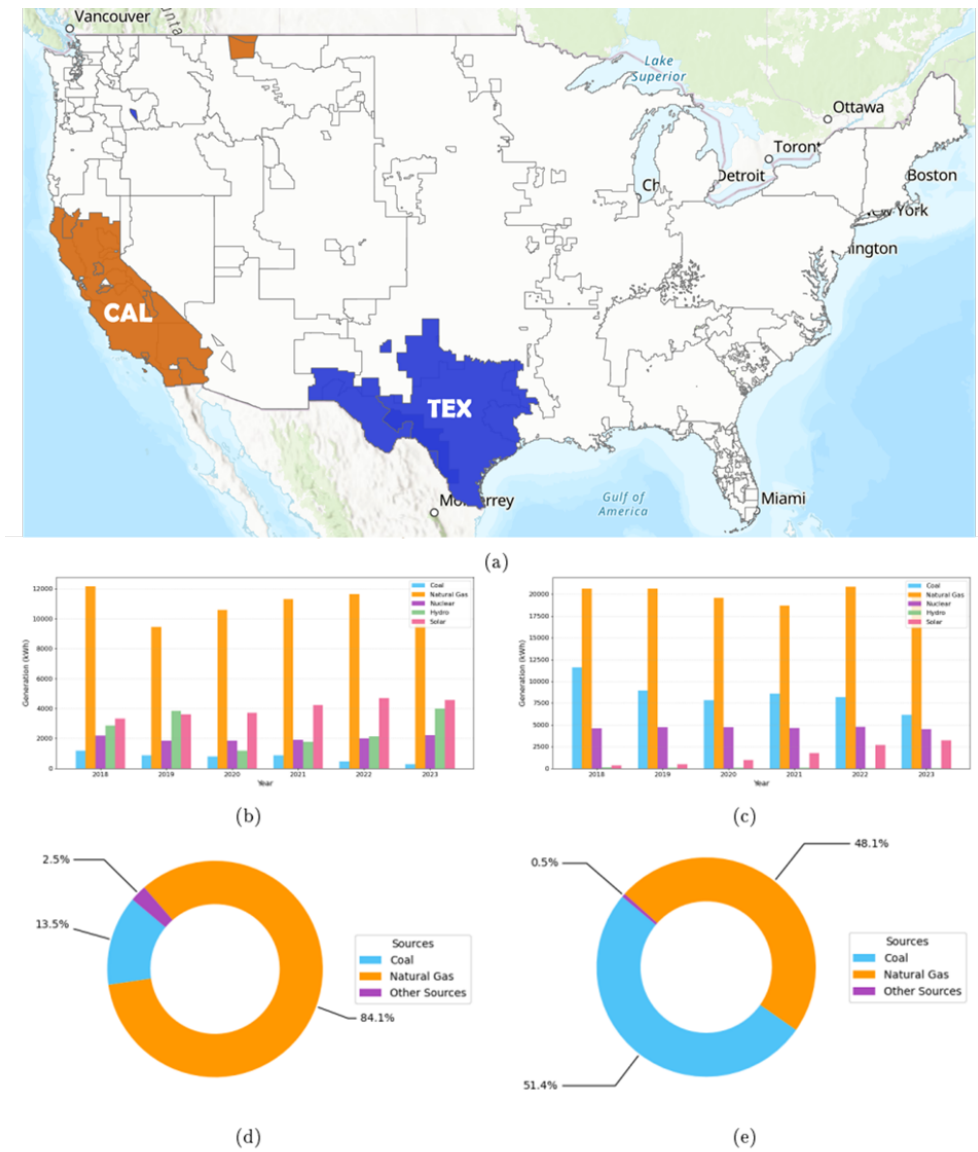


Fig. 8 (a)Approximate operational coverage of the EIA CAL and TEX balancing authority regions used in this study. The highlighted regions represent balancing authority operational areas rather than exact political state boundaries, (b) Comparison of annual average energy generation from various sources for California, (c) Comparison of annual average energy generation from various sources for Texas, (d) proportions of CO_2 emissions by different f uel s ource c ategories f or C alifornia, a nd (e) proportions of CO_2 emissions by different fuel source categories for Texas.

