ORIGINAL RESEARCH **Open access**

Multi-Energy Load Forecasting with Mixed Time Scales Using Multi-Task Learning and Stacked Cross-Attention

Hassan Rahman[1*], Tariq Mahmood[1]

Abstract

Reliable prediction of multiple energy loads stands as an essential condition for matching supply to demand within integrated energy systems. Yet the contrasting reaction dynamics and usage profiles separating electrical, thermal, and cooling loads give rise to two core difficulties: inconsistent temporal resolutions and unequal predictive reliability from one load category to another. To surmount these hurdles, the present work advances a multi-task learning structure underpinned by stacked cross-attention. Within this structure, a dedicated alignment unit reconciles the disparate time granularities of individual loads, while Informer encoder stacks are assigned as domain experts to isolate characteristics unique to each load. Stacked cross-attention, operating as the soft sharing conduit, fluidly merges expert-derived representations at the sequential level, promoting tighter inter-task cooperation and responsiveness. The overall design boosts aggregate predictive fidelity under mixed temporal resolutions and simultaneously narrows the accuracy gap across load types. Side-by-side experimental evidence reveals that the variant employing stacked cross-attention registers the strongest predictive outcomes and shrinks the imbalance metric by 79.17 %. Beyond this, the Informer-based expert encoder blocks secure a greater-than-30.09 % reduction in MAPE when set against substitute expert realizations. In comparison with multi-gate mixture-of-experts frameworks, long-established forecasting benchmarks, and recently published advanced solutions, the introduced framework consistently reaches higher forecasting reliability, corroborating both its practical value and forward-looking character.

Keywords Load forecasting, Integrated energy systems, Multi-task learning, Mixed time scales, Feature fusion

*Correspondence:
Hassan Rahman
hrahman@uok.edu.pk
[1] Department of Artificial Intelligence and Power Engineering, Faculty of Engineering, University of Karachi, Karachi, Pakistan

Introduction

Steady gains in societal living standards have been accompanied by mounting fossil fuel depletion and carbon output. In response, the Paris Agreement codifies the ambition of holding the worldwide temperature increase below 2 °C relative to pre-industrial levels by the close of the century [1]. Integrated energy systems (IES) represent a cornerstone mechanism for delivering energy thrift and emission cuts [2, 3]. By knitting together disparate energy vectors and capitalizing on cross-vector synergies, an IES raises overall efficiency and accelerates the integration of renewables [4-6]. Dependable multi-energy load forecasting constitutes a foundational prerequisite for demand-responsive energy dispatch in IES, serving an indispensable function in orchestrating multi-energy coordination and informed system governance.

Related works

Contemporary approaches to load forecasting broadly fall into two groupings: physics-driven white-box paradigms reliant on simulation environments, and data-driven black-box paradigms built atop machine learning [7]. Prominent simulation

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

environments encompass EnergyPlus [8] and DeST [9]. Such physics-based routes offer considerable transparency, yet they require granular engineering inputs and laborious model construction, and they falter when faced with rapid operational fluctuations [10]. In parallel, machine-learning-centric forecasting pipelines mine consumption regularities from rich historical archives, constructing input–output mappings between load profiles and their drivers [11]. Classical machine learning regressors, including Support Vector Regression (SVR) [12], Decision Tree (DT) [13], and Artificial Neural Network (ANN) [14], have been widely used. For example, Neto and Fiorelli [15] used ANN to estimate facility energy consumption and compared its accuracy against detailed white-box baselines, observing that ANN yielded superior short-horizon predictions. The escalating sophistication of data streams and usage regimes has imposed stricter expectations on predictive architectures. Deep learning (DL) paradigms, spanning Long Short-Term Memory (LSTM) [16], Gated Recurrent Unit (GRU) [17], and Convolutional Neural Network (CNN) [18], emerged as a response to these heightened demands. Models formulated around such paradigms customarily adopt deeper hierarchical designs and exhibit stronger representational power. Ciechulski and Osowski [19] fielded LSTM for short-range load prediction in the Polish Power System, reporting elevated accuracy across both 1-hour and 24-hour forecast lead times. Chalapathy *et al*. [20] constructed and contrasted DL-based against non-DL-based architectures under varying forecast horizons, and the collected results indicated that DL-driven models attained the best performance in terms of both absolute deviation and percentage error.

Together with the maturation of IES, multi-energy load forecasting has attracted intensified scholarly focus. Multi-task learning (MTL), distinguished by its capacity to manage a suite of forecasting objectives concurrently, has risen to prominence when coupled with DL algorithms [21]. Recent contributions have enlisted diverse DL algorithms as domain-specific specialists inside MTL configurations to tease out per-task load signatures. Wang *et al*. [22] employed an LSTM in the specialist role and constructed a high-accuracy multi-energy prediction engine. In the investigation by Li *et al*. [23], the candidate model was pitted against counterparts drawing on CNN-LSTM and LVSVM specialists, revealing marginal accuracy gains. Critically, the evidence also exposed sizeable swings in predictive quality linked directly to specialist algorithm selection. While the studies cited above affirm the viability of learned feature extraction, their effectiveness plateaus once they confront intricate time-series data streams. Informer [24], a state-of-the-art temporal modeling tool, is expressly built to master high-dimensional, extended-duration sequences. Gong *et al*. [25] fielded Informer to anticipate heating demand inside district-scale heating networks and established its superiority over LSTM and RNN formulations. It follows that to push the envelope in feature extraction proficiency, MTL frameworks must absorb ever more powerful temporal modeling backbones.

The inter-task information-sharing blueprint constitutes another pivotal lever governing MTL efficacy, as it dictates both the pathway and the intensity with which tasks exchange representations. Sharing blueprints are conventionally split into rigid hard sharing and flexible soft sharing. Under a hard sharing regime, all tasks share segments of a common computational graph, while task-distinguishing components are routed through parallel specialist arms. Under a soft sharing regime, each task instead preserves its own secluded network—often labeled an expert module—and relies on dedicated mechanisms such as learned convex combinations or gating circuits to perform representation merging and information dissemination. This latter paradigm is more congenial to heterogeneous task collections and enlarges the model's expressive range. Although multi-energy streams within an IES are interconnected to some extent, the constituent loads—electrical, heating, and cooling—exhibit divergent dynamic responses and physical substrate properties. Imposing hard sharing upon such heterogeneous tasks risks triggering harmful interference, manifested as unbalanced predictive fidelity across task channels. Soft sharing thus emerges as the more fitting design choice for IES multi-energy forecasting. By way of illustration, Tan *et al*. [26] introduced a composite correlation strength index to bestow static weight coefficients upon the various experts, thereby materializing a soft sharing scheme. Zhao and Wu [27] quantified coupling similarity scores to drive weighted aggregation of load feature streams. Extending this line, Guo *et al*. [28] fashioned a communal representation tier tasked with collecting task-level feature signals originating from every task branch. A kindred approach was pursued by Li *et al*. [29], who implemented fully connected layers as communal representation tiers. Hu *et al*. [30] operationalized attention operations to amalgamate salient variability signatures sourced from diverse load classes. Even so, most reported methodologies remain tethered to global weighting or coarse fusion rules, neither of which affords the plasticity to calibrate locally to instance- and task-specific idiosyncrasies. This

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

rigidity can lead to the bypassing of crucial cross-task interactions. Hence, there is an urgent need for a soft sharing construct that both formally captures cross-task relational structure and permits fine-grained, context-responsive feature melding. Such a construct would unlock richer inter-task collaboration and drive further accuracy uplifts.

Beyond the above, the matter of incongruent forecast time scales across electrical, heating, and cooling loads within IES has drawn relatively little scrutiny. The mainstream of the current literature imposes a uniform temporal granularity upon multi-energy prediction exercises. For example, Guo *et al*. [28], Liu *et al*. [31], and Niu *et al.* [32] consider the entire set of energy loads—electrical, heating, and cooling alike—at the hourly mark. However, because the electrical and thermal branches of an IES operate under fundamentally different dynamic regimes, their respective forecast horizon requirements part ways significantly. The electrical branch, characterized by rapid control loops, necessitates more granular predictive time steps, commonly situated in the minutes-to-sub-hour band [33, 34]. The thermal branch, by contrast, governed by storage effects such as thermal inertia and fluid transport lag, evolves on a slower control timescale and therefore calls for sparser predictive time steps, usually at or above the hourly level [35, 36]. Relying on a homogeneous time grid cannot satisfy the simultaneous predictive needs of differing load classes and may precipitate either supply–demand discord or gratuitous computational expense. Moreover, forecasting across heterogeneous time grids begets unequal numbers of target steps for each load, layering additional complexity onto model formulation. For these reasons, confronting multi-energy load forecasting under mixed timescale settings is of first-order importance for the competent scheduling and supervisory control of IES.

In sum, despite considerable headway in multi-energy load forecasting, further concerted work is still needed to tackle mixed-time-scale prediction to meet practical deployment demands. In parallel, more formidable expert backbones and more nuanced soft-sharing constructs remain essential for delivering high-fidelity and equitably balanced multi-energy load predictions.

Main contributions

To close the investigatory voids identified above, the current study puts forward an original MTL framework that weaves together a temporal resolution reconciliation unit, specialist blocks built upon Informer encoder stacks, and stacked cross-attention for multi-energy load prediction under non-uniform time granularities, with the capacity to secure both elevated precision and a more uniform distribution of predictive performance. The chief contributions delivered by this work are as follows:

(1) The introduced framework tackles the heterogeneous temporal resolution requirements inherent to multi-energy load prediction within IES. By reconciling and consolidating disparate time granularities, it empowers 15-minute forecasting for electrical loads alongside hourly forecasting for heating and cooling loads across the ensuing hour.

(2) An original soft sharing construct, designated the SCA module, is presented to formally characterize cross-task relational dynamics while accommodating context-sensitive and high-resolution feature blending. It promotes collaborative exchange across tasks and retention of task-unique signatures, which together yield greater accuracy and balance in multi-energy load forecasts.

(3) The introduced framework embeds Informer encoder stacks as domain specialists within an MTL paradigm. Harnessing their recognized proficiency in processing elongated sequences and isolating salient representations, this integration produces substantial accuracy gains in multi-energy load prediction tasks.

(4) An exhaustive set of comparative assessments and performance audits of the introduced SCA-MTL framework is undertaken. When set alongside leading-edge competitor architectures, the framework consistently recorded top-tier predictive accuracy, bearing out both its practical soundness and its forward-looking design.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Integrated Energy System and Multi-Energy Load Characteristics

An IES ordinarily couples several energy subsystems through conduit networks and energy conversion plants [37], affording versatile mobilization of varied energy media to meet end consumers' multi-energy requirements, as illustrated in **Figure 1**. Given the dense integration and mutually reinforcing character of IES, its operational planning and fine-tuning rest upon trustworthy forecasts of multi-energy loads [38], which constitutes the cornerstone for assuring resource-efficient and disturbance-resilient system functioning.

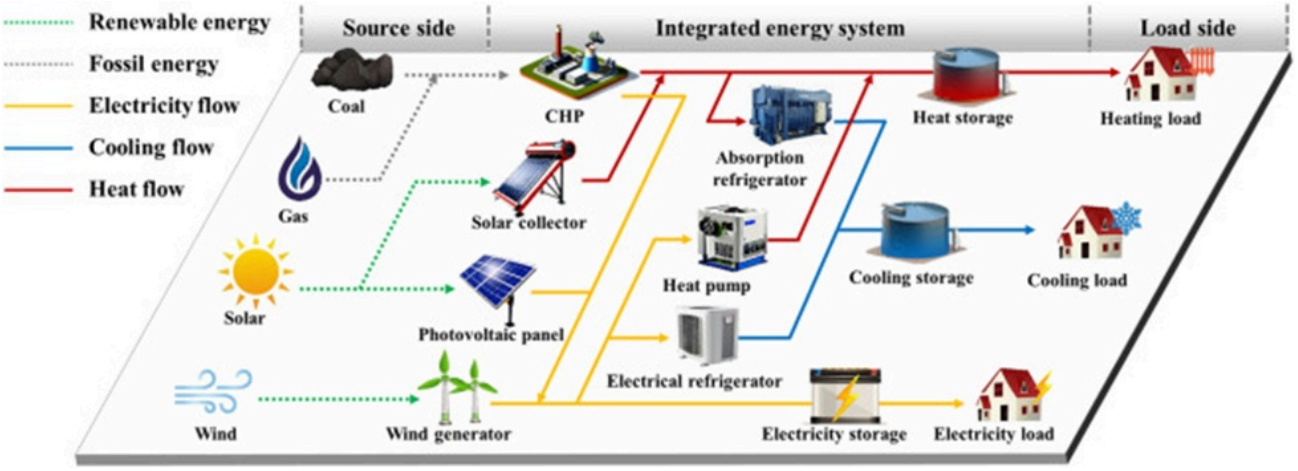


Figure 1. Schematic of IES.

Inside an IES, energy flows are deeply intertwined, fostering strong reciprocal dependencies among load categories and yielding intricate load behaviors. An alteration in a single load class can reverberate throughout the entire arrangement. Furthermore, the dynamical properties and control cadences of interconnected subsystems are not uniform, resulting in non-identical forecasting temporal resolutions. These attributes conspire to make multi-energy load prediction an especially formidable endeavor. Realizing both accurate and equitably balanced multi-energy load forecasts under mixed temporal resolutions holds decisive significance for the IES discipline.

Materials and Methods

The full-cycle workflow for multi-energy load forecasting is depicted in **Figure 2**. First, data scrutiny and preparatory steps were executed, comprising dataset merging, interdependency evaluation, measurement unit conversion, rescaling, and partition into training and evaluation subsets. Afterward, model optimization and multi-energy load forecasting were conducted. This phase entailed hyperparameter selection and the institution of an early termination criterion. The multi-energy load records were initially harmonized to a common temporal grid and subsequently passed into the specialist blocks for load representation extraction. The harvested representations were then refined and combined via the SCA module to suit the distinct needs of each prediction target. These combined representations were then relayed to the output module to generate the conclusive load forecasts. Ultimately, leveraging the optimized model, a suite of side-by-side comparisons and appraisals was carried out on the withheld test partition. Ablation trials, benchmarking against substitute specialist configurations, and parallel evaluations against state-of-the-art (SOTA) models were orchestrated using the defined performance indicators to substantiate the strengths of the proposed framework.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

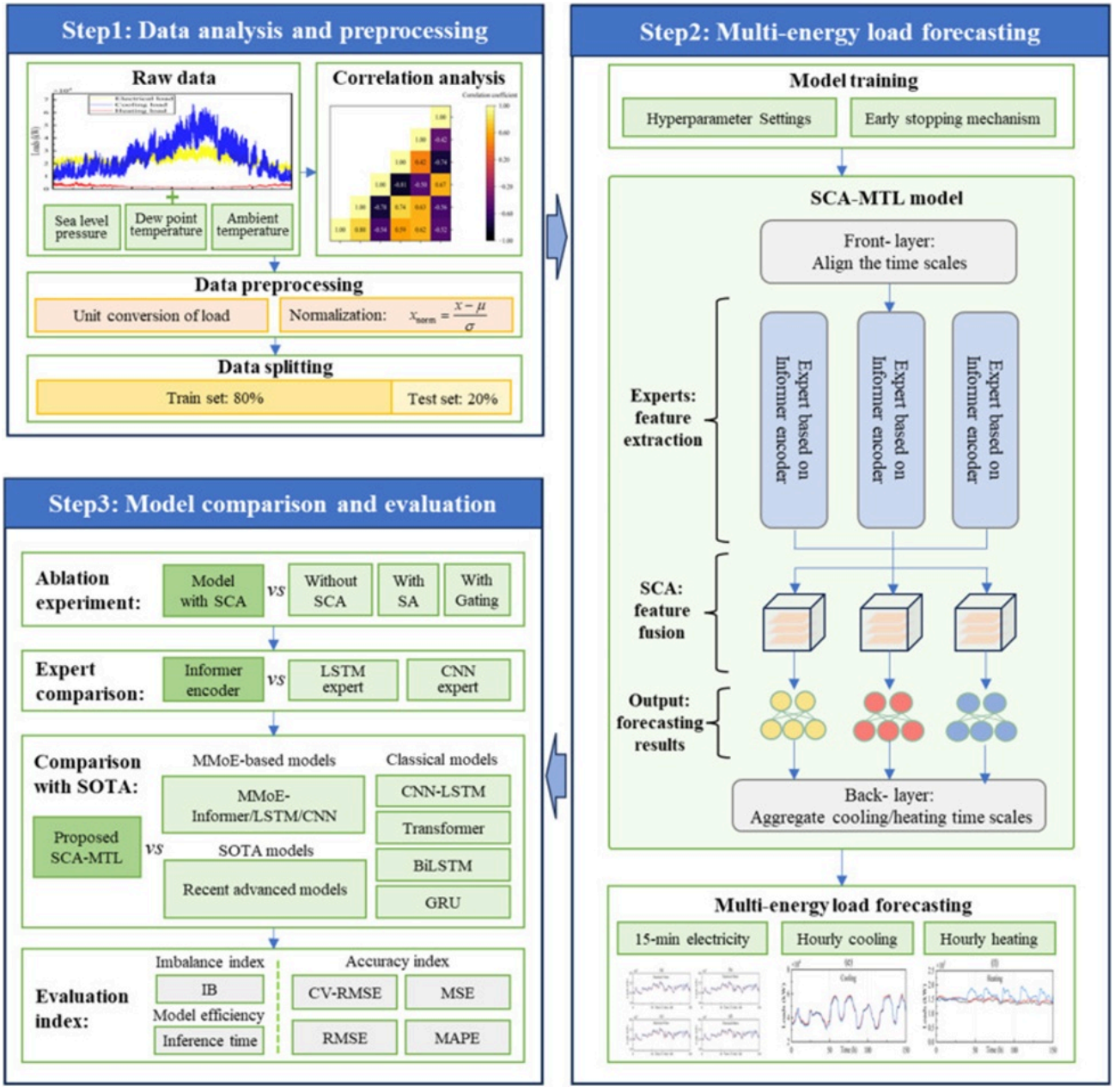


Figure 2. Flow chart of multi-energy load forecasting.

