Energy 305 (2024) 132344

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A reinforcement learning-based online learning strategy for real-time short-term load forecasting

Xinlin Wang [a], Hao Wang [b,c,*], Shengping Li [d], Haizhen Jin [e]

[a] CSIRO Energy, Newcastle, Australia
[b] Department of Data Science and AI, Faculty of IT, Monash University, Melbourne, Australia
[c] Monash Energy Institute, Monash University, Melbourne, Australia
[d] School of Design and the Built Environment, Curtin University, Perth, Australia
[e] Industrial Systems Engineering and Management, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Short-term load forecasting
Reinforcement learning
Online learning
Quantile regression
Open-source

ABSTRACT

Real-time Short-Term Load Forecasting (STLF) is crucial for energy management and power system operations. Conventional Machine Learning (ML) methodologies for STLF are often challenged by the inherent variability in energy demand. To tackle the challenge associated with inherent variability, this paper presents a novel Reinforcement Learning (RL)-enhanced STLF method. Different from conventional methods, our method dynamically improves the STLF model by selecting optimal training data to capture recent power usage trends and possible variations in demand patterns. By doing so, our method can significantly reduce the impact of unforeseen fluctuations in real-time forecasting. In addition to the novel RL-enhanced STLF method, we propose a comprehensive evaluation framework, encompassing three key dimensions: accuracy, runtime efficiency, and robustness. Tested on three distinct real-world energy datasets, our RL-enhanced method demonstrates superior forecasting performance across three evaluation metrics by achieving accurate and robust predictions in real-time under varying scenarios. Furthermore, our approach provides uncertainty bounds for practical predictions applications. These results underscore the significant advancements made by our RL-based method in forecasting precision, efficiency, and robustness. We have made our algorithm openly accessible online to promote continued development and advancement of STLF methods.

1. Introduction

Short-Term Load Forecasting (STLF) serves as an essential tool for predicting energy demand over short horizons, which can range from minutes to several hours, depending on the resolution of data and specific operational requirements [1–3]. With the growing mix of energy demands from different sectors and wide electrification, STLF plays an increasingly important role in supporting power system operations, energy management, and electricity pricing [3–6]. In particular, microgrid and virtual power plant (VPP) operators can leverage STLF to develop real-time operational strategies for managing their energy production and consumption scheduling [7]. Accurate STLF plays a pivotal role in informing operational decisions that enhance system efficiency, reduce costs, and lower carbon emissions [5,8]. Furthermore, utility companies depend on the precision of STLF for making critical daily operational decisions, such as managing electricity supply, implementing prosumer power curtailment, and overseeing asset management [3,4,9]. STLF is essential for these stakeholders to maintain a balance between energy supply and demand, ensuring grid reliability and stability.

Although daily load patterns may possess a certain degree of regularity, the demand exhibits notable fluctuations and intricate details in both short-term and long-term [6,8], making STLF a challenging task. Due to the rapid advancements in machine learning (ML) technology, many ML-based strategies have been applied to enhance the effectiveness of STLF. For example, Rubasinghe et al. in [10] recently introduced an enhanced Empirical Mode Decomposition (EMD) method, called Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN). It aims to provide a deeper understanding of the electrical load, through identifying different energy components embedded inside the load data, so as to enhance the forecasting performance. Similarly, Yi et al. in [11] proposed a comprehensive prediction framework using a convolutional neural network (CNN) and a long short-term memory network (LSTM) for feature extraction and learning. Additionally, a self-attention mechanism is incorporated to

* Corresponding author at: Department of Data Science and AI, Faculty of IT, Monash University, Melbourne, Australia.
E-mail addresses: xinlin.wang@csiro.au (X. Wang), hao.wang2@monash.edu (H. Wang), shengping.li@curtin.edu.au (S. Li), jinhz@nus.edu.sg (H. Jin).

https://doi.org/10.1016/j.energy.2024.132344
Received 10 December 2023; Received in revised form 28 April 2024; Accepted 6 July 2024
Available online 9 July 2024

further enhance the learning results. The above representative studies showcase the primary approaches adopted by state-of-the-art ML-based STLF methods, which encompass tailored feature extraction techniques and the deployment of various advanced ML models. For a more comprehensive literature review of ML-based STLF methodologies, please refer to Section 2.