ARTICLE IN PRESS

A key modelling assumption underlying this approach is that historical fuel mix patterns provide a valid basis for forecasting CO_2 emissions. This is justified by the structure of the EIA-930 dataset, which records the observed hourly generation by fuel source for each balancing authority, rather than relying on static or assumed fuel compositions. Consequently, the engineered fuel mix features (Renewable_Pct, Fossil_Pct, Renewable_Demand, and Fossil_Demand) reflect the actual, hour-by-hour evolution of the generation mix within each region, including any gradual shifts in the supplier composition over time. While abrupt structural changes - such as the commissioning of a large new generation asset - may reduce short-term predictive accuracy, the five-year training period captures sufficient variation in the fuel mix to enable robust generalisation. The ablation study in Table 4 further confirms the validity of this approach: removing fuel mix features causes a disproportionately large degradation in CO_2 forecasting accuracy (MAPE increasing from 3.41% to 16.98% in the CAL region), demonstrating that these features carry genuine predictive signal rather than spurious correlation.

To capture the dynamic nature of the generation fuel mix, features representing the hourly proportions of renewable energy (Renewable_Pct, computed as hydro plus solar generation divided by total generation) and fossil fuel energy (Fossil_Pct, computed as coal plus natural gas divided by total generation) were calculated. Additionally, interaction terms Renewable_Demand and Fossil_Demand were derived by multiplying these percentages by the corresponding hourly demand, allowing the model to learn how the fuel mix composition at different demand levels influences CO_2 emissions. These features enable the TAL-Net to adapt to the hour-by-hour and seasonal variations in generation mix shown in Figures 8(b) and 8(c).

Further, features representing the energy generation mix were created, including the proportions of renewable and fossil fuel generation, as well as their respective contributions to total demand. Hour-to-hour and day-to-day differences in demand were also calculated to account for variability and abrupt changes in load. Boolean columns were converted to integers, and rows with missing values introduced during the creation of lagged and rolling features were removed. Finally, the dataset was sorted chronologically within each region and rounded to three decimal places to ensure consistency and accuracy.

All 27 features are fed into the TAL-Net's LSTM layer as the input vector (Table 6) at each timestep, allowing the attention mechanism to dynamically weight the importance of fuel mix variations alongside temporal patterns.

Deep Learning Models

This study investigates the use of advanced deep learning models to jointly forecast electricity demand and CO_2 emissions as a regression task, using their capability to model complex temporal dependencies and nonlinear patterns in time-series data [18]. We implemented and compared five deep learning models' architectures: Recurrent Neural Network (RNN)[35, 36], Long Short-Term Memory (LSTM), Restricted Boltzmann Machine combined with Multilayer Perceptron (RBM+MLP)[37, 38], Restricted Boltzmann Machine combined with LSTM (RBM+LSTM), and a Hybrid TAL-Net

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 6 Description of the 27 engineered input features grouped by category, including temporal indicators, lagged values, rolling statistics, energy generation mix proportions, and demand differencing terms.

Category	**Features**	**Count**
Temporal	Hour, Month, DayOfWeek, Is_Weekend, DayOfYear, WeekOfYear	6
Seasonal (one-hot)	Season_Autumn, Season_Spring, Season_Summer, Season_Winter	4
Demand Lags	Demand_Prev_Hour, Demand_Yesterday_Same_Hour, Demand_Last_Week_Same_Hour	3
Emission Lags	Emission_Prev_Hour, Emission_Yesterday_Same_Hour, Emission_Last_Week_Same_Hour	3
Rolling Statistics	Rolling_Mean_3H, Rolling_Mean_24H, Rolling_Mean_3H_Emission, Rolling_Mean_24H_Emission	4
Energy Mix	Total_Gen (MWh), Renewable_Pct (%), Fossil_Pct(%), Renewable_Demand, Fossil_Demand	5
Differencing	Demand_diff1, Demand_diff24	2
Total		27

model. Below briefly described the theoretical foundations and motivation behind each of the first four models before focusing in detail on the primary TAL-Net approach.