Time scale alignment module

Different load classes exhibit distinct reaction behaviors, introducing mismatches in required forecast granularities and native data sampling intervals. These mismatches pose obstacles to modeling under mixed time resolutions, including hindrances to learning load interdependencies and possible erosion in forecast fidelity. In response, the temporal resolution reconciliation unit is devised to resolve the difficulty of non-uniform time granularities across multi-energy loads. This unit comprises a front-end phase and a back-end phase. The front-end phase brings loads onto a shared time grid and enforces uniform sequence extent. The back-end phase reverts the temporal resolution to quantify forecast error against observed truth. As shown in **Figure 3**, the hourly-registered heating and cooling loads, together with the 15-minute-registered electrical load sourced from Arizona State University [39], are used to illustrate the operation of the temporal resolution reconciliation unit. Within the front-end phase, successive hourly heating and cooling load records are

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

enriched with finer-grained estimates through linear interpolation, thereby equalizing temporal resolutions while maintaining the veracity of the initial measurements, as given by **Eq. 1.**

$$L_{i,t} = L_i + \left(\frac{t}{60}\right)\cdot(L_{i+1} - L_i), t = 15, 30, 45 \quad (1)$$

Where i denotes the hourly index, t denotes the interpolated minute position, and L denotes the heating or cooling load. The multi-energy load sequences, now possessing matching length and temporal granularity, are directed into the forecasting model. Within the back-end phase, the electrical load predictions are left unaltered. In contrast, the heating and cooling load estimates are combined by averaging from the 15-minute granularity up to the hourly granularity to produce the final forecasted values.

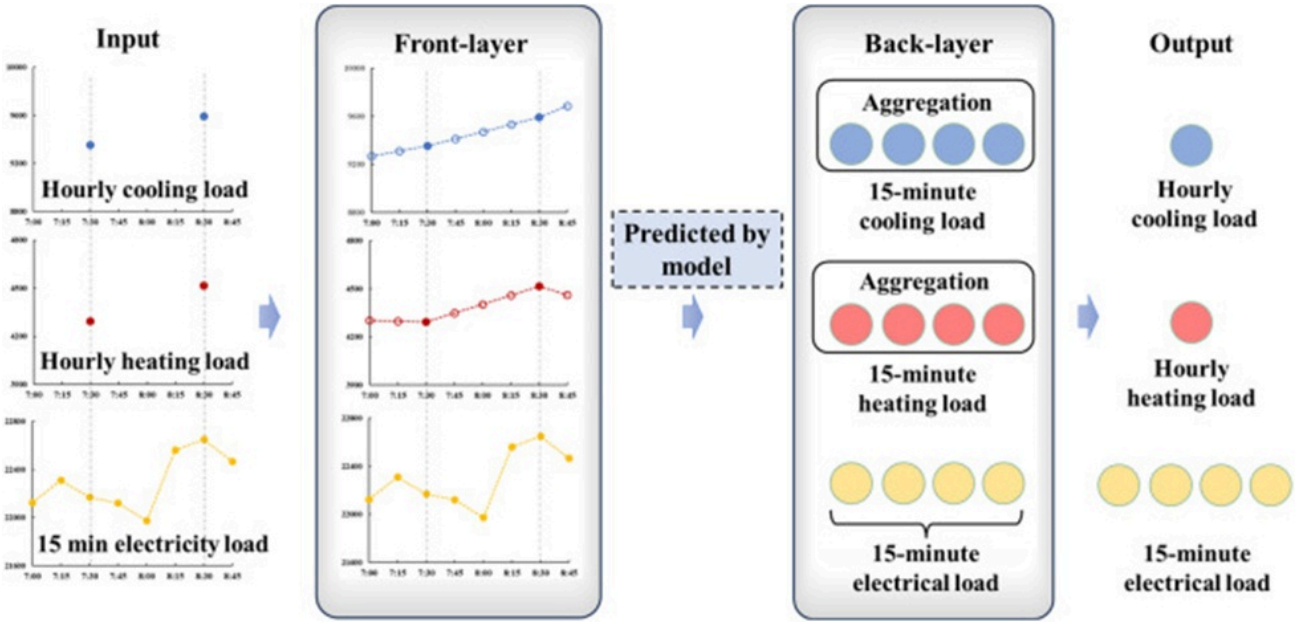


Figure 3. Procedure of time scale alignment module.

Through the temporal resolution reconciliation unit, the forecasting model can manage load sequences with disparate temporal granularities and extents, facilitating more effective extraction of coupling relationships among multi-energy loads.

Expert module based on informer encoder

The specialist block plays a critical function in isolating temporal signatures and interdependency structures among dissimilar load types within multi-energy load forecasting contexts. Informer constitutes a leading-edge algorithm purpose-built for long-duration time-series characterization [24]. To amplify representational extraction power, the encoder portion of Informer is incorporated as the specialist element within the MTL architecture.

The cornerstone of Informer is the sparse self-attention operation, as rendered in **Figure 4a**. It allows every temporal step in the sequence to engage with every other, thereby capturing far-reaching dependencies and long-span interactions. It realizes this by computing attention coefficients via dot-product calculations between queries (Q) and keys (K), each obtained through linear projection of the input. These coefficients are then applied to weigh the values (V), likewise drawn from the input, to yield a context-conditioned feature representation [40]. The definition of self-attention is written as follows:

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$$A(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where,d_k represents the dimensionality of K, and $\sqrt{d_k}$ functions to temper gradient fluctuations.

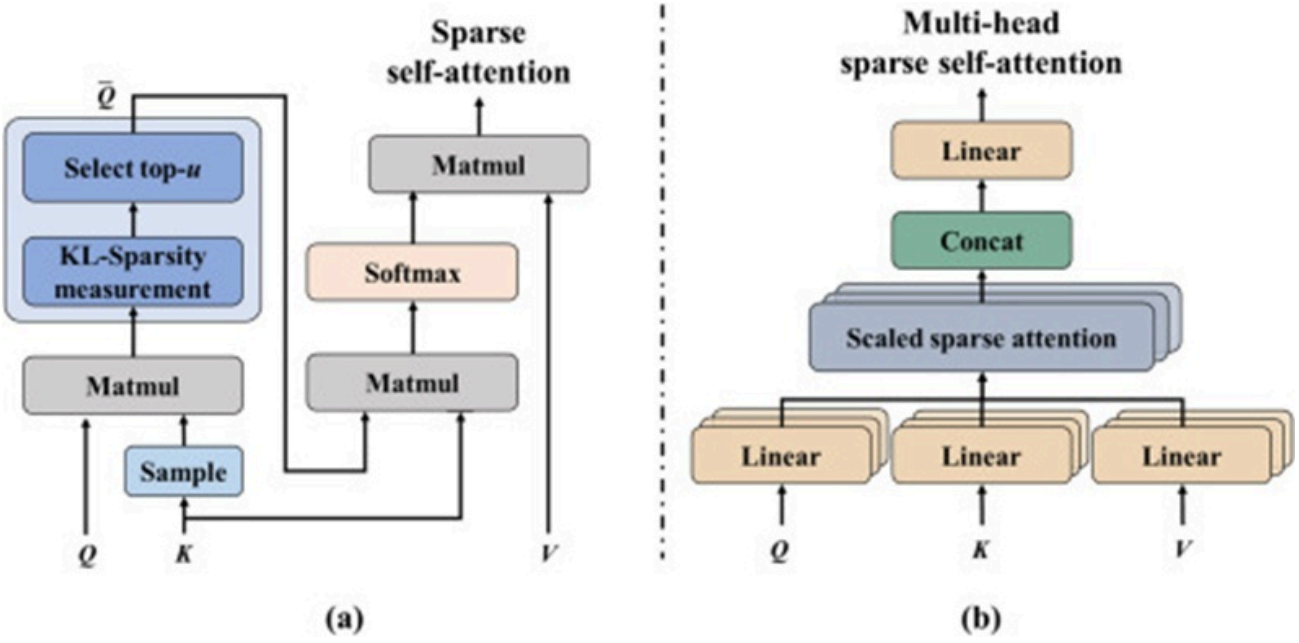


Figure 4. Schematics of sparse self-attention and multi-head sparse self-attention.

To lessen the computational load that accompanies exhaustive time-step processing, the Kullback-Leibler divergence is deployed as a sparsity metric to single out dominant queries, with the sparsity corresponding to the i-th query formulated as follows:

$$\overline{M}(q_i,K) = \max_j\left\{\frac{q_i k_j^T}{\sqrt{d_k}}\right\} - \frac{1}{L_k}\sum_{j=1}^{L_k}\frac{q_i k_j^T}{\sqrt{d_k}} \quad (3)$$

Where,q_i labels the ith query, k_j labels the j-th key, L_k denotes the quantity of randomly drawn keys, and $\overline{M}\ k$ denotes the sparsity measurement. Only the top-u selected queries proceed to the subsequent attention computation, thereby contracting the computational footprint. To further augment the model's ability to discern manifold motifs within temporal data streams, multi-head sparse self-attention is employed. As visualized in **Figure 4b**, Q, K, and V are split across several heads, which bolsters the robustness and the representational flexibility of the attention mechanism.

The Informer encoder is primarily assembled from an embedding block, an encoding block, and a distillation process, as exhibited in **Figure 5**. The embedding block projects the scalar historical observations, the localized timestamp signals, and the global timestamp signals into a common high-dimensional representation space, simplifying model reasoning and numerical processing. The embedding block computation is specified as follows:

$$X_{feed[i]}^t = \alpha u_i^t + PE_{(L_x\times(t-1)+i)} + \sum_p [SE_{(L_x\times(t-1)+i)}]p \quad (4)$$

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Where, u_i^t corresponds to the scalar historical data mapping; PE corresponds to localized timestamp embeddings; SE corresponds to global timestamp embeddings; α acts as a blending weight; and p designates the timestamp category. The encoding block is founded upon multi-head sparse self-attention to encode sequential input. The distillation process further isolates dominant features and contracts dimensionality through a convolutional operation.

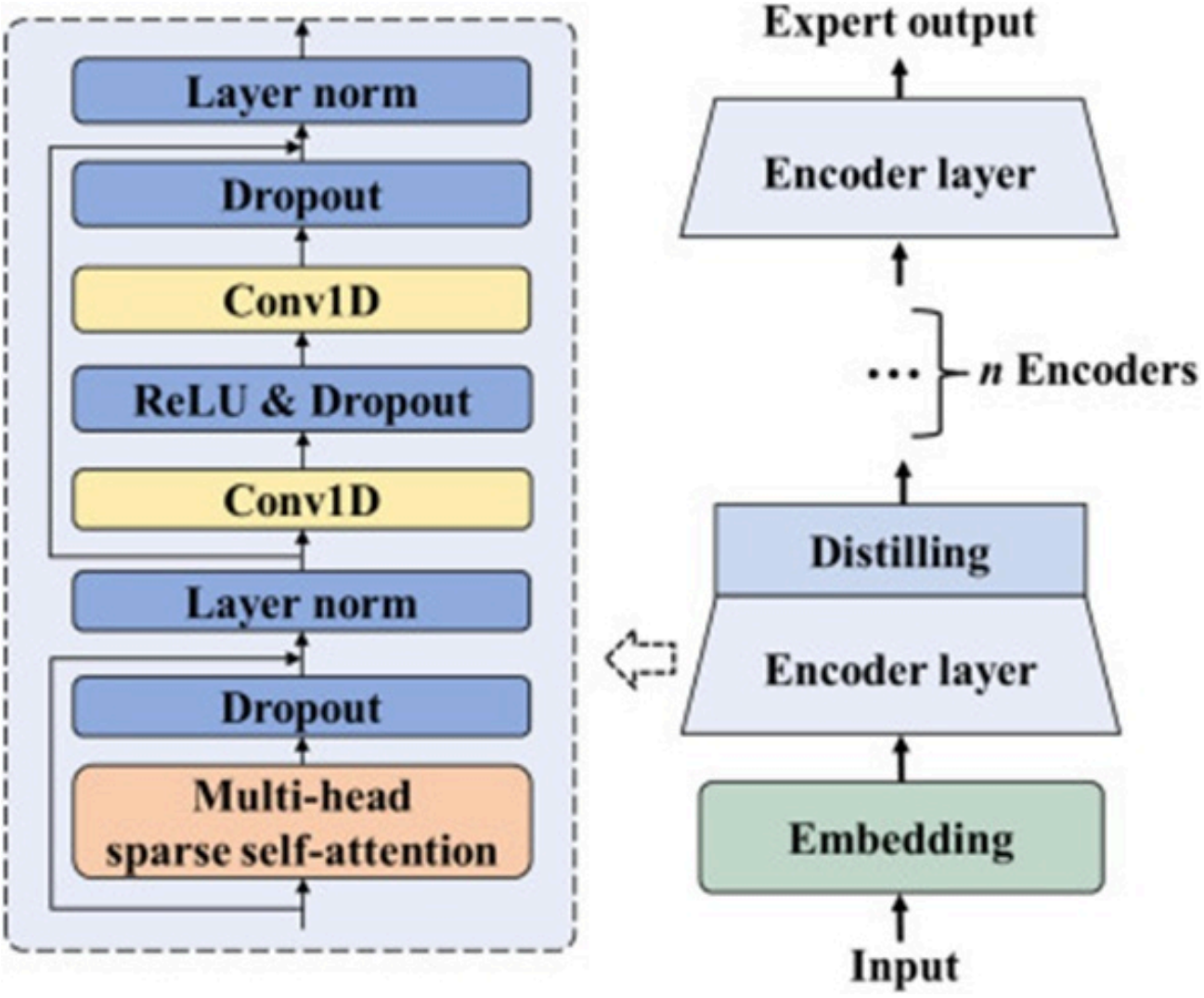


Figure 5. Schematic of Informer encoder.

Stacked cross-attention module for feature fusion

To foster cooperative dynamics among specialists and accommodate the distinctive requirements of each forecasting objective, a stacked cross-attention (SCA) module is put forward to merge the extracted representations originating from several specialists at the sequential level. Functioning like a conductor, the SCA fluidly resolves which specialist-generated signals deserve precedence according to the demands of the associated forecasting task. By enabling task-aware, high-granularity synthesis of multiple specialist representations, the SCA ensures steady and equitable behavior across distinct forecasting targets. It meaningfully counters the predictive disparity issue that frequently surfaces in IES stemming from uneven forecasting complexity among energy load categories.

As delineated in **Figure 6**, the SCA is assembled from several cross-attention tiers, each of which seizes the interplay between a pair of specialist representation streams. Cross-attention is defined analogously to self-attention, except that Q

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

issues from a separate origin than K and V [41]. The output of the SCA module is obtained by summing the outcomes of several cross-attention tiers, as expressed in **Eq. 5.**

$$SCA = \sum_{i=1}^{n} CA_i(Q_A, K_x, V_x), x \in \{A, B, \ldots\} \quad (5)$$

where, x is drawn from the collection of specialists, and the count of elements within that collection is n.