While the aforementioned advancements present a promising direction to enhance STLF performance, certain challenges endure when applying these ML methodologies in practical real-world scenarios. A central challenge arises from the fluctuating nature of demand, which can exhibit non-stationary behavior, particularly for medium and small consumers like precincts and buildings, in contrast to bulk power systems. These variations are primarily driven by consumer behaviors, which are notoriously challenging to predict, as consumer behaviors are influenced by factors such as weather conditions, lifestyle choices, and the diverse range of electrical appliances in use [6]. These resulting intricacies directly contribute to the complexities observed in demand data [8]. Nonetheless, the efforts made by the state-of-the-art ML-based STLF methods inadvertently lead to larger and more complex learning frameworks, which, in turn, introduce new challenges related to the need for diverse feature data, substantial computational resource needs, elongated training duration, and intricately designed system frameworks [2,7,12]. Firstly, obtaining access to rich and diverse feature data, such as temperature, humidity, and cloud cover as weather data along with other consumer-related features, can be challenging [5,7,12]. It is evident that the predictive capabilities of ML models can be improved by integrating an extensive array of input features and utilizing customized and intrinsic methods for feature extraction approaches. However, obtaining such multifaceted data is often impractical due to practical limitations [8,12]. Secondly, there is a growing dependence on sophisticated hardware, necessitated by increased computational demands from vast data volumes and complex learning structures, which may limit the real-world applicability of these methodologies [7,8]. Thus, instead of increasing reliance on these cutting-edge ML technologies, it is imperative to propose effective and feasible solutions that are tailored to tackle the stochastic nature of energy consumption data.

To address these challenges in STLF and devise a practical strategy, this study introduces an innovative online training data selection strategy leveraging reinforcement learning (RL). The motivation for employing RL stems from its capability to dynamically interact with the environment, optimize decisions across multiple criteria simultaneously, and its model-free nature [13,14]. Considering the intrinsic nature of demand data, the dynamic training data selection criteria are proposed to guide RL to optimize its decision-making process. The criteria not only measure forecasting accuracy but also analyze the reasons for any inaccuracies. Informed by these evaluations, RL dynamically recalibrates its training data selection strategies to better fit the current power usage situations, thus improving forecasting accuracy. For example, when RL identifies that the current incorrect forecasts are caused by unanticipated variations in load, it autonomously adjusts the selection of training data, accommodating more various power usage patterns to mitigate the impact of fluctuations.

Following this, it is crucial to highlight that the proposed forecasting method employs a suite of quantile regression models to learn from the updated training data, rendering a dynamic prediction interval at each step. Unlike traditional regression techniques that focus primarily on the mean or central tendency, quantile regression provides an approach to analyze the entire distribution of the forecasted data. This is particularly beneficial when dealing with demand forecasts where sudden fluctuations can dramatically influence prediction outcomes [15]. Consequently, these intervals intrinsically reflect the automatic adjustments made by RL by varying their sizes and values to capture uncertainties in the forecast.

By providing the demand data feature-tailored dynamic data selection criteria and integrating them with RL, the proposed method is capable of autonomously analyzing the step-by-step forecasting results and optimizing its actions for accurate forecasting. The proposed method exclusively relies on historical meter data as its sole learning reference to deliver accurate predictions. The widespread availability of its target dataset, coupled with its model-free RL-based framework, creates a straightforward forecasting method, promising to broaden horizons in practical applications. To comprehensively discuss the performances of the proposed method, we adopt an innovative evaluation paradigm that covers three dimensions: **Accuracy**, **Runtime efficiency**, and **Robustness**. We especially emphasize that the proposed robustness assessment is motivated by the intricate and multifarious nature of energy demand. We point out that a practical STLF method should consistently maintain accuracy and timeliness across varying target datasets, exemplifying its robustness against different forecasting scenarios. Consequently, we have selected three extensive real-world datasets. The first two datasets consist of years of electrical load data from two campuses in the United States, the University of California, Irvine (UCI) [8] and the University of California, San Diego (UCSD) [16]. The UCSD dataset is publicly available, but the UCI dataset it not. The third dataset, obtained from Open Power System Data (OPSD) and also publicly available, presents energy load from power plants across Europe, encompassing the aggregated energy demands of large infrastructures and regional grids [17]. This diverse selection of datasets enables our study to test representative load types with varying complexity and potential applications across both small and large-scale systems. Our comprehensive evaluations across multiple datasets confirm that our method consistently surpasses benchmarks in forecasting accuracy and runtime efficiency, thus demonstrating its robustness. Not only does it deliver superior prediction results when compared to various ML/DL-based benchmarks, but it also requires fewer training data, ensuring near real-time forecasts. Most notably, irrespective of the fluctuating complexities inherent to target datasets, our approach consistently yields accurate results. This demonstrates its clear advantage over contemporary methods and underscores its significant potential in diverse real-world scenarios. We make the proposed STLF method available in open source to promote further research in the field [18].