The RNN is a classical sequence learning architecture that processes input sequences step-by-step, maintaining a hidden state that captures information from previous time steps. Although RNNs are conceptually simple, but tend to struggle with longer sequences due to gradient instability, which often limits their accuracy on time-series with long-term dependencies.

The LSTM is a type of RNN specifically designed to capture long-range dependencies in sequential data by mitigating the vanishing gradient problem inherent in standard RNNs. Its gating mechanisms allow it to effectively learn patterns across varying time scales, making it a strong baseline for time-series forecasting tasks like electricity demand and emissions prediction.

The RBM+MLP model combines an unsupervised Restricted Boltzmann Machine for learning a compressed latent representation of the input features, followed by a feedforward multi-layer perceptron that performs the supervised prediction of electricity demand and CO_2 emissions. This pipeline improves initialization and helps capture complex feature interactions.

The RBM+LSTM extends this idea by feeding the latent representation learned by RBM into an LSTM network, combining the advantages of unsupervised feature learning and temporal sequence modeling. This hybridization aims to improve the model's ability to capture both feature-level and temporal dependencies in the data.

While these baseline models provide valuable insights into the capabilities of traditional and hybrid deep learning architectures, each has limitations in fully capturing intricate temporal dependencies and dynamically assigning importance to critical time

ARTICLE IN PRESS

ARTICLE IN PRESS

steps. To overcome these challenges, a TAL-Net model has been developed, augmenting the LSTM's sequential modeling capacity with an attention mechanism that selectively focuses on the most relevant time steps during prediction. The following subsection details the theoretical background, architectural design, and advantages of the proposed TAL-Net model.

Temporal Attention LSTM Network (TAL-Net)

The hybrid Temporal Attention LSTM Network (TAL-Net) proposed in this study extends the standard sequence-to-vector deep learning framework commonly applied to time-series forecasting. In its conventional form, such a framework employs a Long Short-Term Memory (LSTM) network to capture long-range temporal dependencies by processing a sequence of multivariate input vectors into a fixed-length hidden representation. An attention mechanism is subsequently applied to dynamically weigh the contributions of individual time steps, enabling the model to focus on the most informative portions of the sequence that may be underrepresented in purely LSTM-based architectures. Finally, fully connected dense layers map these learned representations to the target output predictions. This architecture is particularly well suited to forecasting tasks characterized by both temporal dependencies and complex interrelationships among multiple input features.

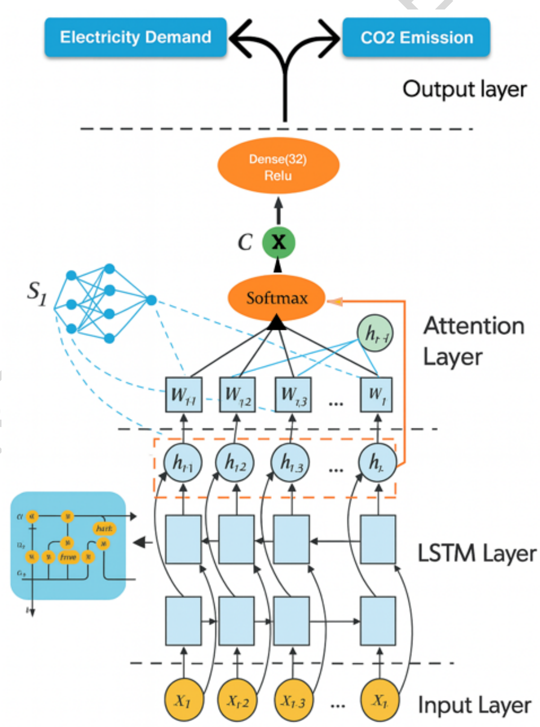


Fig. 9 Architecture of the proposed hybrid TAL-Net model for joint forecasting of electricity demand and CO_2 emissions.