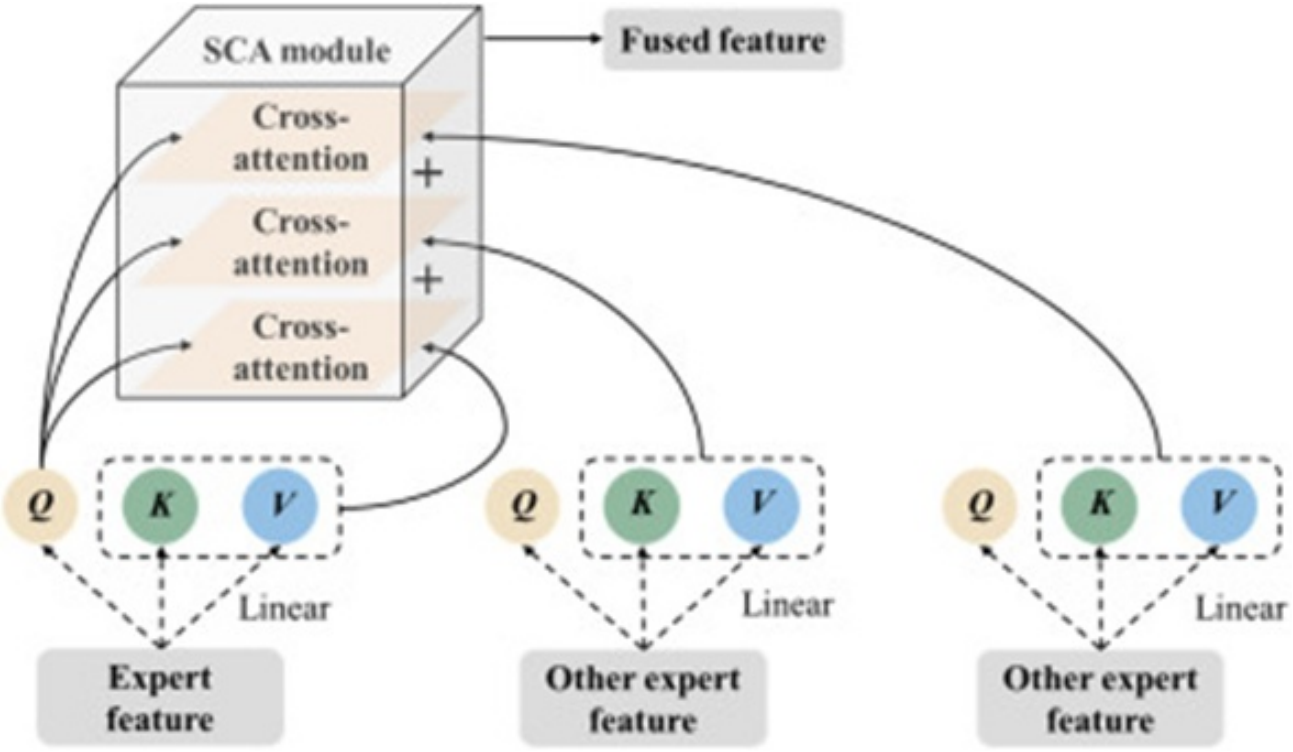


Figure 6. Schematic of SCA module.

Figure 7 renders the schematic of cross-attention; here Q is drawn from the representation stream of specialist A, while K and V are drawn from a separate specialist B. The computation defining cross-attention is formulated as follows:

$$CA(Q_A, K_B, V_B) = \text{Softmax}\left(\frac{Q_A K_B^T}{\sqrt{d_k}}\right) V_B \quad (6)$$

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

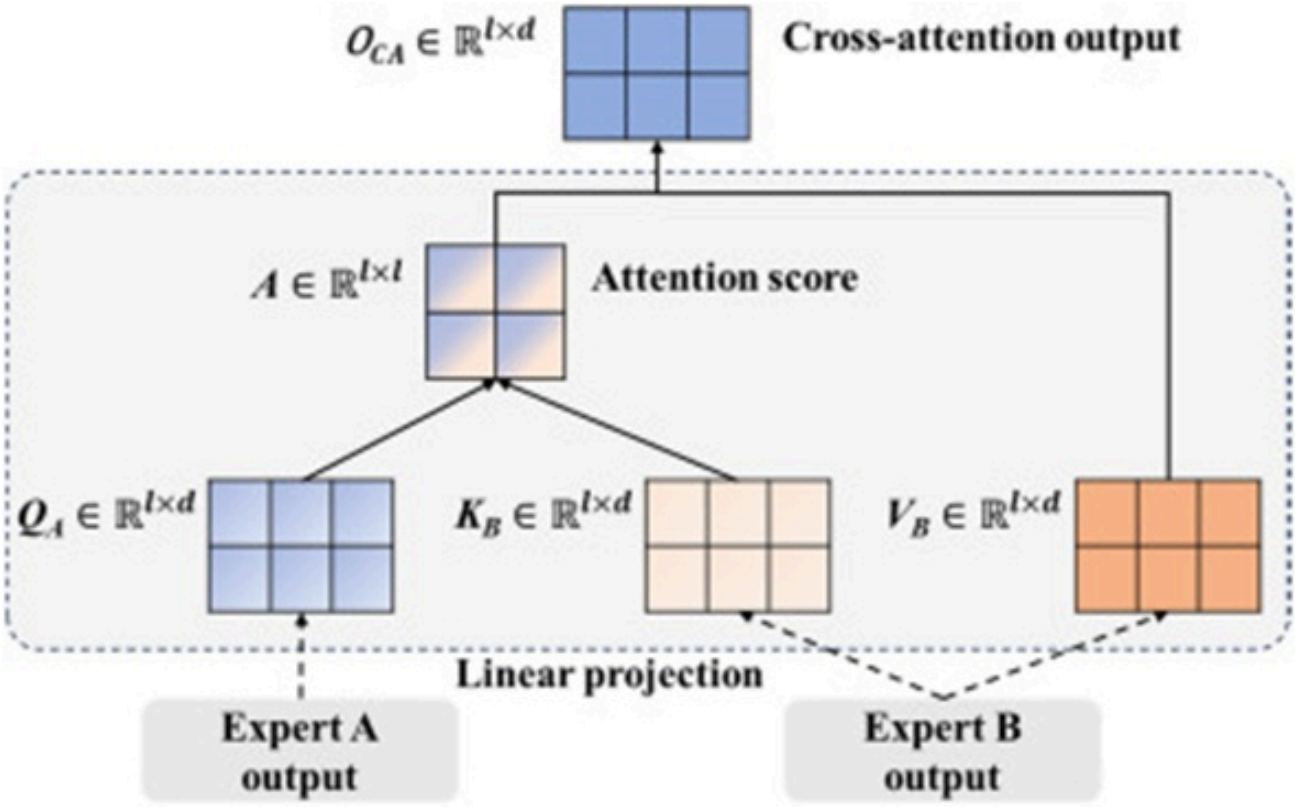


Figure 7. Schematic of cross-attention.

The SCA module enables collaborative participation among specialists and seizes interactions at the sequential level, establishing the basis for acquiring coupling motifs and relational patterns spanning interrelated load forecasting assignments. This elevates generalization capacity and improves the architecture's overall predictive quality.

Proposed SCA-MTL model

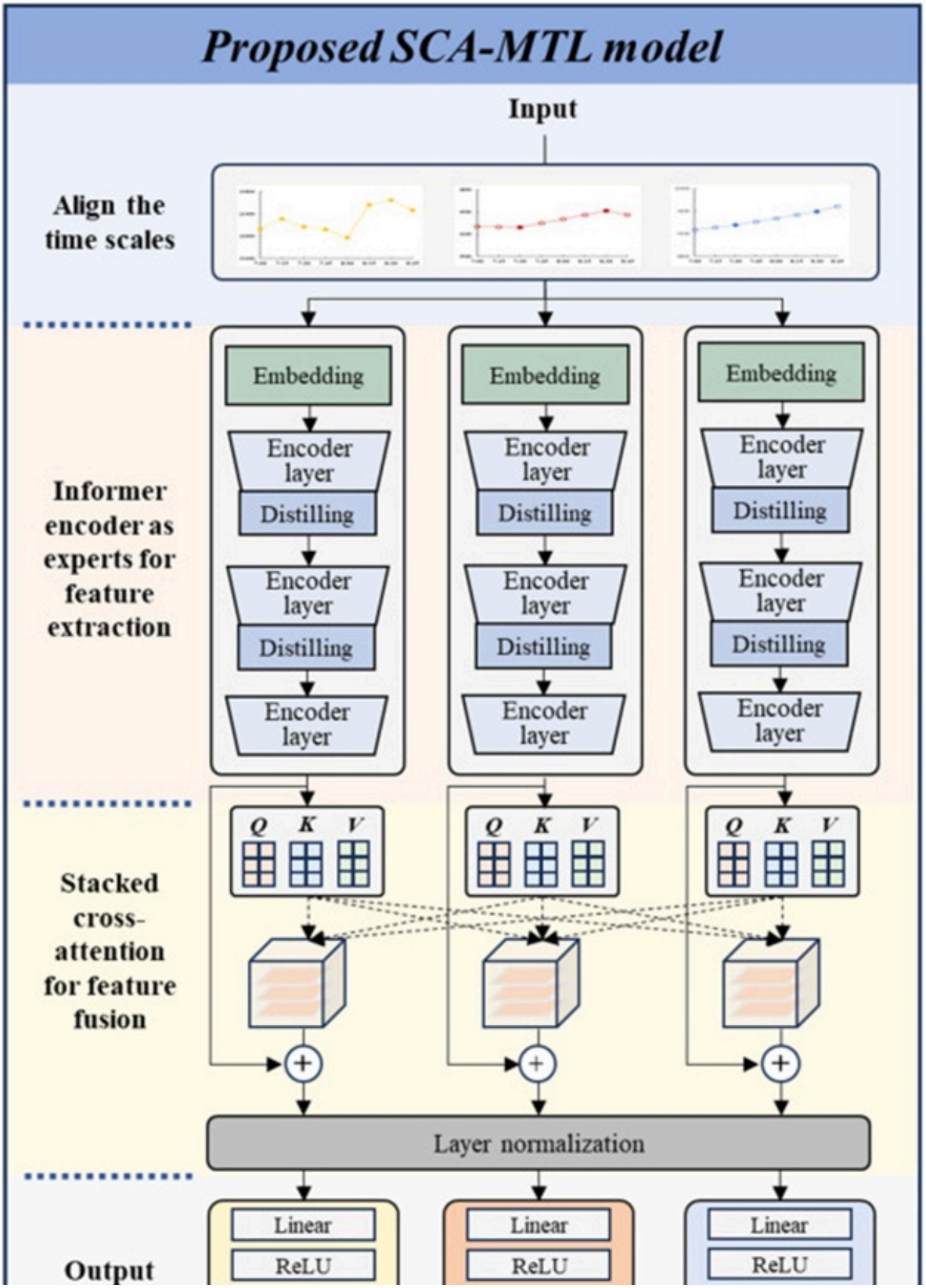
A novel SCA-MTL architecture is introduced for multi-energy load prediction under mixed temporal resolutions. It assembles a temporal resolution reconciliation unit, a specialist module resting on Informer (i.e., Informer expert), an SCA module, and an output module.

The computational flow of the SCA-MTL architecture is shown in **Figure 8**. At the outset, loads with non-uniform temporal resolutions are routed into the front-end phase to equalize the time scales. Following this, the load sequences are fed into several Informer specialists to encode the sequential data. In the present study, three specialists are employed, reflecting standard convention in multi-task learning research [42]. To enable task-tailored feature exploitation, each specialist is followed by an SCA module that fuses the representation streams of the specialists, allowing each forecasting task to selectively draw on information from the others. Residual connections [43] are inserted at the SCA module to guard against the hazards of vanishing or exploding gradients in deeply stacked networks, specified as follows:

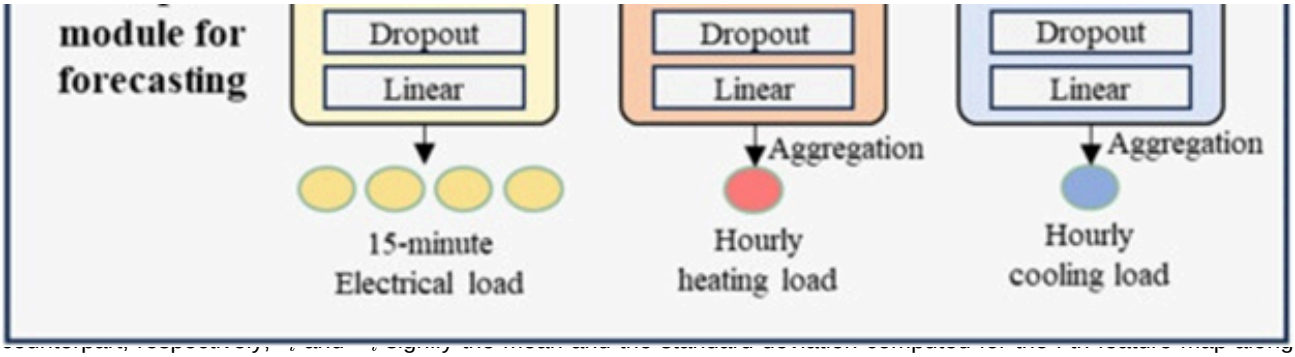
$$H(x_l) = x_{l+1} = F(x_l, W_l) + x_l \quad (7)$$

Where $H(x_l)$ denotes the output of layer l, equivalently the input supplied to layer $l + 1$; $F(x_l, W_l)$ stands for the residual transformation; and denotes the input fed to layer l.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



the channel dimension; and ε acts as a tiny constant to avert division by zero.

Ultimately, the representations specific to each forecasting objective are converted into predicted values through the output block, which is built from a pair of fully connected layers activated by ReLU and interspersed with a Dropout layer for regularization. Independent output blocks are constructed for the electrical, heating, and cooling streams so that each generates predictions tailored to its respective task. The prediction horizon spans 4 time steps, which aligns with 15-minute interval projections covering the next one-hour window. The heating and cooling estimates are subsequently collapsed back to an hourly granularity within the back-end phase, enabling discrepancy computation relative to the observed records. The overall prediction framework can be expressed as follows:

$$\begin{aligned}
&\hat{y}_{e,j},\ldots,\hat{y}_{e,j+4},\hat{y}_{h,j},\hat{y}_{c,j}=f(Y_e,Y_h,Y_c,T,D,P)\\
&\begin{cases}
Y_e=\left[y_{e,j-1},y_{e,j-2},\ldots,y_{e,j-n}\right]\\
Y_h=\left[y_{h,j-1},y_{h,j-2},\ldots,y_{h,j-m}\right]\\
Y_c=\left[y_{c,j-1},y_{c,j-2},\ldots,y_{c,j-m}\right]\\
T=\left[t_{i-1},t_{i-2},\ldots,t_{i-n}\right]\\
D=\left[h_{i-1},h_{i-2},\ldots,h_{i-n}\right]\\
P=\left[p_{i-1},p_{i-2},\ldots,p_{i-n}\right]
\end{cases}
\end{aligned} \tag{9}$$

Where, f stands for the introduced model; Ye denotes the antecedent electrical load sequence sampled every 15 minutes, with n representing its temporal length, fixed at 48; Yh denotes the antecedent hourly heating load sequence, with m representing its temporal length, fixed at 12; Yc denotes the antecedent hourly cooling load sequence; T denotes the outdoor air temperature; D denotes the dew point temperature; P denotes the atmospheric pressure at sea level; $\hat{y}$ denotes the forecasted load value; and the subscripts e, h, and c refer to electrical, heating, and cooling loads, respectively.

Evaluation criteria

The set of performance indicators adopted for assessment comprises the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Variation of the Root Mean Square Error (CV-RMSE), and Mean Absolute Percentage Error (MAPE). Their respective formulas are provided below:

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \quad (10)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (11)$$

$$\mathrm{CV-RMSE} = \frac{100\%}{\bar{y}_i} \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (12)$$

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \quad (13)$$

Where, $\hat{y}_i$ and y_i designate the forecasted and the ground-truth values, respectively; and n indicates the total number of data instances. Within this quartet of indicators, MAE and RMSE are contingent on the magnitude scale of the data, whereas CV-RMSE and MAPE are magnitude-agnostic.

Beyond these, an imbalance (IB) measure is introduced in the present work to capture the degree of unevenness in the model's predictive performance across the different energy types. It is determined by taking the standard deviation of the MAPE scores obtained for each separate load forecasting task. The definition of this imbalance measure is written as:

$$\mathrm{IB} = \sqrt{\frac{\sum_{j=1}^{k} (\mathrm{MAPE}_j - \overline{\mathrm{MAPE}})^2}{k}} \quad (14)$$

Where MAPEj is the MAPE registered for the j-th load forecasting task; $\overline{(\mathrm{MAPE})}$ is the arithmetic mean of the MAPE values over all tasks; and k is the total count of forecasting tasks, set to 6 in this study.

Experimental Settings

This section covers the dataset particulars, the steps taken for data preparation, the hardware and software environment, and the chosen hyperparameter values.