The remainder of this paper is organized as follows. Section 2 discusses the related works. Section 3 introduces the proposed method. Section 4 describes our experimental methodology. Section 5 exhibits our experimental results. Section 6 summarizes our findings and contributions.

2. Related work

Recent innovations have led to many new methods in STLF, primarily focusing on customized feature extraction techniques, cutting-edge ML models, and well-designed multi-model-based forecasting frameworks. In this section, we focus solely on related works published from 2020 up to the date of this study.

The feature extraction strategies are proposed to capture or extract important characteristics from the data, enabling ML-based prediction methods to learn more effectively [2,10,19]. Among them, time series decomposition stands out as a dominant technique for feature extraction in time series data mining. In 2023, Rubasinghe et al. refined the original EMD and provided the improved EMD, called ICEEMDAN [10]. This innovation was found to enhance the efficacy of ML/DL models, as evidenced by performance comparisons between algorithms like LSTM and Back Propagation using both the ICEEMDAN and traditional EMD. Concurrently, Yue et al. proposed the Ensemble Empirical Mode Decomposition (EEMD) to boost the precision of LSTM-based forecasting models [20]. Their study also explored permutation entropy, feature selection, and Bayesian optimization to gauge their combined effect on accuracy. These efforts resonated with Rubasinghe et al. demonstrating marked improvements in LSTM's performance when integrated with EMD and EEMD. Beyond EMD, Variational Mode Decomposition

(VMD) has also gained attention in STLF [2,19]. In the work of Yang et al. the integration of VMD with learning strategies was explored using XGBoost, SecureBoost, and a federated k-means clustering algorithm [2]. Through evaluations with various datasets, the efficacy and advantages of incorporating VMD were demonstrated. In parallel, Wang et al. introduced a Bald Eagle Search-optimized VMD, aiming to further boost the learning outcomes of ML/DL techniques [19]. By comparing the performance of several DL models with and without the integration of the optimized VMD, the study underlines the importance of incorporating VMD into the model architecture. These findings collectively emphasized the significant role of advanced feature extraction techniques in heightening predictive accuracy.

Parallel to the evolution in feature extraction, recent efforts have been made to pioneer learning techniques to refine forecasting outcomes. For instance, Jiang et al. harnessed the Deep-Autoformer Neural Network, a novel deep learning (DL) model, to foster superior STLF [1]. Building on the foundation of the original Autoformer, their study incorporated additional Multi-Layer Perceptron layers, thus enhancing the deep information extraction capabilities of Autoformer. Empirical results from testing data demonstrate that the novel DeepAutoformer not only achieves high-precision STLF but also effectively mitigates overfitting issues. Additionally, RL, a dynamic approach in the ML paradigm, has also found its application in STLF. Both Park et al. and Feng et al. have recently developed advanced forecasting strategies leveraging RL [21,22]. Feng et al. leveraged RL to establish a multiple ML model results selection framework where various ML/DL models are simultaneously trained for forecasting [21]. Then, an RL agent is designed to dynamically select the most potentially accurate forecast output from these models. On the other hand, Park et al. utilized RL to dynamically select the most similar training data from the entire dataset, thus improving the STLF accuracy [22]. Nonetheless, in this study, we point out that the online learning approach, which consistently selects the most similar historical data as the training dataset, may struggle with unexpected variations in electrical load patterns that stem from the inherent unpredictability and variability of real-world datasets [7,22]. For practical and accurate STLF, the online training method should be comprehensive enough to mitigate the impacts of unexpected changes in demand.