ARTICLE IN PRESS

Figure 9 illustrates the architecture of the proposed hybrid TAL-Net model developed for joint forecasting of electricity demand and CO_2 emissions. The diagram in Figure 9 depicts the end-to-end pipeline of the TAL-Net model, with data flowing from bottom to top. At the bottom, the Input Layer receives multivariate feature vectors $(X_1, X_{t-2}, X_{t-3}, \ldots, X_T)$. These are processed by the LSTM Layer (shown in orange), which produces a sequence of hidden states $(h_{t-1}, h_{t-2}, h_{t-3}, \ldots, h_t)$. The Attention Layer (centre) computes learned weights $(W_{t1}, W_{t2}, W_{t3}, \ldots, W_t)$ via a softmax operation and aggregates the hidden states into a context vector C. This context vector passes through a Dense Layer with 32 neurons and ReLU activation before reaching the Output Layer, which produces two joint predictions: electricity demand and CO_2 emissions. To adapt this architecture, a set of 27 engineered features capturing temporal, seasonal, lagged, and rolling statistics was selected to form the feature vector, with electricity demand and CO_2 emissions as the two continuous targets. The features and targets were scaled independently to the range $[0, 1]$ to normalise the feature vector. The scaled data was then transformed into overlapping sequences of $T = 30$ timesteps to capture recent history, with each sequence paired to its corresponding next-step targets.

The model begins with the Input Layer, which accepts sequences of length $T = 30$, each timestep represented by 27 engineered features x_t. These multivariate time-series inputs are structured into a three-dimensional tensor of shape $(30, 27)$. The input sequence is then processed by the LSTM Layer, shown in the lower portion of the figure, which consists of 64 hidden nodes. The LSTM captures temporal dependencies by maintaining a memory of past timesteps through its recurrent structure, producing a sequence of hidden state vectors $H = [h_1, h_2, \ldots, h_T]$, as formalized in Equation (1):

$$h_t = \mathrm{LSTM}(x_t; \theta_{\mathrm{LSTM}}) \tag{1}$$

These hidden states h_t encode the temporal patterns and are passed to the Attention Layer, highlighted in the middle of Figure 9. The attention mechanism computes an alignment score e_t for each hidden state by applying a trainable linear projection followed by a tanh activation:

$$e_t = \tanh(W h_t + b) \tag{2}$$

where W and b are learned parameters. The scores e_t are normalized using a softmax function to produce attention weights α_t, which indicate the relative importance of each timestep:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \tag{3}$$

The attention weights are visualized in the figure as the arrows connecting the hidden states h_t to the context vector c through a softmax operation. The context vector aggregates the hidden states into a single summary representation of the sequence by

ARTICLE IN PRESS

ARTICLE IN PRESS

computing a weighted sum of the hidden states:

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{4}$$

The context vector c, as depicted in the center of Figure 9, captures the most relevant temporal information, emphasizing important time steps dynamically. This vector is then passed to a fully connected Dense Layer with 32 neurons and ReLU activation, which projects c into a higher-level feature space, enabling the model to learn nonlinear interactions:

$$z = \text{ReLU}(W_c c + b_c) \tag{5}$$

Finally, the output z is passed to the Output Layer, as shown at the top of the figure, where a second dense layer produces the two continuous target variables, forecasted electricity demand and CO_2 emissions, as a two-dimensional vector:

$$\hat{y} = W_o z + b_o \tag{6}$$

The TAL-Net architecture jointly predicts both electricity demand and CO_2 emissions through a shared representation learning approach. Rather than treating emissions as a simple function of demand (which would merely rescale one output from the other), the model learns from 27 input features that include distinct lag and rolling statistics for both demand and emissions, as well as fuel mix proportions. This means the attention mechanism can assign different weights to timesteps that are informative for demand versus emissions - for instance, weighting recent renewable generation percentages more heavily when predicting emissions while focusing on temperature-related temporal patterns for demand. The shared LSTM hidden states capture the correlation structure between the two targets, while the separate pathways in the final dense layer allow for target-specific adjustments.