Data description and preprocessing

The data employed in this research originate from the campus metabolism project hosted at Arizona State University [39], zeroing in on the 2020 records of campus-wide hourly heating and cooling demand along with the 15-minute electrical demand, as plotted in **Figure 9.** A marked positive association was evident linking electrical and cooling demand, yielding a correlation coefficient of 0.80; meanwhile, an inverse association was evident between cooling and heating demand, yielding a correlation coefficient of −0.78, as plotted in **Figure 10**. The correlation coefficient is computed according to:

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$$r = \frac{\mathrm{cov}(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}} \quad (15)$$

Where cov symbolizes covariance, σ symbolizes standard deviation, and Xi and Xj denote two distinct input variable sequences. This pattern likely reflects the dominant role of air-conditioning equipment in meeting campus cooling needs, which produces tight coupling among the loads. All load quantities were uniformly expressed in kilowatts (kW), and their summary statistics—minimum (Min), maximum (Max), average (Mean), and standard deviation (Std)—are collated in **Table 1**. Because Arizona State University is located in a subtropical climate zone, the electrical and cooling loads both reach elevated levels, with mean magnitudes of 23,038 kW and 25,930 kW, respectively; the heating load, in contrast, stays at a far lower average of only 2176 kW. The weather variables fed into the model were obtained from the National Centers for Environmental Information (NCEI) database for Phoenix during 2020 [44]. They included outdoor air temperature, dew point temperature, and sea-level atmospheric pressure, each of which showed sizable correlations with the load variables and was therefore supplied as a predictor input.

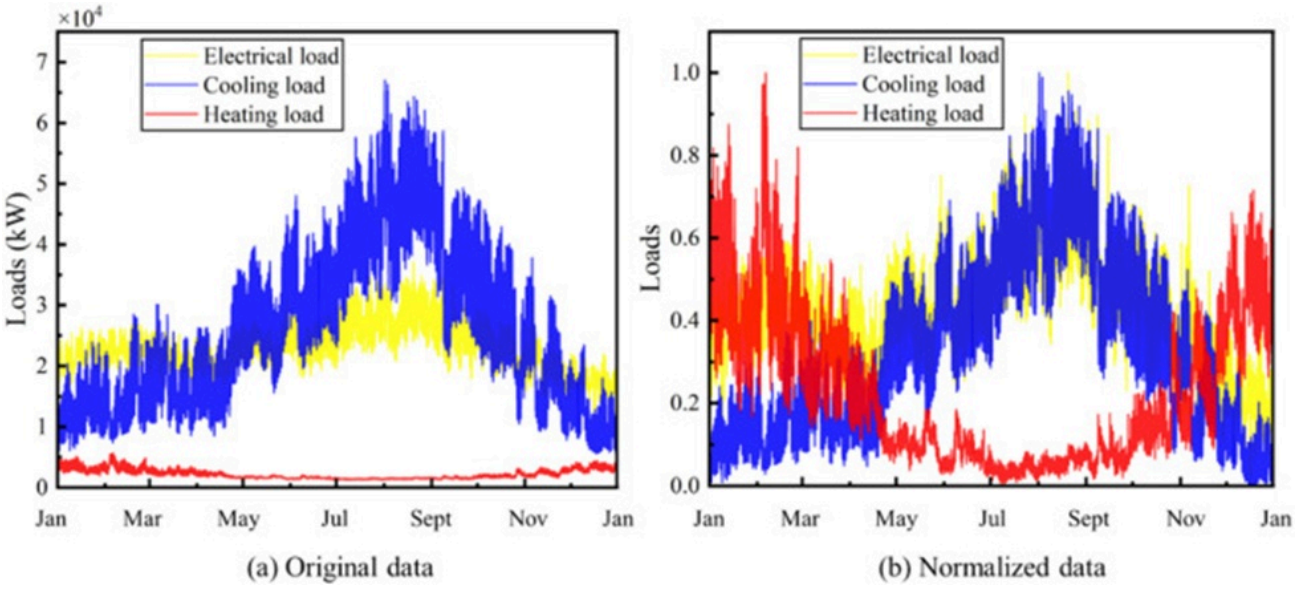


Figure 9. Multi-energy load curves utilized in this study [39].

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

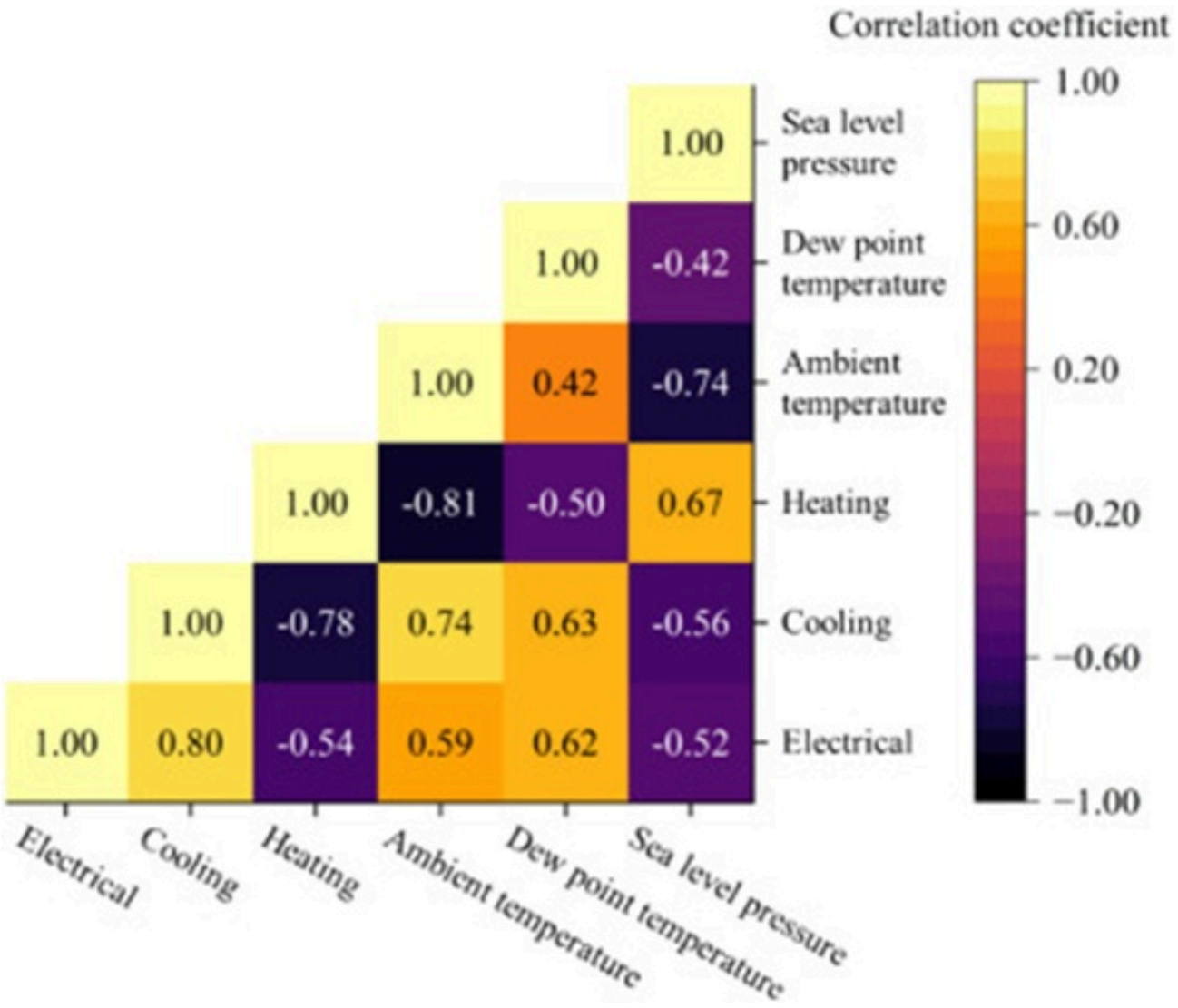


Figure 10. Correlation coefficients of different inputs.

Table 1. Descriptive statistics of loads.

Statistics	Heating load (kW)	Cooling load (kW)	Electrical load (kW)
Min	1241.59	5607.36	12,154.00
Max	5662.96	66,999.06	37,116.46
Mean	2176.09	25,930.67	23,038.51
Std	778.63	13,700.53	3670.36

The full dataset was stochastically divided into a training portion and a held-out test portion using an 8:2 proportion. Normalization was applied during preprocessing to dampen the influence of mismatched variable ranges.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Experimental configurations

All trials were run on a system built around a high-performance server housing an Intel Xeon Platinum 8255C CPU clocked at 2.5 GHz, 40 GB of system memory, and an NVIDIA RTX3080 graphics processor. The software stack comprised PyTorch version 1.11.0 operating over Python 3.8. The hyperparameter specification adopted for the proposed model is itemized in **Table 2.**

Table 2. Hyperparameters of the proposed model.

Module	Description
Informer expert	*d_model* = 512, *n_heads* = 8, *e_layers* = 3
SCA	*d_model* = 512, *n_heads* = 8, *number of cross-attention* =3
Output	*linear*1_ *out_features*=32, *linear*2_ *out_features*=1, *dropout*=0.2

Over the course of training, the Mean Squared Error (MSE) served as the optimization objective, and the Adam updater was applied to iteratively refine the model weights with a step size of 10–5. An early termination safeguard was activated to halt training before overfitting set in, with the patience counter configured to 3 in this study.

Results and Discussion

Throughout the following section, every evaluation was performed in a mixed-time-resolution setting, entailing 15-minute electrical load forecasts and hourly heating and cooling load forecasts across the upcoming hour. The benefits conferred by the SCA component and the Informer-based specialists were substantiated through ablation trials and comparative experiments. In addition, the presented SCA-MTL framework was measured against MMoE-style architectures, conventional forecasting benchmarks, and recently proposed high-performing models to confirm its forward-looking nature. Forecast traces, error metric analyses, and uncertainty assessments were utilized to interpret the findings, with interpretive discussion integrated alongside them.

Comparisons for time scale alignment modules

To more effectively reconcile the disparate temporal granularities characterizing electrical, heating, and cooling load records, four candidate strategies were trialed as the time scale alignment component: cubic spline interpolation, quadratic interpolation, linear interpolation, and direct reuse of the original coarse values. The three interpolation-centered strategies were selected due to their prevalent application in the imputation of missing values [45, 46]. In contrast, the direct-reuse baseline was incorporated to examine whether interpolation itself introduces any counterproductive distortion.

The MAPE scores recorded by the model under each alternative time scale alignment approach are assembled in **Table 3**. Linear interpolation consistently secured the top predictive accuracy across most time steps and load categories, as indicated by bold highlighting. This outcome suggests that its inherent simplicity and smoothness are particularly compatible with the typical temporal profiles found in multi-energy load series, especially where abrupt discontinuities are rare. By contrast, although cubic spline and quadratic interpolation can yield more fluid transitions, they risk overfitting or injecting spurious undulations when the underlying trend is fairly stable. Compared with the direct-reuse baseline, linear interpolation better equips the model to perceive evolving tendencies and temporal linkages, thereby improving predictive precision. Owing to its consistently stronger showing, linear interpolation is selected to fulfill the time scale alignment function within the model.

Table 3. MAPEs of model under different time scale alignment modules.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Alignment module	Heating	Cooling	Electrical-60	Electrical-45	Electrical-30	Electrical-15
Cubic spline interpolation	3.88 %	2.50 %	2.50 %	2.22 %	1.99 %	1.81 %
Quadratic interpolation	**3.32 %**	2.66 %	2.45 %	2.16 %	1.98 %	1.87 %
Original data filling	3.61 %	2.57 %	2.54 %	2.23 %	2.02 %	1.74 %
Linear interpolation	3.63 %	**2.39 %**	**2.30 %**	**2.02 %**	**1.76 %**	**1.53 %**

Ablation experiments of SCA module

To establish both the indispensability and the performance edge of the SCA module, a set of ablation trials was orchestrated. In addition to contrasting the full model against a version lacking the SCA component, two other widely recognized soft sharing strategies were brought into the comparison: a vanilla self-attention block and the gating circuit drawn from the Multi-gate Mixture-of-Experts (MMoE) design. The model variants under test were arranged as follows: 1) Model A: vanilla self-attention operating as the soft sharing mechanism; 2) Model B: MMoE-style gating circuit operating as the soft sharing mechanism; 3) Model C: the SCA module entirely removed; 4) Proposed Model: the SCA module operating as the soft sharing mechanism. Outside of these designated differences, every other architectural element was held fixed across all variants.

Within Model A, the concatenated outputs discharged by all specialist networks were jointly routed through a self-attention block, after which the resulting representations were distributed to each output head for final prediction. The self-attention block was parameterized with d_model = 512 and n_head = 8. Within Model B, every task was paired with its own gating unit, which linearly projected the task-specific input into a vector of scalar weights assigned across the pool of specialists. These weights then governed a weighted average over the specialist outputs, thereby constructing a bespoke feature representation for each task.

Given that summer and winter impose markedly distinct load regimes, the predictive behaviors observed during these two seasonal windows on the test partition were singled out for granular examination. **Figure 11** portrays the multi-energy load forecasting results delivered by the different soft sharing variants throughout the summer period. The full proposed model, armed with the SCA module, reliably outperforms the alternatives across the entire set of forecasting objectives, generating estimates that adhere more closely to the ground-truth trajectories, particularly through episodes of pronounced load volatility. This enhanced fidelity can be attributed to the SCA module's capacity to formally encode cross-task relational structures and to conduct representation exchanges at the sequential granularity. Relative to Model B, which applies a static set of mixture weights to specialist outputs via a gating operation, and Model A, which conducts wholesale representation merging through a standard self-attention operation, the SCA module proves markedly more effective at nurturing inter-task synergy and at distilling task-distinctive signal content.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

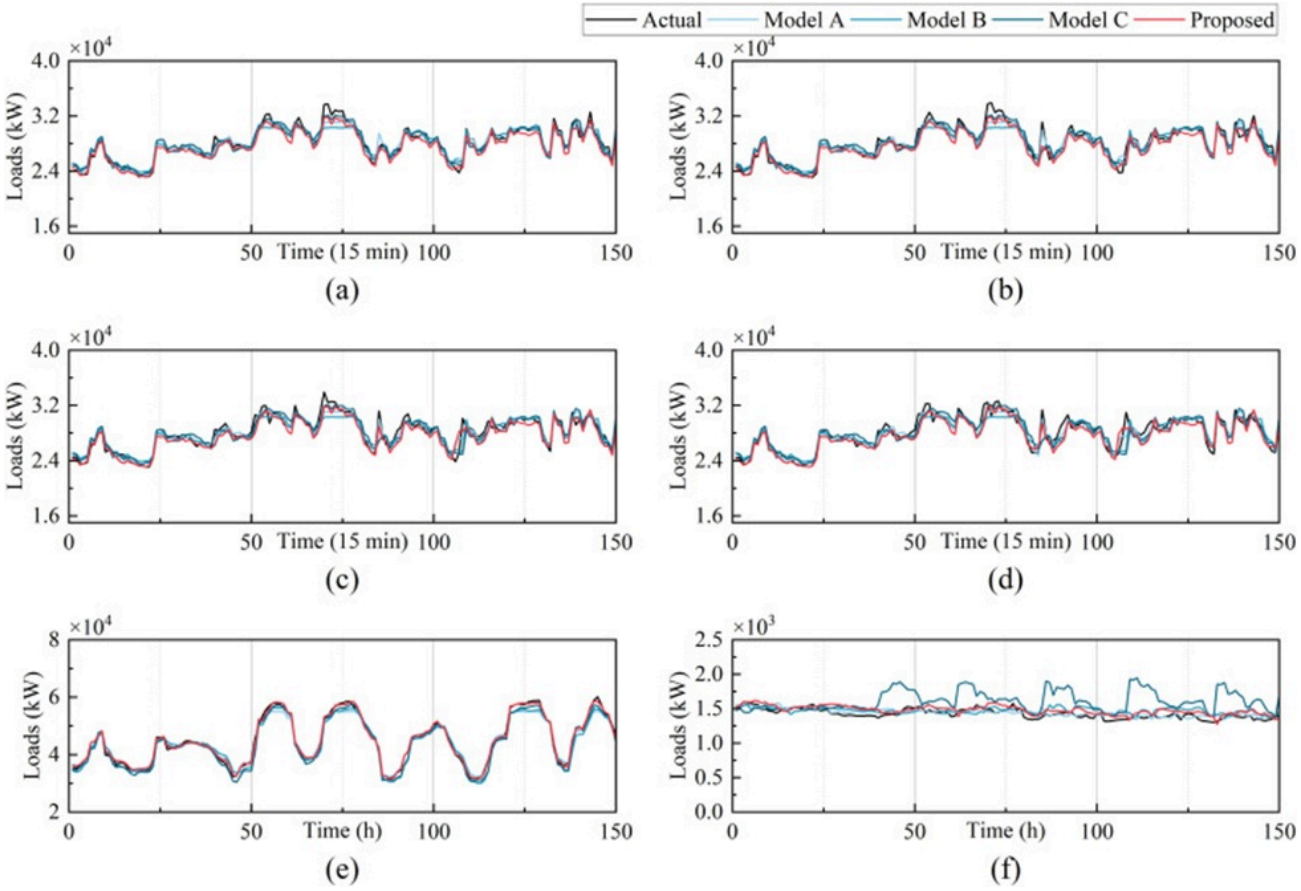


Figure 11. Multi-energy load forecasting performance of models with different soft sharing mechanisms in summer.