In addition to standalone models, frameworks that integrate multiple distinct forecasting models are garnering interest. For instance, Jia et al. devised a multifaceted prediction structure involving Convolutional Neural Network (CNN) and LSTM for STLF [23]. Similarly, Javed et al. presented a novel hybrid prediction system for STLF, using a unique combination of Short Receptive Field-based Dilated Causal Convolutional network and improved LSTM [24]. An extensive evaluation by Zulfiqar et al. introduced the platform integrating multiple ML/DL algorithms to enhance prediction [25]. Such hybrid frameworks have consistently demonstrated improved forecasting accuracy, highlighting the evolving landscape of predictive modeling techniques.

In sum, while existing literature offers a myriad of advancements in STLF methodologies, ranging from feature extraction to multi-model-based learning frameworks, challenges still persist, especially with regard to the unpredictability of energy demands and resource constraints. Our work builds on these foundations, particularly emphasizing the adaptability of RL. By integrating RL in a novel manner, we aim to address the gaps prevalent in current forecasting methodologies and contribute to a more adaptive, real-time solution to the dynamic landscape of STLF.

3. The proposed methodology

The proposed work introduces an RL-based online training strategy that proactively evaluates the meter data and the previous forecasts. Based on the assessment, the proposed method dynamically adjusts its training data selection strategies, mitigating the impact of unpredictable fluctuations in demand data. At the core of our methodology is a training data selection criterion, which is specifically centered on power consumption data features. The RL-based training data selector refines its decision-making process by continually adapting to these multifaceted criteria. We begin by outlining these criteria. Subsequently, we describe the data preparation, exhibiting the data restructuring process according to our proposed criteria. We then detail our RL-driven forecasting approach. Fig. 1 provides an overview of our proposed framework, which is composed of two primary components.

1. This RL-based training data selector dynamically chooses an optimal training dataset. Let t denote the wall-clock time when the data is captured. Accordingly, the obtained output is $\mathcal{D}(t)$.
2. Real-time load predictor: Once the training dataset $\mathcal{D}(t)$ is determined, the predictor is trained on it. We design the predictor to encompass two parts. The first is a set of quantile regression-based load prediction models. Then, to further enhance the prediction accuracy, a prediction result adjustment module is introduced.

Considering our framework employs a set of quantile regression models to precede, the predicted interval, denoted as $R(t+1)$, represents the possible range for the electrical load at $t+1$, ensuring a probabilistic coverage of the true load value. To further enhance accuracy, a prediction result adjustment mechanism is introduced. This approach uses recent forecasting outcomes to predict a potential forecasting error, $e(t+1)$, which then adjusts the original outputs from the quantile regression-based models. Consequently, the refined output is the amended prediction interval, $R_{adjusted}(t+1)$, where $R_{adjusted}(t+1) = R(t+1) + e(t+1)$. We note that to enhance forecast accuracy, our STLF strategy's prediction horizon is set to be one step ahead, according to different time resolutions of the test dataset.

3.1. Online training data selection criteria

Given the inherent diversity and complexity of the power load, the optimal training data set should consistently strike a fine balance. That is, while representing the most recent user activities from the historical data, the training data should also encompass a broad array of potential usage patterns to mitigate fluctuations. At the same time, considering computational cost, the size of the training data should be appropriate. Therefore, it is a trade-off between representing the most similar patterns to the current power usage activities and selecting potential usage patterns to cushion the variations.

Building on these insights, this work provides four pivotal training data selection criteria to generate the most suitable training dataset $\mathcal{D}(t)$ at each step.

1. The training data should accurately reflect the current power usage characteristics and patterns, ensuring a representation of both periodic and prevalent energy consumption patterns aligned with the current situation.
2. Training data should incorporate a range of diverse scenarios. While some of these scenarios may not have a strong correlation with the current energy consumption patterns, they serve to anticipate and address unexpected variances in power usage, thereby enhancing the reliability of predictions.
3. The training data should be appropriately sized to ensure efficient training.
4. Since we rely on quantile regression models, training samples should have a suitable distribution to maintain a practical prediction interval size.

Thus, in accordance with the proposed criteria, the 3.2 is introduced to represent the historical meter data for observation and selection by RL.