The architecture in Figure 9 provides a clear visual representation of the pipeline described by Equations (1)–(6). It shows how the LSTM layer models sequence dynamics, the attention layer focuses on salient timesteps to compute the context vector, and the dense layers transform this information to produce accurate joint predictions for electricity demand and emissions. This combination of long-term memory, dynamic weighting, and nonlinear projection makes the model well-suited for the complex forecasting task at hand.

To prevent overfitting and ensure generalization, several strategies were employed. First, the models were trained for only 10 epochs, which was determined through early experimentation to provide stable convergence without overfitting (training loss continued to decrease while validation loss stabilized). Second, MinMax scaling to [0, 1] was applied independently to features and targets, preventing scale-related training instabilities. Third, the chronological train-test split preserves the temporal ordering of the data, ensuring that the model is evaluated on genuinely future observations rather than interpolating between randomly sampled points. Finally, the consistent

ARTICLE IN PRESS

ARTICLE IN PRESS

performance across five different train-test splits (80-20 through 60-40) in Tables 2-3 demonstrates that the low errors are not artifacts of a particular data partition but reflect genuine predictive capability.

Performance Evaluation

The performance of all five forecasting models developed in this study, LSTM, RNN, RBM+MLP, RBM+LSTM, and the hybrid TAL-Net, was evaluated using three widely adopted regression metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), as these are commonly used in the energy forecasting literature.

RMSE (Equation 7) measures the square root of the mean squared difference between predicted and actual values, penalising larger errors more strongly and highlighting cases where extreme deviations occur:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

MAE (Equation 8) represents the average of the absolute differences between predicted and actual values, providing a straightforward interpretation of the typical error magnitude, expressed in the same units as the target variable:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{8}$$

MAPE (Equation 9) expresses the average absolute error as a percentage of the actual values, making it a scale-independent measure that facilitates comparison of forecast accuracy between different models and regions:

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i + \epsilon}\right| \tag{9}$$

where ϵ is a small constant added to prevent division by zero.

These metrics were applied consistently to evaluate all five models on both electricity demand and CO_2 emissions forecasts for California and Texas. Together, RMSE (Equation 7), MAE (Equation 8), and MAPE (Equation 9) offer a comprehensive and widely accepted framework for assessing the predictive performance of energy forecasting models, enabling objective comparison of results.

MAPE was chosen as the primary metric because it provides a scale-independent measure that allows direct comparison between regions with different d emand and emission magnitudes, as well as between the two target variables. While MAPE can be sensitive to near-zero values, this is mitigated in our dataset because California's minimum hourly CO_2 emissions (∼1,500 tons) and Texas's (∼6,000 tons) are sufficiently above zero to avoid numerical instability. Nevertheless, we also report MAE

ARTICLE IN PRESS

and RMSE in Tables 2-3, which provide complementary perspectives: MAE indicates the average absolute error in the original units, while RMSE penalizes larger deviations more heavily. The consistency of model rankings across all three metrics confirms that our conclusions are not artifacts of the chosen evaluation measure.

Declarations

Data availability. The dataset is a publicly available dataset, published by the U.S. Energy Information Administration (EIA) between 1st July 2018, and 30th June 2023 [33]. Available doi: *https://doi.org/10.7910/DVN/OKEATQ*

Code availability. The code developed for this study, is publicly available on GitHub at: *https://github.com/Ann-Mary-Thomas/talnet-usa-energy-co2.git*. The implementation was developed in Python 3.9.13.

Acknowledgements. No external funding was received for this work.

Author Contributions. AMT: Data curation; software; investigation; methodology; formal analysis; visualisation; writing original draft. MD: Conceptualisation, Resources, Investigation, Software, Methodology, Validation, Visualisation, Writing and reviewing. SPR: Visualisation, Software, Investigation, Validation, Formal analysis, Reviewing. All authors have read and approved the manuscript.

Competing interests. The authors declare no competing financial or non-financial interests.

References

[1] Hu, Y. & Man, Y. Energy consumption and carbon emissions forecasting for industrial processes: Status, challenges and perspectives. *Renewable and Sustainable Energy Reviews* **182**, 113405 (2023).