Worth particular emphasis is the observation that Model C, which operates bereft of the SCA module, registers large-magnitude errors in its heating load forecasts, an outcome plausibly traceable to contamination from the overriding cooling load signatures. In the summer season, cooling demand is both larger in scale and richer in structured variation, whereas heating demand remains subdued and relatively featureless. An architecture that leans on a solitary specialist with no soft sharing provisions lacks the capacity to disentangle the faint heating load dynamics under such lopsided conditions. By contrast, the SCA module draws upon the diversified competencies of several specialists, opening the door to considerably more faithful predictions.

Figure 12 displays the multi-energy load prediction outcomes obtained by the distinct soft sharing designs over the winter interval. The introduced architecture equipped with the SCA component preserves its commanding advantage. Departing from the summer regime, winter conditions involve heightened thermal demand that follows more conspicuous and well-ordered temporal rhythms. Winter conditions involve heightened thermal demand that follows more conspicuous and well-ordered temporal rhythms. As a result, Model C registers a discernible uptick in heating load prediction fidelity. Even so, sizable residual inaccuracies remain evident. These observations jointly bear out both the indispensable function and the practical value of the SCA component.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

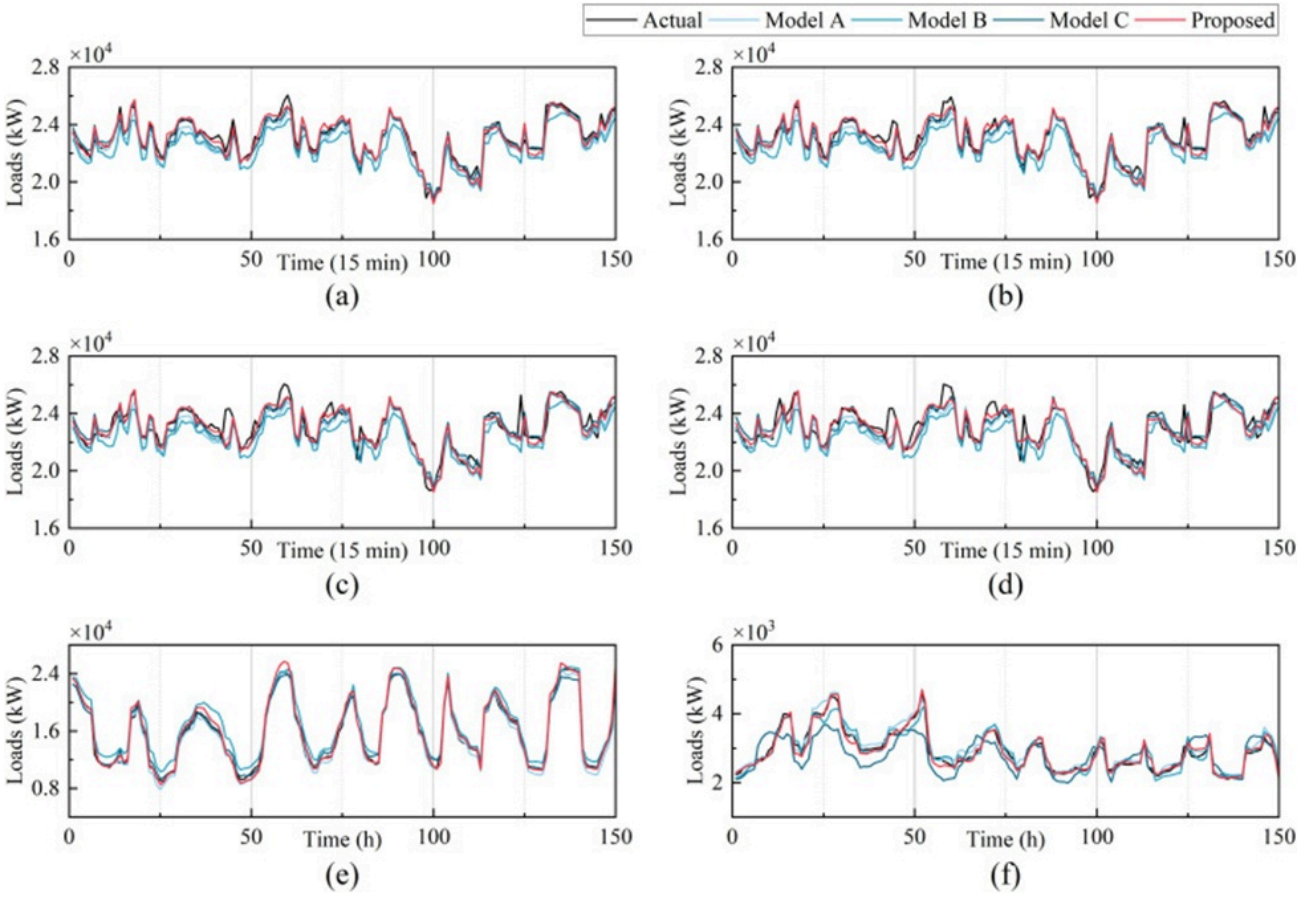


Figure 12. Multi-energy load forecasting performance of models with different soft sharing mechanisms in winter.

Figure 13 catalogues the MAE, RMSE, CV-RMSE, and MAPE scores logged by the competing soft sharing variants when assessed against the entirety of the test partition. A diminished metric value signals enhanced model competence. Barring the single instance where Model A posts an incrementally superior MAPE on the heating load, the introduced model attains the most advantageous scores across every other measured dimension. When attention turns to the scale-agnostic indicators CV-RMSE and MAPE, the performance asymmetry afflicting Model C becomes substantially more glaring. To cite a concrete example, the MAPE readings for Model C remain under 4% on electrical and cooling loads, yet surge dramatically to 11% on the heating load. To render the extent of predictive asymmetry numerically tractable, the imbalance index (IB) put forward was brought to bear. The computed IB figures for Model A, Model B, Model C, and the Proposed Model stand at 7.52×10^{-3}, 1.00×10^{-2}, 3.24×10^{-2}, and 6.74×10^{-3}, respectively. The introduced model exhibits the faintest degree of predictive asymmetry, securing a 79.17% contraction when measured against Model C, which functions bereft of the SCA component.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

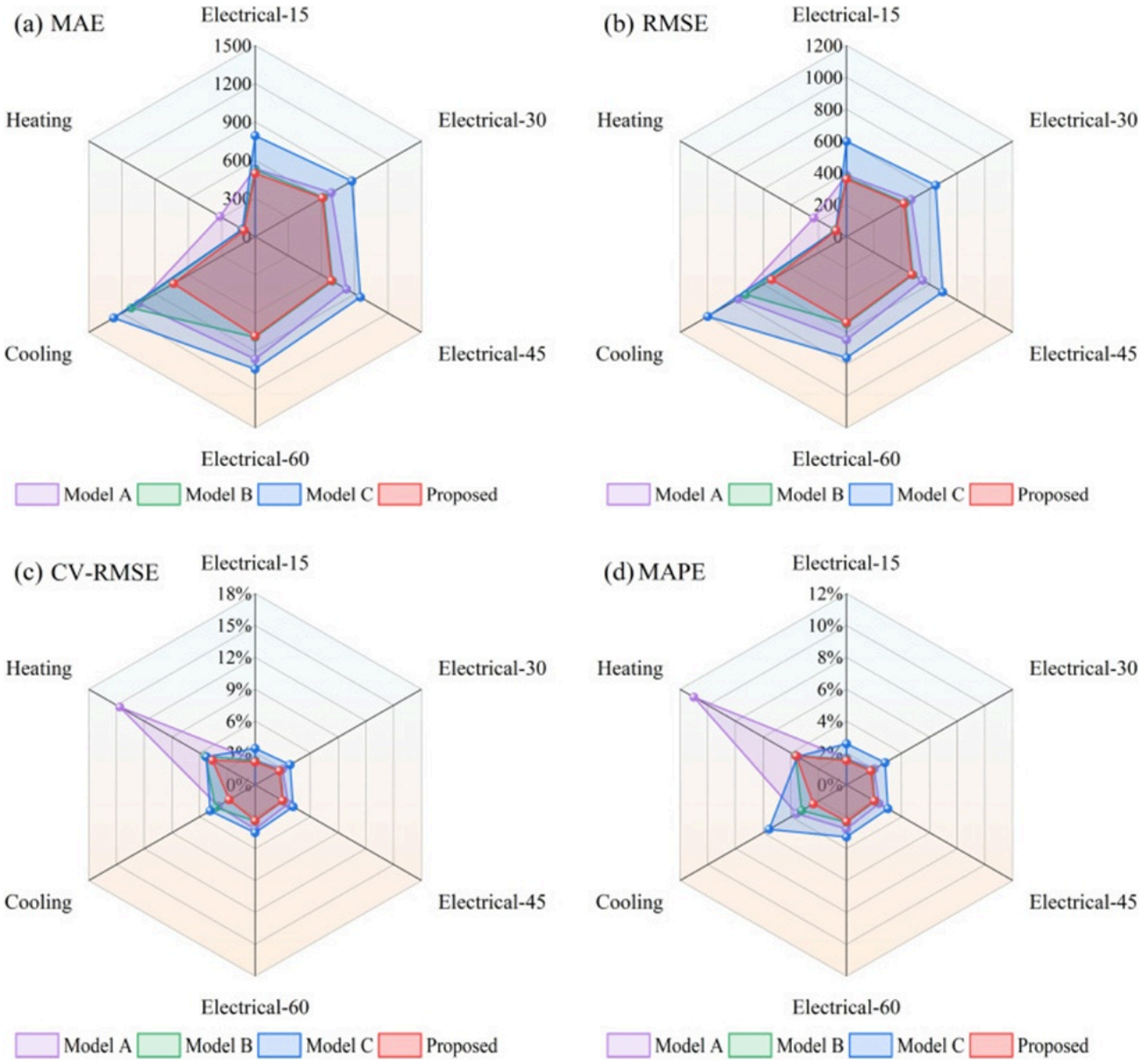


Figure 13. Comparisons of models on different evaluation criteria.

Overall, the SCA component enables the model to generate multi-energy load forecasts that are more faithful to observations and more uniformly reliable across tasks, corroborating both its essentiality and its performance ascendancy.

Analysis for the improvement of informer-based expert module

To substantiate the edge held by Informer when cast in the specialist role, the specialist blocks within the comparator architectures were swapped out for CNN and LSTM, two workhorses routinely applied to sequential data processing. Every other constituent of the comparator architectures was kept rigorously identical to the introduced SCA-MTL framework to eliminate confounding factors.

Figure 14 renders the multi-energy load forecasting traces produced by architectures fitted with the alternative specialist designs across the summer season. With Informer occupying the specialist position, the generated predictions closely follow the ground-truth trajectories, revealing a clear performance advantage. When LSTM is instead stationed as the

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

specialist, the forecasts stray noticeably from the observed values, with the divergence most acute during episodes of vigorous load fluctuation. When the specialist role is handed to CNN, predictive capability degrades to merely shadowing coarse directional movements, with substantial mismatches materializing at most time indices. Since this portion of the investigation isolates the effect attributable to the specialist choice, the SCA component was preserved uniformly across all versions. The evidence indicates that while the specialist architectures differ in forecasting accuracy, they each maintain a comparatively even performance spread across the different load modalities.Informer harnesses the self-attention operation to jointly reason over the entire input span in parallel, a property that confers intrinsically more powerful representational reach relative to LSTM. In addition, energy load trajectories are characterized by strong serial dependence, and the self-attention operation proves inherently more adept at latching onto such structure. These factors account for why the Informer specialist comes out ahead in load prediction contexts. CNN, on the other hand, operates by sliding learned filters across the input to harvest localized patterns [47]. Its forte lies principally in spatial motif detection, leaving it comparatively ill-equipped to model sequential dependencies. The upshot is that when CNN is called upon as the specialist, the resultant architecture turns in relatively lackluster multi-energy load forecasts.

The multi-energy load prediction results yielded by architectures housing the different specialist designs over the winter season are collected in **Figure 15**. The aggregate performance ordering among the specialist choices echoes the summer pattern, settling into the sequence Informer first, followed by LSTM, and lastly CNN, with the separation especially pronounced for heating load forecasts. When model behavior is contrasted between the summer and winter windows, deploying Informer as the specialist module consistently produces the most favorable outcomes, reinforcing its prowess in isolating informative signatures from multi-energy load streams.

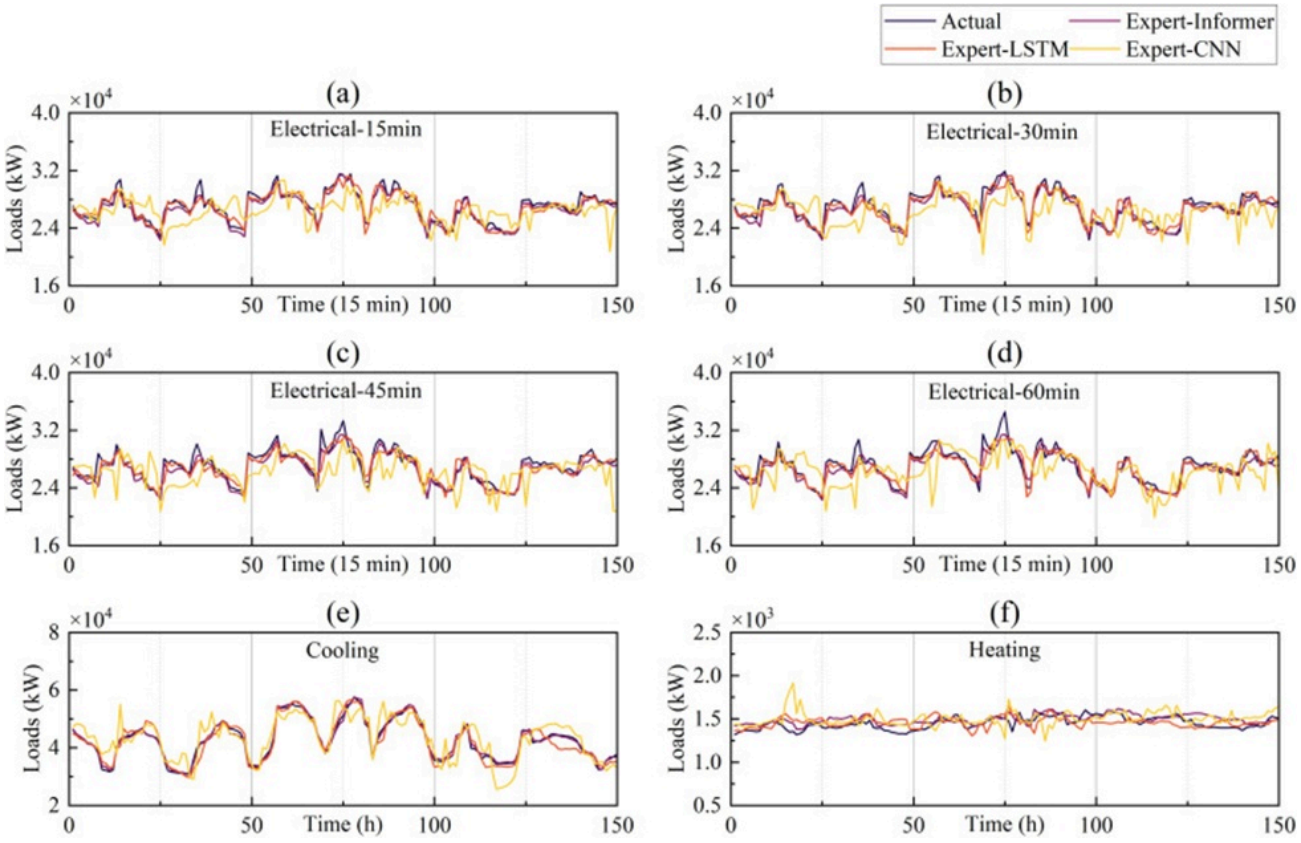


Figure 14. Multi-energy load forecasting performance of models with different expert modules in summer.