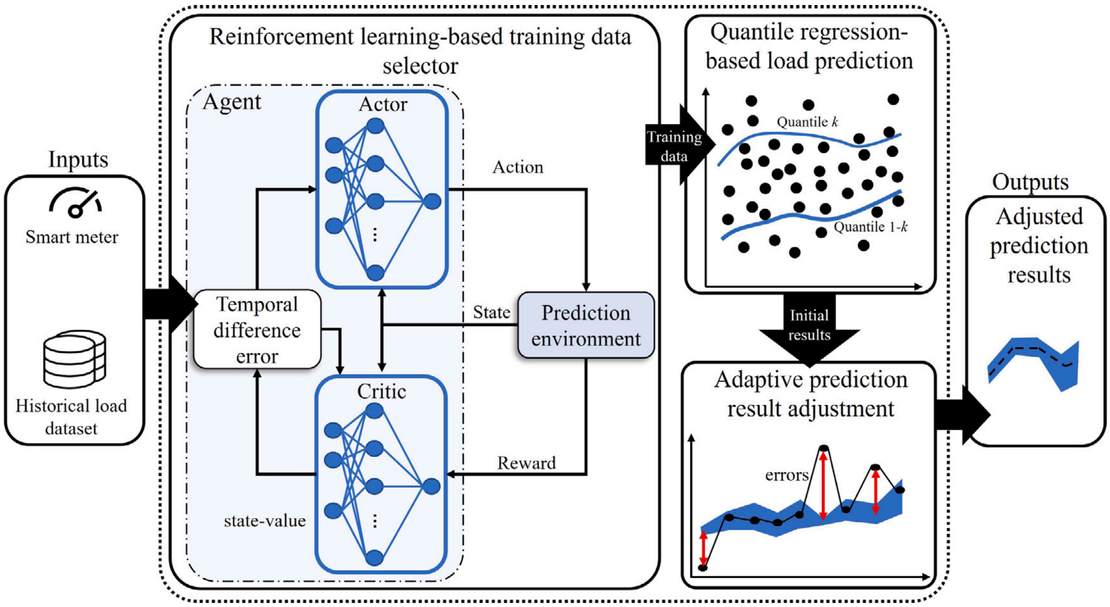

Fig. 1. Flow chart of the proposed unified framework.

3.2. Data preparation

The data preparation phase represents the initial step in our framework, tasked with automatically generating a well-structured collection of candidate training data samples for RL to select. To balance the reflection of daily activity patterns with the efficiency of training, we set the length of observation data at 24 h. That can effectively capture the daily periodicity inherent in human activities. Namely, for a given data capture instant t, the observation data used for prediction spans from $L(t-24\text{ h})$ to $L(t)$. Here, L represents the meter data in the database. Then, considering the influence of varied weather patterns across the four seasons on power consumption, our framework is designed to access an entire year's historical dataset as its training data selection pool. This means that while the immediate observation window is 24 h for capturing daily patterns, our model can leverage data spanning from $L(t-1year)$ to $L(t)$ to observe the impacts of seasonal factors on demand and select the optimal training data.

Building on this, our framework generates power load series for each 24-h segment preceding time t throughout the dataset's weekly representation. For example, when the current moment t is noon on a Thursday. Our approach selectively extracts load data from the exact 24-h blocks, mirroring the present time span, for each week within the dataset. In other words, the proposed method captures data from every 'Thursday noon to Friday noon' segment across the entire year. Therefore, it results in a collection of 52 sub-sequences, each offering a weekly snapshot of this specified 24-h window.

3.2.1. Dynamic time warping-based training data restructure

To effectively manage the obtained 52 distinct 24-h load series in an easier way for RL to select, our method then measures the similarity of each sub-sequence relative to the current observation data, $L(t-24\text{ h}),\ldots,L(t)$. We employ Dynamic Time Warping (DTW) as our similarity measure. DTW is a widely recognized technique for gauging similarity between time series [8]. It calculates a distance-like metric between sequences, with smaller values indicating higher similarity [8]. The choice of DTW stems from its remarkable ability to provide a robust similarity measure even in the presence of temporal shifts and distortions that can be commonplace in time series data, especially in the context of electrical load profiles. Such shifts may arise due to factors such as changes in daily human activity patterns or slight variations in equipment operations. DTW can inherently account for these shifts, ensuring that two sequences that are similar but out of phase in their patterns can still be recognized as alike. This characteristic makes DTW particularly suitable for our scenario, where capturing the essence of similarity is crucial.