[2] Akhmat, G. *et al.* The challenges of reducing greenhouse gas emissions and air pollution through energy sources: evidence from a panel of developed countries. *Environmental Science and Pollution Research* **21**, 7425–7435 (2014).

[3] Shurpali, N., Agarwal, A. K. & Srivastava, V. Greenhouse gas emissions. *Energy, environment, and sustainability. Singapore: Springer* (2019).

[4] Aderibigbe, A. O., Ani, E. C., Ohenhen, P. E., Ohalete, N. C. & Daraojimba, D. O. Enhancing energy efficiency with ai: a review of machine learning models in electricity demand forecasting. *Engineering Science & Technology Journal* **4**, 341–356 (2023).

[5] Pimenow, S., Pimenowa, O. & Prus, P. Challenges of artificial intelligence development in the context of energy consumption and impact on climate change. *Energies* **17**, 5965 (2024).

ARTICLE IN PRESS

ARTICLE IN PRESS

[6] Ukoba, K., Olatunji, K. O., Adeoye, E., Jen, T.-C. & Madyira, D. M. Optimizing renewable energy systems through artificial intelligence: Review and future prospects. *Energy & Environment* **35**, 3833–3879 (2024).

[7] Balat, M., Balat, H. & Acici, N. Environmental issues relating to greenhouse carbon dioxide emissions in the world. *Energy exploration & exploitation* **21**, 457–473 (2003).

[8] Chukwunonso, B. P. *et al.* Predicting carbon dioxide emissions in the united states of america using machine learning algorithms. *Environmental Science and Pollution Research* **31**, 33685–33707 (2024).

[9] Kang, K., Zhu, M., Sun, G. & Guo, X. Fossil fuels versus biofuels: Perspectives on greenhouse gas emissions, energy consumptions, and projections (2019). In: Fuel Processing and Energy Utilization, pp. 1–14. Chapman and Hall/CRC.

[10] Kumari, S. & Singh, S. K. Machine learning-based time series models for effective co_2 emission prediction in india. *Environmental Science and Pollution Research* **30**, 116601–116616 (2023).

[11] Namboori, S. *Forecasting carbon dioxide emissions in the United States using machine learning.* Ph.D. thesis, Dublin, National College of Ireland (2020).

[12] Quadrelli, R. & Peterson, S. The energy–climate challenge: Recent trends in co2 emissions from fuel combustion. *Energy policy* **35**, 5938–5952 (2007).

[13] Rivera-González, L., Bolonio, D., Mazadiego, L. F. & Valencia-Chapi, R. Long-term electricity supply and demand forecast (2018–2040): A leap model application towards a sustainable power generation system in ecuador. *Sustainability* **11**, 5316 (2019).

[14] Liu, Y., Yu, S., Zhu, Y., Wang, D. & Liu, J. Modeling, planning, application and management of energy systems for isolated areas: A review. *Renewable and Sustainable Energy Reviews* **82**, 460–470 (2018).

[15] Ma, X., Wang, J., Huang, J. & Ke, Y. Energy consumption and carbon emission modeling and forecasting study with novel deep learning methods. *Expert Systems with Applications* **290**, 128314 (2025).

[16] Nejat, P., Jomehzadeh, F., Taheri, M. M., Gohari, M. & Majid, M. Z. A. A global review of energy consumption, co2 emissions and policy in the residential sector (with an overview of the top ten co2 emitting countries). *Renewable and sustainable energy reviews* **43**, 843–862 (2015).

[17] Daioglou, V. *et al.* Energy demand and emissions of the non-energy sector. *Energy & Environmental Science* **7**, 482–498 (2014).

ARTICLE IN PRESS

ARTICLE IN PRESS

[18] Qureshi, M., Arbab, M. A. & Rehman, S. u. Deep learning-based forecasting of electricity consumption. *Scientific Reports* **14**, 6489 (2024).

[19] Abbasimehr, H., Shabani, M. & Yousefi, M. An optimized model using lstm network for demand forecasting. *Computers & industrial engineering* **143**, 106435 (2020).