Informer harnesses the self-attention operation to jointly reason over the entire input span in parallel, a property that confers intrinsically more powerful representational reach relative to LSTM. Added to this, energy load trajectories are characterized by strong serial dependence, and the self-attention operation proves inherently more adept at latching onto

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

such structure. These factors account for why the Informer specialist comes out ahead in load prediction contexts. CNN, on the other hand, operates by sliding learned filters across the input to harvest localized patterns [47]. Its forte lies principally in spatial motif detection, leaving it comparatively ill-equipped to model sequential dependencies. The upshot is that when CNN is called upon as the specialist, the resultant architecture turns in relatively lackluster multi-energy load forecasts.

The multi-energy load prediction results yielded by architectures housing the different specialist designs over the winter season are collected in **Figure 15**. The aggregate performance ordering among the specialist choices echoes the summer pattern, settling into the sequence Informer first, followed by LSTM, and lastly CNN, with the separation especially pronounced for heating load forecasts. When model behavior is contrasted between the summer and winter windows, the deployment of Informer as the specialist module unfailingly produces the most favorable outcomes, reinforcing its prowess in isolating informative signatures from multi-energy load streams.

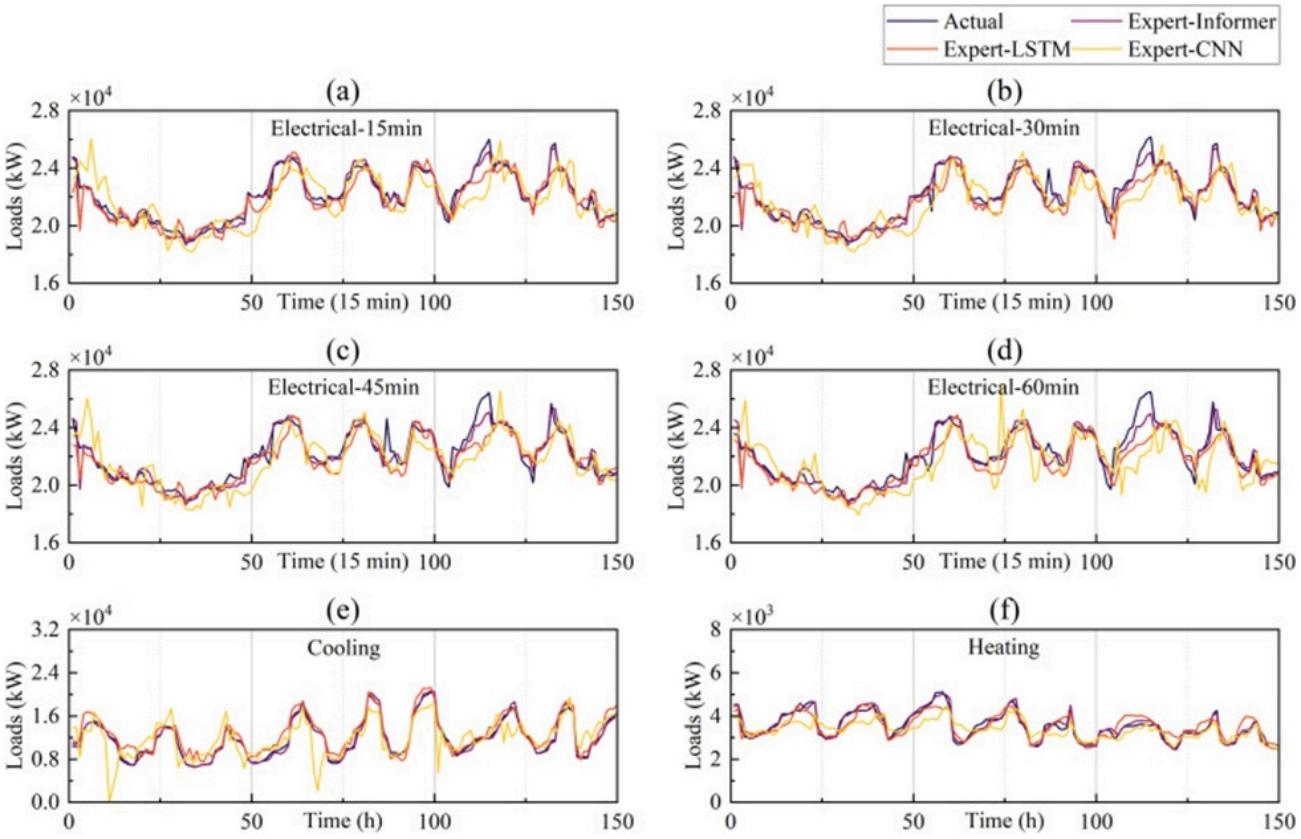
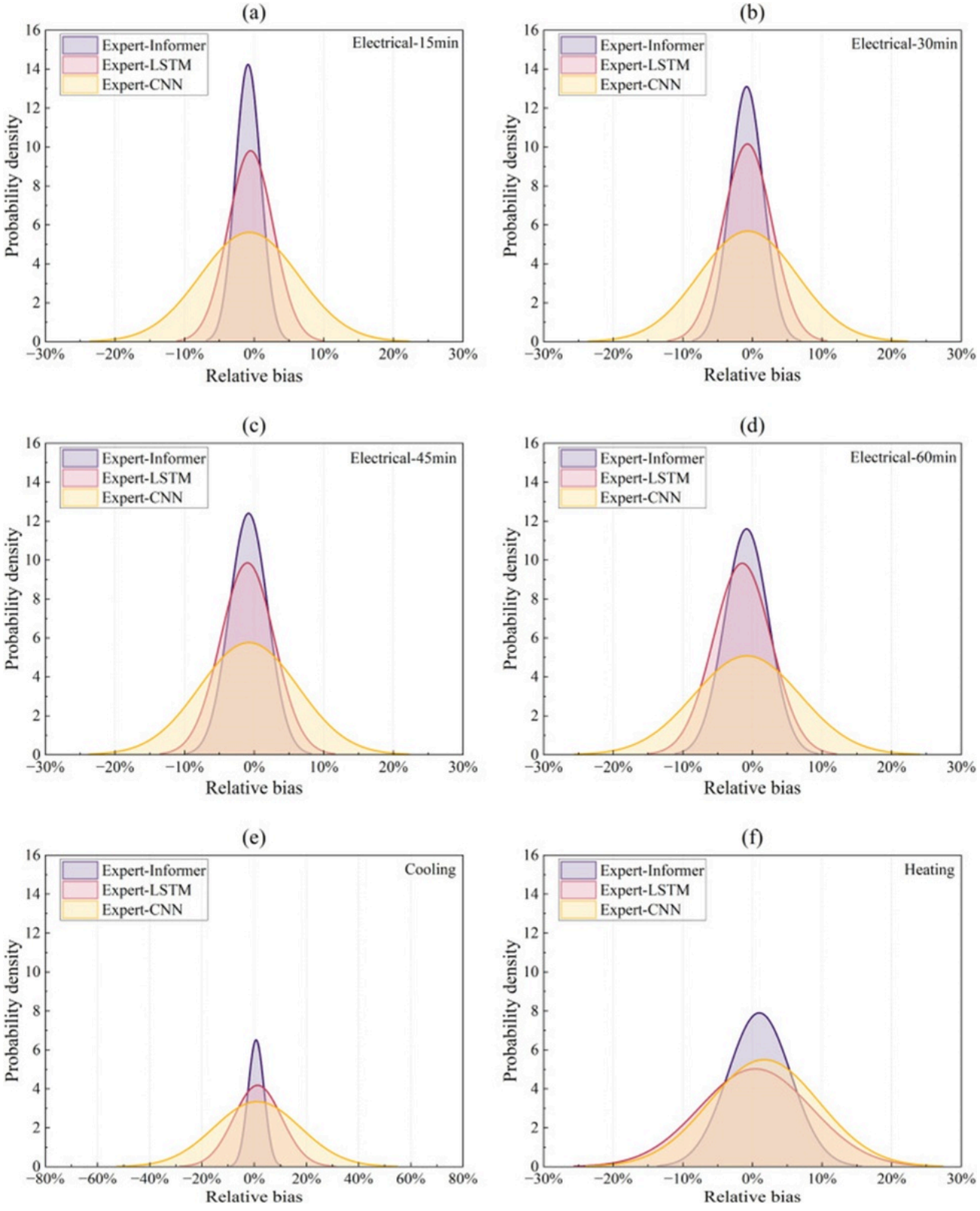


Figure 15. Multi-energy load forecasting performance of models with different expert modules in winter.

Uncertainty characterization was carried out to inspect the relative bias distributions of architectures constructed around the various specialist choices when applied to the full test corpus, as graphed in **Figure 16**. This characterization provides a lens for comparing model resilience. The vertical dimension maps probability density, while the horizontal dimension maps relative bias. The Informer specialist concentrates the greatest probability mass in the immediate neighborhood of zero, indicative of the smallest overall relative deviation. When LSTM and CNN are slotted into the specialist position, the density amassed around zero is comparatively meager and dispersed across a wider band. To illustrate, within **Figure 16a**, the bias envelope for Informer, LSTM, and CNN spans ± 7.5%, ± 10%, and ± 25%, respectively. This signals that with LSTM or CNN acting as the specialist, the architecture suffers from diminished sturdiness, a frailty that can manifest as outsized errors under particular conditions. Beyond this, the Informer specialist yields a mean relative bias very near zero across the full set of forecasting tasks, indicating unbiased estimation. LSTM and CNN, in contradistinction, display marked asymmetry in the fourth-step electrical load prediction and the heating load prediction, respectively. The

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

performance statistics tallied for architectures equipped with the various specialist designs are assembled in **Table 4**. When measured against LSTM, the Informer specialist effects MAPE contractions spanning 30.09% to 61.64%; when measured against CNN, the contractions span 42.65% to 78.74%.



© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 16. Relative bias distributions of models with different expert modules.

Table 4. Evaluation criteria of models with different expert modules.

Criteria	Loads	Expert-CNN	Expert-LSTM	Expert-Informer
MAE	Electrical-15	1054.23	581.67	**361.12**
	Electrical-30	1067.59	620.13	**415.79**
	Electrical-45	1067.93	684.24	**472.98**
	Electrical-60	1202.46	765.48	**539.19**
	Cooling	1611.15	1231.61	**539.48**
	Heating	132.15	140.08	**75.45**
RMSE	Electrical-15	1394.26	792.71	**496.92**
	Electrical-30	1411.66	862.14	**607.75**
	Electrical-45	1409.72	954.24	**688.70**
	Electrical-60	1538.67	1054.39	**780.01**
	Cooling	2073.69	1600.07	**734.58**
	Heating	187.64	194.84	**99.85**
CV-RMSE	Electrical-15	6.04 %	3.43 %	**2.15 %**
	Electrical-30	6.12 %	3.73 %	**2.63 %**
	Electrical-45	6.11 %	4.13 %	**2.98 %**
	Electrical-60	6.67 %	4.57 %	**3.38 %**
	Cooling	7.93 %	6.12 %	**2.81 %**
	Heating	8.68 %	9.01 %	**4.62 %**
MAPE	Electrical-15	4.55 %	2.49 %	**1.53 %**
	Electrical-30	4.60 %	2.65 %	**1.76 %**
	Electrical-45	4.61 %	2.93 %	**2.02 %**
	Electrical-60	5.17 %	3.29 %	**2.30 %**
	Cooling	8.01 %	6.23 %	**2.39 %**
	Heating	5.81 %	6.25 %	**3.63 %**

In summary, the Informer-anchored specialist module provides the strongest predictive capability while preserving resilience and stability, a combination that supports the engineering of Informer to operate in expert capacity.

Comparisons with SOTA models

The discussion now turns to benchmarking the presented SCA-MTL framework against three families of comparator approaches. The first family draws from MMoE-style designs, a class of architectures that has found broad adoption within the multi-task learning community. The second family assembles a collection of time-tested forecasting baselines

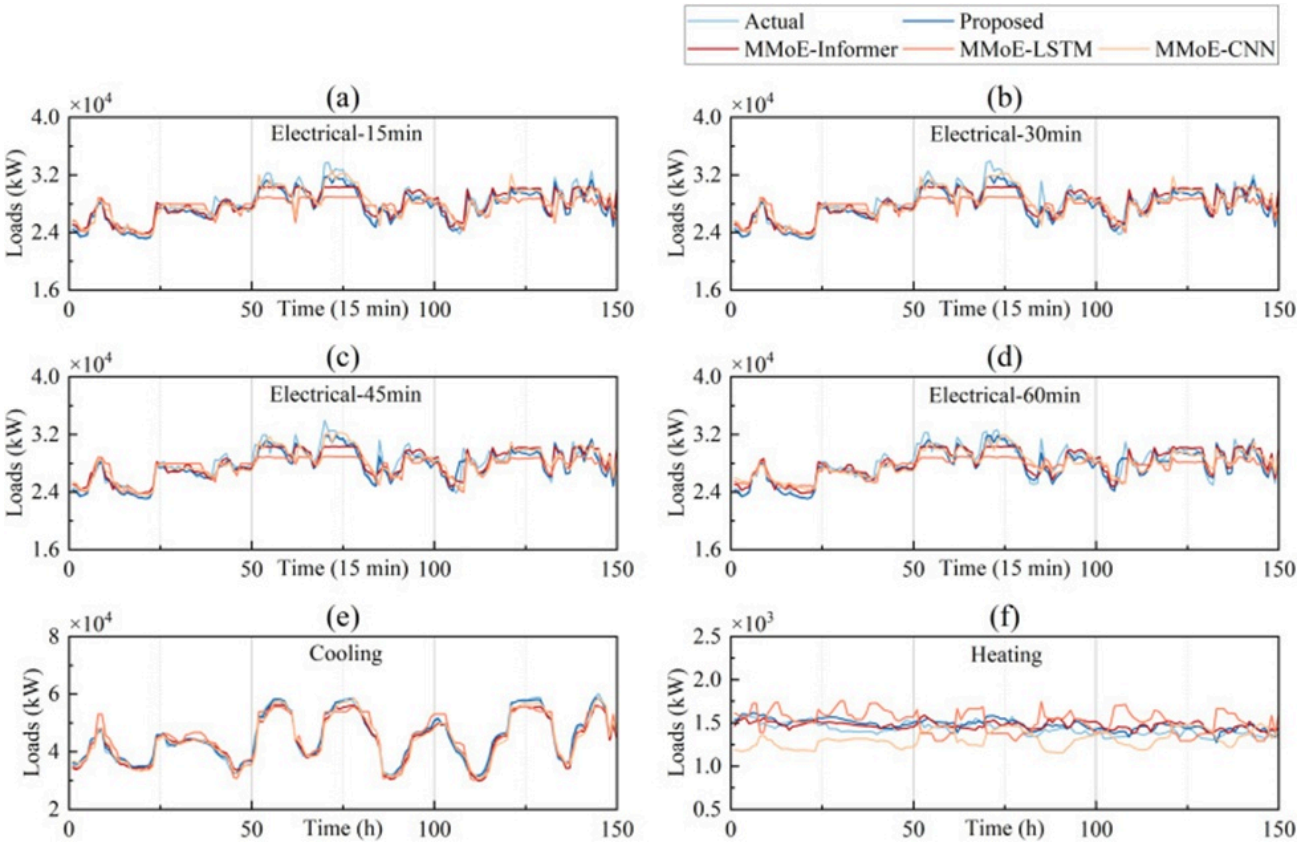
© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

intended to function as general-purpose yardsticks. The third family includes recently devised cutting-edge architectures to situate the presented work relative to the current performance frontier.

Comparisons with MMoE-based models

The MMoE paradigm constitutes a multi-task learning formulation that has delivered compelling outcomes across numerous application domains [42]. Its foundational concept rests on employing a gating function that metes out scalar mixing coefficients to the outputs of distinct specialist sub-networks, thereby addressing the bespoke requirements posed by each prediction target. For the analysis in this subsection, a trio of MMoE-grounded variants, each harnessing a different class of specialist module, was assembled to enable an exhaustive side-by-side appraisal against the SCA-MTL framework put forward in this paper.

Figure 17 captures the multi-energy load prediction trajectories yielded by the presented SCA-MTL framework together with those of the three MMoE-grounded variants throughout the summer months. The presented SCA-MTL framework claims the top position within this head-to-head evaluation. Among the MMoE cohort, the variant termed MMoE-Informer approaches the predictive caliber of the presented SCA-MTL design most closely. Its generated forecasts generally align with the observed values, yet they remain second-best relative to the SCA-MTL framework when demand reaches its zenith. The underlying reason stems from the SCA-MTL framework's capacity to navigate the interplay among representations sourced from multiple specialists at the granularity of individual temporal positions; this permits a more nuanced excavation of cross-specialist relationships and a more responsive adaptation to the exigencies of multi-energy load prediction. The variants MMoE-LSTM and MMoE-CNN lag substantially further behind. As inspection of **Figure 17a** reveals, the forecast trace shows no inflection whatsoever during peak loading intervals, with the predicted magnitudes situated well beneath the actual observations. This finding dovetails with the earlier analytical conclusion that positioning Informer in the specialist role begets the strongest predictive results.



© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 17. Multi-energy load forecasting performance of the proposed SCA-MTL model and three MMoE-based models in summer.

Figure 18 charts the multi-energy load prediction behavior recorded over the winter interval, where the presented SCA-MTL framework once again demonstrates the most advantageous forecasting characteristics. In contrast to what was observed under summer conditions, the MMoE-grounded variants post improved accuracy on the electrical and cooling load forecasts. This improvement can be ascribed to the reduced amplitude of electrical and cooling consumption during winter, paired with the greater prevalence of comparable samples throughout the dataset. These circumstances render the governing load dynamics more amenable to discovery.

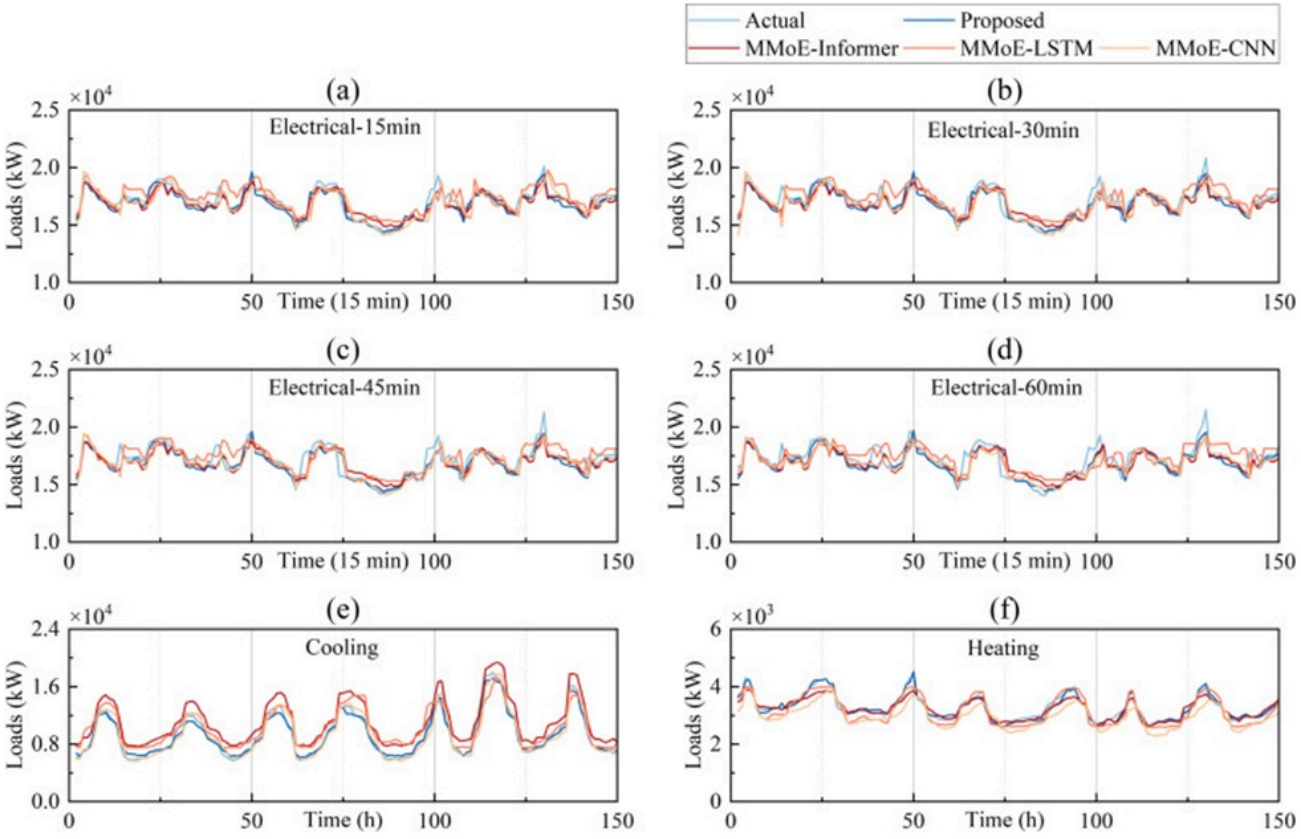


Figure 18. Multi-energy load forecasting performance of the proposed SCA-MTL model and three MMoE-based models in winter.

A violin-plot representation is enlisted to juxtapose the absolute deviation distributions characterizing the various architectures, as rendered in **Figure 19.** Within each paired violin, the left-hand lobe maps to the presented SCA-MTL framework, while the right-hand lobe maps to the three MMoE-grounded variants. The spread of the violin at any ordinate value conveys the density of deviations falling at that magnitude. Spanning the complete suite of prediction tasks, the presented SCA-MTL framework amasses a greater proportion of its absolute deviations tightly clustered around the zero mark. Taking a broader view, however, the MMoE-grounded architectures exhibit larger absolute deviations, along with more broadly splayed violin contours. Moreover, with the lone exception of the heating load estimates generated by MMoE-Informer, the MMoE-grounded variants veer away from a mean deviation centered at zero over several prediction targets, an effect most striking in the cooling and heating load forecasts, where a pronounced asymmetry in the distribution is visible.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

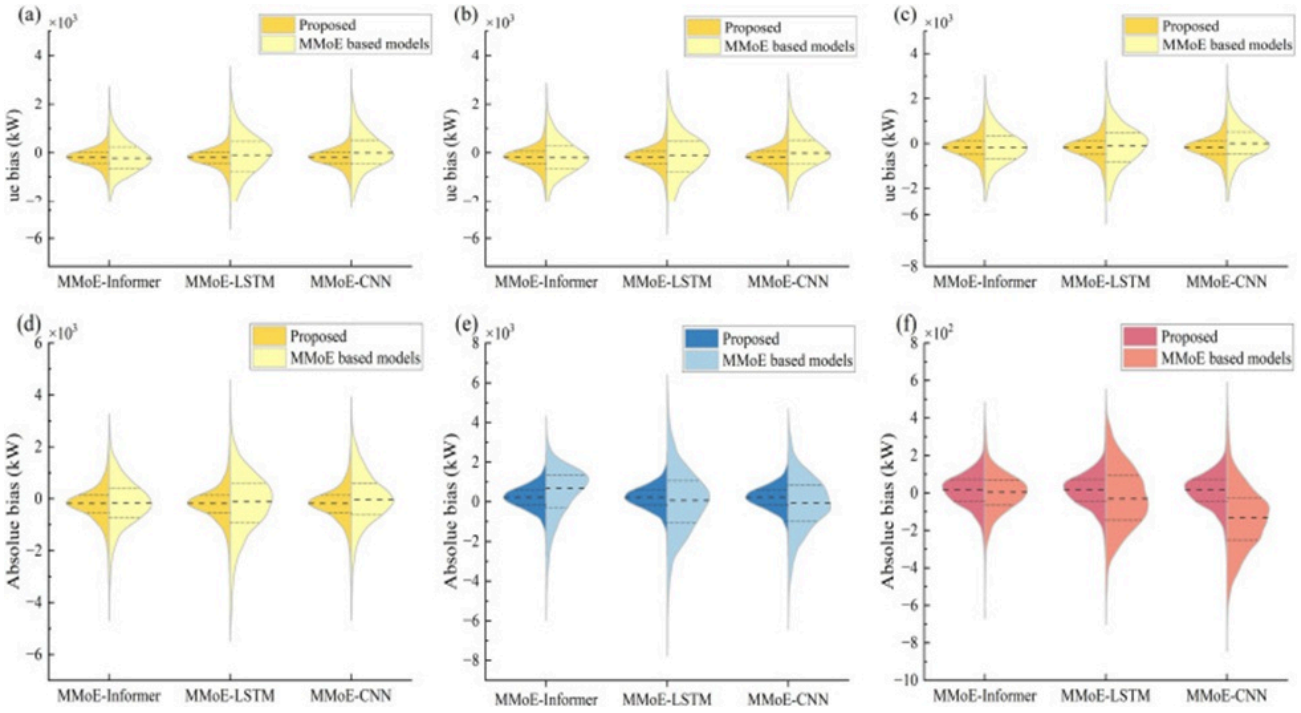


Figure 19. Absolute bias distributions of the proposed SCA-MTL model and three MMoE-based models.

The suite of performance indicators computed for the presented SCA-MTL framework and the three MMoE-grounded variants is organized in **Table 5**. When placed alongside MMoE-Informer, which attains the top marks within the MMoE family, the presented SCA-MTL framework suffers only a negligible deficit on heating load prediction; across all remaining tasks, its MAPE readings are cut by margins ranging from 29.66% to 57.17%.

Table 5. Evaluation criteria of the proposed SCA-MTL model and three MMoE-based models.

Criteria	Loads	MMoE-CNN	MMoE-LSTM	MMoE-Informer	Proposed
MAE	Electrical-15	832.23	862.65	598.08	**361.12**
	Electrical-30	836.81	876.31	644.07	**415.79**
	Electrical-45	825.10	882.02	693.62	**472.98**
	Electrical-60	858.58	999.09	760.16	**539.19**
	Cooling	749.34	1173.30	1001.86	**539.48**
	Heating	167.65	139.15	76.98	**75.45**
RMSE	Electrical-15	1132.04	1234.47	791.26	**496.92**
	Electrical-30	1135.84	1248.66	872.88	**607.75**
	Electrical-45	1121.01	1255.09	949.23	**688.70**
	Electrical-60	1147.59	1352.43	1038.47	**780.01**
	Cooling	943.62	1571.39	1275.08	**734.58**
	Heating	209.40	183.38	113.80	**99.85**

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

CV-RMSE	Electrical-15	4.90 %	5.35 %	3.43 %	**2.15 %**
	Electrical-30	4.92 %	5.41 %	3.78 %	**2.63 %**
	Electrical-45	4.86 %	5.44 %	4.11 %	**2.98 %**
	Electrical-60	4.97 %	5.86 %	4.50 %	**3.38 %**
	Cooling	3.61 %	6.01 %	4.87 %	**2.81 %**
	Heating	9.69 %	8.49 %	5.27 %	**4.62 %**
MAPE	Electrical-15	3.68 %	3.62 %	2.58 %	**1.53 %**
	Electrical-30	3.70 %	3.69 %	2.78 %	**1.76 %**
	Electrical-45	3.65 %	3.71 %	2.99 %	**2.02 %**
	Electrical-60	3.80 %	4.24 %	3.27 %	**2.30 %**
	Cooling	4.05 %	5.46 %	5.58 %	**2.39 %**
	Heating	8.30 %	6.76 %	**3.54 %**	3.63 %

Taking stock of the evidence, the presented SCA-MTL framework yields more capable multi-energy load prediction, a property that flows from its capacity to adjudicate the entanglement of distinct specialist representations at the resolution of each time step through the cross-attention-driven SCA construct. MMoE, by comparison, assigns a single global scalar weighting to each specialist's contribution, informed solely by the input; this tactic proves insufficiently nimble when confronting the labyrinthine and volatile character of multi-energy load prediction, ultimately manifesting as subpar performance on select forecasting objectives.

Comparisons with classical forecasting models

A set of established forecasting architectures—namely Transformer, BiLSTM, CNN-LSTM, and GRU—were pressed into service as benchmarks against which the presented SCA-MTL framework was evaluated. **Figure 20** chronicles the MAPE scores registered by the rival architectures on the electrical, cooling, and heating load prediction targets. The presented SCA-MTL framework posts the lowest MAPE figure on every single forecasting objective, a result that corroborates its forward-looking design. The architecture that falls nearest behind is Transformer, which likewise uses self-attention to process load sequences and demonstrates a degree of competitiveness within the comparison cohort. That said, when the focus shifts to cooling and heating load prediction, the MAPE differentials become considerably wider, a manifestation of the bounded representational power that a solitary model can bring to bear on the tangled problem of multi-energy load forecasting.

Furthermore, stripped of the capacity to unearth the relational fabric connecting prediction targets in the fashion that multi-task learning affords, these single-target architectures betray a forecasting imbalance phenomenon, most vividly illustrated by outsized errors concentrated on the heating load. BiLSTM traces a broadly analogous contour, marked by inflated error metrics and lopsided predictive reliability. CNN-LSTM and GRU labor under still higher aggregate inaccuracies and, in an anomalous departure, register curiously large errors specifically on the cooling load forecast, a liability that disqualifies them from satisfactorily discharging the prediction remit.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

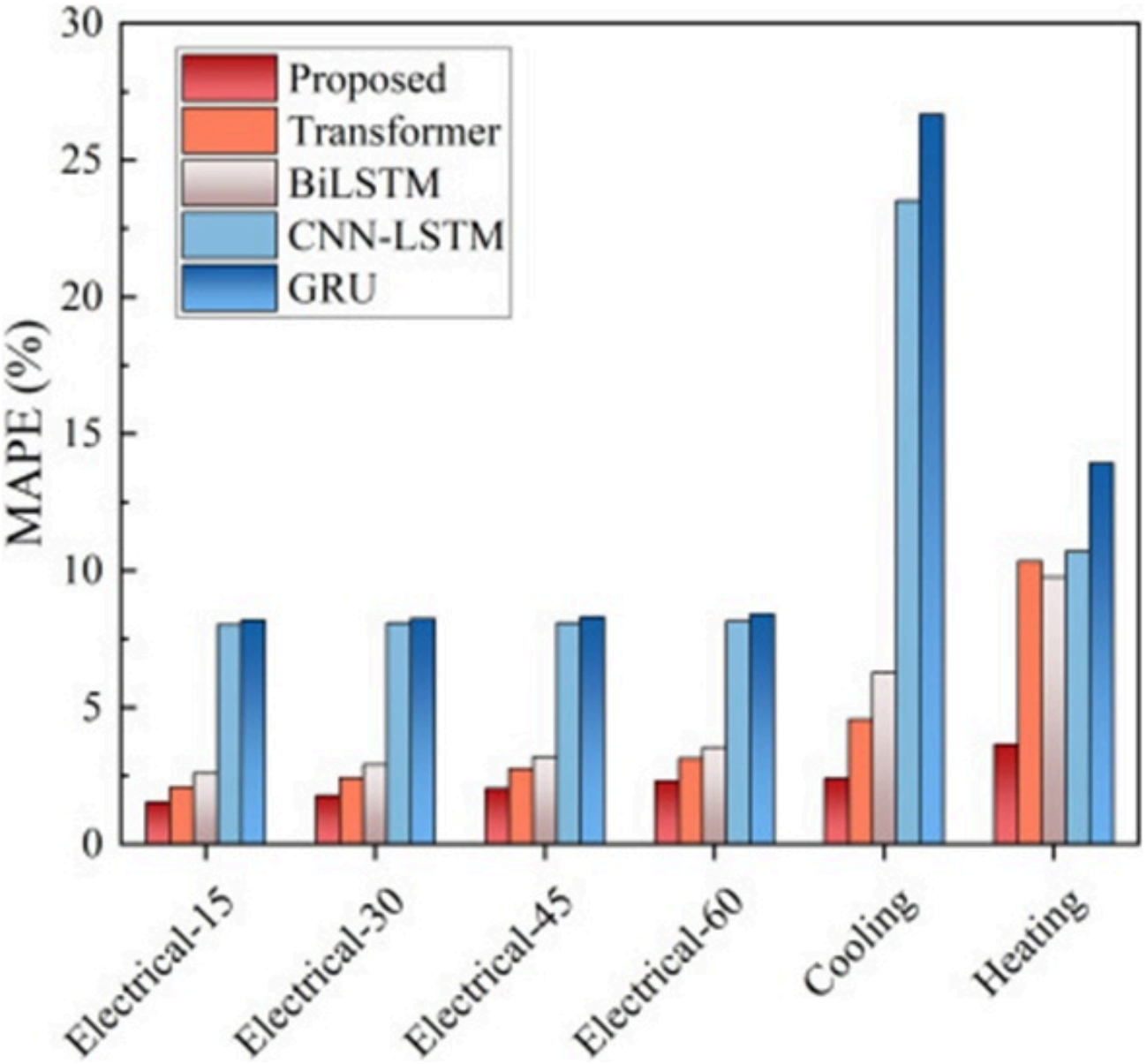


Figure 20. MAPEs of the proposed SCA-MTL model and classical forecasting models.