[20] Aryai, V. & Goldsworthy, M. Day ahead carbon emission forecasting of the regional national electricity market using machine learning methods. *Engineering Applications of Artificial Intelligence* **123**, 106314 (2023).

[21] Chang, Y.-S. *et al.* An lstm-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research* **11**, 1451–1463 (2020).

[22] Hamdan, A., Al-Salaymeh, A., AlHamad, I. M., Ikemba, S. & Ewim, D. R. E. Predicting future global temperature and greenhouse gas emissions via lstm model. *Sustainable Energy Research* **10**, 21 (2023).

[23] Choi, E., Cho, S. & Kim, D. K. Power demand forecasting using long short-term memory (lstm) deep-learning model for monitoring energy sustainability. *Sustainability* **12**, 1109 (2020).

[24] Singhal, D., Ahuja, L. & Seth, A. Forecasting smart grid stability using a hybrid cnn bi-lstm approach. *SN Computer Science* **5**, 539 (2024).

[25] Alhussein, M., Aurangzeb, K. & Haider, S. I. Hybrid cnn-lstm model for short-term individual household load forecasting. *Ieee Access* **8**, 180544–180557 (2020).

[26] Cavus, M. & Allahham, A. Spatio-temporal attention-based deep learning for smart grid demand prediction. *Electronics* **14**, 2514 (2025).

[27] Ciechulski, T., Paś, J., Stawowy, M. & Duer, S. Using artificial intelligence to predict power demand in small power grids—problem analysis as a method to limit carbon dioxide emissions. *Sustainability* **17**, 3694 (2025).

[28] Chen, C.-Y., Chai, K. K. & Lau, E. Ai-assisted approach for building energy and carbon footprint modeling. *Energy and AI* **5**, 100091 (2021).

[29] Dey, M., Wickramarachchi, D., Rana, S. P., Simmons, C. V. & Dudley, S. Power grid frequency forecasting from micro-PMU data using hybrid vector-output LSTM network (2023). In: 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), pp. 1–5. IEEE.

[30] Dey, M., Rana, S. P., Wylie, J., Simmons, C. V. & Dudley, S. Detecting power grid frequency events from µpmu voltage phasor data using machine learning (2022). In: IET Conference Proceedings CP811, pp. 125–129. IET.

ARTICLE IN PRESS

ARTICLE IN PRESS

[31] Parameswaran, S., Dey, M. & Patel, P. Micro-pmu data-driven anomalous voltage event detection for the power distribution system (2024). In: International Conference on Frontiers of Intelligent Computing: Theory and Applications, pp. 145–157. Springer.

[32] Dey, M., Rana, S. P. & Dudley, S. Automated terminal unit performance analysis employing x-RBf neural network and associated energy optimisation–a case study based approach. *Applied Energy* **298**, 117103 (2021).

[33] Biswas, A., Mork, D., Qiu, M., Braun, D. & Dominici, F. Five-year dataset depicting electric power generation and co_2 emissions within the u.s. electricity sector, spanning from july 2018 to june 2023 (2023).

[34] Biswas, A., Qiu, M., Braun, D., Dominici, F. & Mork, D. Quantifying effects of solar power adoption on co2 emissions reduction. *Science Advances* **11**, eadq5660 (2025).

[35] Amalou, I., Mouhni, N. & Abdali, A. Multivariate time series prediction by rnn architectures for energy consumption forecasting. *Energy Reports* **8**, 1084–1091 (2022).

[36] Aseeri, A. O. Effective rnn-based forecasting methodology design for improving short-term power load forecasts: Application to large-scale power-grid time series. *Journal of Computational Science* **68**, 101984 (2023).

[37] Moayedi, H. *et al.* Prediction of co2 emission for the central european countries through five metaheuristic optimization techniques helping multilayer perceptron. *Engineering Applications of Computational Fluid Mechanics* **18**, 2327437 (2024).

[38] Babu, V. *et al.* An innovative method for electricity load forecasting in the agriculture sector using a restricted boltzmann machine (rbm) model (2024). In: 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), pp. 778–783. IEEE.

ARTICLE IN PRESS