A computational throughput audit pitting the presented architecture against the foregoing baseline models was also performed. Specifically, both the cumulative inference latency across the entire test partition and the latency attributable to a single sample were tracked as proxies for model efficiency, with the figures collated in **Table 6**. The measurements reveal that the presented architecture expends a mere 3.71 × 10–4 s to process each sample. While this duration exceeds that logged by the baseline comparators, the added cost remains negligible within operational deployment settings such as 15-minute or hourly load forecasting cycles. This slender increase in computational expenditure is judged to represent a reasonable trade-off for the meaningful enhancement in prediction fidelity that the presented architecture delivers.

Table 6. Computational efficiency comparison of models.

Model	Inference time per sample	Inference time on the test set
MGMOE-Informer	3.71×10^{-4} s	2.61 s

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

MGMOE-CNN	6.06×10^{-5} s	0.43 s
MGMOE-LSTM	1.24×10^{-4} s	0.87 s
Transformer	3.50×10^{-5} s	0.25 s
BiLSTM	6.74×10^{-5} s	0.47 s
CNNLSTM	3.27×10^{-5} s	0.23 s
GRU	5.03×10^{-5} s	0.35 s
Proposed	5.39×10^{-4} s	3.78 s

Comparisons with recent advanced models

To gain a more complete appraisal of how the presented framework measures up against the contemporary frontier, its predictive accuracy was set alongside that of several recently introduced high-performing architectures tailored to multi-energy load forecasting, drawing upon MAPE statistics reported in the originating publications. For works that evaluated performance on a seasonal basis, the average MAPE across all four seasons for each load category was used as the reference. Given that the presented framework issues four sequential 15-minute-ahead electrical load estimates for the coming hour, the mean of these four values was used as the summary indicator for the electrical prediction component.

Table 7 provides a consolidated overview of the comparative exercise. The presented framework attains the strongest predictive exactness on both the electrical and cooling load categories. Even though its heating load prediction accuracy lands fractionally below that registered by certain competing approaches, the presented framework nonetheless delivers the most competitive all-around performance characteristics. Worth underlining is the fact that the presented framework functions within a mixed-time-resolution prediction paradigm that subdivides into six constituent sub-tasks—encompassing four 15-minute electrical forecasts alongside single hourly forecasts for heating and cooling respectively—a setup that renders the overall prediction problem markedly more taxing than the set of three single-temporal-resolution tasks entertained within the comparator studies referenced above. The collected evidence substantiates both the competitive footing and the forward-leaning nature of the presented framework within this area of inquiry.

Table 7. Overall comparison with recent advanced models.

Model	Average	Cooling	Heating	Electrical
DAM-MSHRN-GRU [48]	4.03 %	4.13 %	4.32 %	3.63 %
MTL-LSTM [26]	2.83 %	3.09 %	3.02 %	2.38 %
MTL-CNN-BiLSTM [49]	2.66 %	2.46 %	**2.85 %**	2.68 %
Adp-3EncInformer-KAN [31]	3.00 %	3.38 %	3.09 %	2.53 %
MTL-TCNHU [46]	3.45 %	4.26 %	3.26 %	2.82 %
Proposed	**2.64 %**	**2.39 %**	3.63 %	**1.90 %**

Across an exhaustive set of head-to-head evaluations against the existing corpus of prediction architectures, the presented SCA-MTL framework demonstrates an ability to realize both superior exactness and a more level distribution of predictive dependability when applied to multi-energy load forecasting under non-uniform temporal resolutions. This dual achievement reflects its overall performance superiority.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Conclusion

IIn this paper, a novel SCA-MTL architecture has been proposed, purpose-built for multi-energy load prediction in settings where temporal granularities differ across load types, and shown to deliver elevated accuracy together with a fairer apportionment of forecasting fidelity among tasks. The value added by the SCA construct and the strengths flowing from the Informer encoder-grounded specialist modules have been examined in depth, with quantitative substantiation supplied at each stage. In addition, MMoE-style designs, long-standing forecasting baselines, and recently published advanced models were mobilized as state-of-the-art benchmarks to bear out the forward-looking character of the presented architecture. The chief findings can be distilled as follows:

• The presented architecture, equipped with a module dedicated to reconciling heterogeneous temporal resolutions, successfully surmounts the obstacle that mixed time granularities pose within multi-energy load forecasting.

• The SCA construct is engineered to conduct detailed manipulation and merging of specialist-derived representations to fulfill the particularized requirements of task-differentiated load prediction. It achieves the strongest forecasting results in the ablation trials. It reduces the forecasting disparity index by 79.17%, highlighting its essentiality and performance edge in boosting model capability while correcting the imbalance common to multi-energy load forecasting.

• The specialist block anchored on the Informer encoder bolsters the representational extraction prowess of the architecture, translating into gains in predictive performance. When set against specialist realizations built upon LSTM and CNN, the Informer-grounded specialist yields MAPE contractions in excess of 30.09 %, all while imparting enhanced sturdiness and consistency to the overall architecture.

• Measured against state-of-the-art reference models, the presented architecture returns the smallest MAPE values across the great preponderance of prediction targets. These outcomes cast into sharp relief the merits of its comprehensive architectural scheme and corroborate the advantages that the presented architecture holds within the multi-energy load forecasting arena.

The presented architecture is amenable to widespread adoption across the operational scheduling contexts of IES, supplying accurate and well-balanced multi-energy load projections under mixed temporal-resolution conditions. In future investigations, tools such as attention weight visualization and SHapley Additive exPlanations (SHAP) could be brought to bear to delve further into how representations stemming from separate specialists are combined and mutually shape one another, thus strengthening the transparency of the architecture. Moreover, although the comparative analyses undertaken in this study have affirmed the architecture’s performance lead, subsequent research should pivot toward embedding the presented forecasting architecture directly within operational IES environments to gauge the benefits that materialize at the system level.

Acknowledgements

None

Conflict of interest

None

Financial support

This work was supported by National Natural Science Foundation of China (Grant Nos. U22B20112 and 52476001), and Beijing Institute of Technology Research Fund Program for Young Scholars.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Ethics statement

None

Received: 28 Mar 2022 Revised: 30 May 2022 Accepted: 16 Jun 2022
Published online: 15 June 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Dai M, Sun M, Chen B, Shi L, Jin M, Man Y, et al. Country-specific net-zero strategies of the pulp and paper industry. Nature. 2024;626:327-34. https://doi.org/10.1038/s41586-023-06962-0.

2. Aziz MJ, Gayme DF, Johnson K, Knox-Hayes J, Li P, Loth E, et al. A co-design framework for wind energy integrated with storage. Joule. 2022;6:1995-2015. https://doi.org/10.1016/j.joule.2022.08.014.

3. Meng K, Yiqing Z, Xin S, Gangcheng W, Fang F. Research on two-stage optimization approach of Community integrated energy system considering load supply reliability. Power Gener Technol. 2023;44:25-35.

4. Yang S, Gao HO, You F. Integrated optimization in operations control and systems design for carbon emission reduction in building electrification with distributed energy resources. Adv Appl Energy. 2023;12:100144. https://doi.org/10.1016/j.adapen.2023.100144.

5. Li J, Peng X, Yang Z, Hu S, Duan Y. Design, improvements and applications of dual-pressure evaporation organic rankine cycles: a review. Appl Energy. 2022;311:118609. https://doi.org/10.1016/j.apenergy.2022.118609.

6. Hao J, Yang Y, Xu C, Du X. A comprehensive review of planning, modeling, optimization, and control of distributed energy systems. Carbon Neutrality. 2022;1:28. https://doi.org/10.1007/s43979-022-00029-1.

7. Zhang Y, Zhou Z, Du Y, Shen J, Li Z, Yuan J. A data transfer method based on one dimensional convolutional neural network for cross-building load prediction. Energy. 2023;277:127645. https://doi.org/10.1016/j.energy.2023.127645.

8. Loutzenhiser PG, Maxwell GM, Manz H. An empirical validation of the daylighting algorithms and associated interactions in building energy simulation programs using various shading devices and windows. Energy. 2007;32:1855-70. https://doi.org/10.1016/j.energy.2007.02.005.

9. Zhu D, Hong T, Yan D, Wang C. A detailed loads comparison of three building energy modeling programs: energyPlus, DeST and DOE-2.1E. Build Simul. 2013;6:323-35. https://doi.org/10.1007/s12273-013-0126-7.

10. Zhang Y, Zhou Z, Liu J, Yuan J. Data augmentation for improving heating load prediction of heating substation based on TimeGAN. Energy. 2022;260:124919. https://doi.org/10.1016/j.energy.2022.124919.

11. Kamel E, Sheikh S, Huang X. Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. Energy. 2020;206:118045. https://doi.org/10.1016/j.energy.2020.118045.

12. Zhang G, Ge Y, Pan X, Sadat Afsharzadeh M, Ghalandari M. Optimization of energy consumption of a green building using PSO-SVM algorithm. Sustain Energy Technol Assessments. 2022;53:102667. https://doi.org/10.1016/j.seta.2022.102667.

13. Tso GKF, Yau KKW. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy. 2007;32:1761-8. https://doi.org/10.1016/j.energy.2006.11.010.

14. Ahmad MW, Mourshed M, Rezgui Y. Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 2017;147:77-89. https://doi.org/10.1016/j.enbuild.2017.04.038.

15. Neto AH, Fiorelli FAS. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build. 2008;40:2169-76. https://doi.org/10.1016/j.enbuild.2008.06.013.

16. Salehi S, Kavgic M, Bonakdari H, Begnoche L. Comparative study of univariate and multivariate strategy for short-term forecasting of heat demand density: exploring single and hybrid deep learning models. Energy AI. 2024;16:100343. https://doi.org/10.1016/j.egyai.2024.100343.

17. Fantini DG, Silva RN, Siqueira MBB, Pinto MSS, Guimarães M, Brasil ACP. Wind speed short-term prediction using recurrent neural network GRU model and stationary wavelet transform GRU hybrid model. Energy Convers Manag. 2024;308:118333. https://doi.org/10.1016/j.enconman.2024.118333.

18. Young TL, Gopsill J, Valero M, Eikevåg S, Hicks B. Comparing four machine learning algorithms for household non-intrusive load monitoring. Energy AI. 2024;17:100384. https://doi.org/10.1016/j.egyai.2024.100384.

19. Ciechulski T, Osowski S. High precision lstm model for short-time load forecasting in power systems. Energies. 2021;14:1-15. https://doi.org/10.3390/en14112983.

20. Chalapathy R, Khoa NLD, Sethuvenkatraman S. Comparing multi-step ahead building cooling load prediction using shallow machine learning and deep learning models. Sustain Energy, Grids Networks. 2021;28:100543. https://doi.org/10.1016/j.segan.2021.100543.

21. Pei J, Liu N, Shi J, Ding Y. Tackling the duck curve in renewable power system: a multi-task learning model with iTransformer for net-load forecasting. Energy Convers Manag. 2025;326:119442. https://doi.org/10.1016/j.enconman.2024.119442.

22. Wang L, Tan M, Chen J, Liao C. Multi-task learning based multi-energy load prediction in integrated energy system. Appl Intell. 2023;53:10273-89. https://doi.org/10.1007/s10489-022-04054-6.

23. Li C, Li G, Wang K, Han B. A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient integrated energy systems. Energy. 2022;259:124967. https://doi.org/10.1016/j.energy.2022.124967.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

24. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. 35th AAAI Conf Artif Intell AAAI 2021. 2021;12B:11106-15. https://doi.org/10.1609/aaai.v35i12.17325.

25. Gong M, Zhao Y, Sun J, Han C, Sun G, Yan B. Load forecasting of district heating system based on Informer. Energy. 2022;253:124179. https://doi.org/10.1016/j.energy.2022.124179.

26. Tan M, Liao C, Chen J, Cao Y, Wang R, Su Y. A multi-task learning method for multi-energy load forecasting based on synthesis correlation analysis and load participation factor. Appl Energy. 2023;343:121177. https://doi.org/10.1016/j.apenergy.2023.121177.

27. Zhao H, Wu Y. Multienergy load forecasting model for integrated energy systems based on coupling auxiliary sequences and multitask learning. IET Gener Transm Distrib. 2024;18:1757-69. https://doi.org/10.1049/gtd2.13023.

28. Guo Y, Li Y, Qiao X, Zhang Z. BiLSTM multitask learning-based combined load forecasting considering the loads coupling relationship for multienergy system. IEEE Trans Smart Grid. 2022;13:3481-92. https://doi.org/10.1109/TSG.2022.3173964.

29. Li K, Mu Y, Yang F, Wang H, Yan Y, Zhang C. Joint forecasting of source-load-price for integrated energy system based on multi-task learning and hybrid attention mechanism. Appl Energy. 2024;360:122821. https://doi.org/10.1016/j.apenergy.2024.122821.

30. Hu R, Zhou K, Lu X. Integrated loads forecasting with absence of crucial factors. Energy. 2025;322:135632. https://doi.org/10.1016/j.energy.2025.135632.

31. Liu X, Yang Z, Guo Y, Li Z, Xu X. A novel correlation feature self-assigned Kolmogorov-Arnold Networks for multi-energy load forecasting in integrated energy systems. Energy Convers Manag. 2025;325:119388. https://doi.org/10.1016/j.enconman.2024.119388.

32. Niu D, Yu M, Sun L, Gao T, Wang K. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. Appl Energy. 2022;313:118801. https://doi.org/10.1016/j.apenergy.2022.118801.

33. Jin N, Yang F, Mo Y, Zeng Y, Zhou X, Yan K, et al. Highly accurate energy consumption forecasting model based on parallel LSTM neural networks. Adv Eng Informatics. 2022;51:101442. https://doi.org/10.1016/j.aei.2021.101442.

34. Fang L, He B. A deep learning framework using multi-feature fusion recurrent neural networks for energy consumption forecasting. Appl Energy. 2023;348:121563. https://doi.org/10.1016/j.apenergy.2023.121563.

35. Fan C, Xiao F, Zhao Y. A short-term building cooling load prediction method using deep learning algorithms. Appl Energy. 2017;195:222-33. https://doi.org/10.1016/j.apenergy.2017.03.064.

36. Zhao A, Mi L, Xue X, Xi J, Jiao Y. Heating load prediction of residential district using hybrid model based on CNN. Energy Build. 2022;266:112122. https://doi.org/10.1016/j.enbuild.2022.112122.

37. Zhang W, Cai Y, Zhan H, Yang M. Multi-energy load forecasting for small-sample integrated energy systems based on neural network Gaussian process and multi-task learning. Energy Convers Manag. 2024;321:119027. https://doi.org/10.1016/j.enconman.2024.119027.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

38. Kahwash F, Barakat B, Maheri A. Coupled thermo-electrical dispatch strategy with AI forecasting for optimal sizing of grid-connected hybrid renewable energy systems. Energy Convers Manag. 2023;293:117460. https://doi.org/10.1016/j.enconman.2023.117460.

39. ASU. Campus metabolism. http://cm.asu.edu/2020.

40. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;2017-Decem:5999-6009.

41. Liu J, Zang H, Cheng L, Ding T, Wei Z, Sun G. A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting. Appl Energy. 2023;342:121160. https://doi.org/10.1016/j.apenergy.2023.121160.

42. Ma J, Zhao Z, Yi X, Chen J, Hong L, Chi EH. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2018;1930-9. https://doi.org/10.1145/3219819.3220007.

43. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016;2016-Decem:770-8. https://doi.org/10.1109/CVPR.2016.90.

44. NCEI. https://www.ncei.noaa.gov/2020.

45. Niako N, Melgarejo JD, Maestre GE, Vatcheva KP. Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with ARIMA and LSTM. BMC Med Res Methodol. 2024;24:320. https://doi.org/10.1186/s12874-024-02448-3.

46. Huang N, Wang X, Wang H, Wang H. Ultra-short-term multi-energy load forecasting for integrated energy systems based on multi-dimensional coupling characteristic mining and multi-task learning. Front Energy Res. 2024;12:1-16. https://doi.org/10.3389/fenrg.2024.1373345.

47. Antonesi G, Cioara T, Anghel I, Papias I, Michalakopoulos V, Sarmas E, et al. Hybrid transformer model with liquid neural networks and learnable encodings for buildings' energy forecasting. Energy AI. 2025;20:100489. https://doi.org/10.1016/j.egyai.2025.100489.

48. Chen W, Rong F, Lin C. A multi-energy loads forecasting model based on dual attention mechanism and multi-scale hierarchical residual network with gated recurrent unit. Energy. 2025;320:134975. https://doi.org/10.1016/j.energy.2025.134975.

49. Li K, Duan P, Cao X, Cheng Y, Zhao B, Xue Q, et al. A multi-energy load forecasting method based on complementary ensemble empirical model decomposition and composite evaluation factor reconstruction. Appl Energy. 2024;365:123283. https://doi.org/10.1016/j.apenergy.2024.123283.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press