After computing the similarity scores using DTW, the 52 sub-sequences are ranked based on their resemblance to the current test data, denoted as $Candidates(t)_{52,l}$, where l is the length of each historical 24-h time series. This ranking assists the RL in making an informed selection among these series according to its dynamic power usage situation at each step. Then, the prediction environment is introduced to describe the observation of our method according to the represented $Candidates(t)_{52,l}$.

3.3. Reinforcement learning-based training data selector

In our data selector, we introduce an automated strategy that persistently observes prediction outcomes and meter data, aiming to select the optimal training data for each forecasting step. A critical question then arises: How to select the most suitable training dataset? In Section 2, we highlighted the limitations of traditional online learning methodologies. These approaches, which routinely select the most similar historical or the most recent data as the training dataset, often falter in the face of unexpected variations in electrical load patterns [7,22]. Such fluctuations are a byproduct of non-stationary energy consumption behaviors, which are presented in real-world datasets. Distinguishing our work from these methods, we introduce our RL-based training data selection mechanism.

3.3.1. Implementation of reinforcement learning

The proposed RL-based forecasting method can be formulated as a Markov Decision Process (MDP), encompassing four main components [21,22]:

$$M = \langle S(t), A(t), R(t), ForecastingModels \rangle \quad (1)$$

where:

- $S(t)$: State space. This set of states encompasses the step-by-step forecasts and meter data available for the agent to observe.
- $A(t)$: Action space. This set represents the decisions available to the agent. In this work, it denotes the selection of different training datasets.

- $R(t)$: Reward function. This function provides a value for the reward corresponding to a particular state–action pair.
- $ForecastingModels$: These models determine the transition from the current state $S(t)$ to the subsequent state $S(t+1)$ based on a chosen action $A(t)$. Upon selecting an action (i.e., a specific training dataset), the forecasting models evaluate this input to generate the next prediction results. These predictions, combined with real-time meter data, define the forthcoming state. Further details regarding the models are elaborated in Section 3.4.

3.3.2. Design of state space: $S(t)$

To align with the training data selection criteria introduced in Section 3.1, our state space is crafted to equip the agent with a holistic view of the previous forecast. This view includes three perspectives:

1. Gauging the accuracy of the most recent forecast,
2. Identifying if inaccuracies stem from significant load fluctuations,
3. Evaluating the suitability of the current prediction interval size.

To realize this, we introduce three evaluative metrics to compose the state space, i.e., $S(t) = \{A(t), F(t), I(t)\}$. Each is explained below.

- Accuracy, $A(t)$: Indicates the forecasting results. $A(t) = 1$ if the prediction is accurate, i.e., the real meter data falls into the prediction interval ($L(t) \in R(t)$), else $A(t) = 0$.
- Fluctuation identification, $F(t)$: Determine if the inaccuracy is attributed to significant load fluctuations. We provide a straightforward approach to analyzing load variations. By calculating the load differences at each step and leveraging quartile-based boundaries, we efficiently identify the fluctuations in load data. Specifically, $F(t) = 1$ when $change(t) \in [Q1(t), Q3(t)]$. Here, $change(t) = L(t) - L(t-1)$, indicating the current load changes. $Q1(t)$ and $Q3(t)$ are the first and third quartiles of the changes in the $Candidates(t)$ at consistent weekly times ($Candidates(t)_{52,l}[:,-1] - Candidates(t)_{52,l}[:,-2]$). Otherwise, $F(t) = 0$.
- Interval size, $I(t)$: Assesses the appropriateness of the prediction interval size. The appropriateness of the prediction interval size mainly depends on the magnitude of the corresponding load value. When observing a big value of the load $L(t)$, a relatively expansive prediction interval can be acceptable. Conversely, for a smaller load value, the prediction interval should be narrower in scope.
Consequently, $I(t) = 1, if(\frac{R(t)}{L(t)} < \alpha)$. $I(t)$ is set to 1 if the ratio of the interval size, $R(t)$, to the current load, $L(t)$, is smaller than a preset threshold, α, implying that the interval size is deemed suitable. Otherwise, $I(t) = 0$. α is a preset threshold, indicating the interval size is considered reasonable. The determination of α is based a data-driven investigation detailed in Section 5.

It is crucial to emphasize that, given the exclusive reliance on historical data as our sole observations, we tailor the proposed States(t) as a simplified representation schema. This is rooted in empirical experiments, and the superiority of our approach will be demonstrated in Section 5.

3.3.3. Design of action space: $A(t)$

Align with the proposed traits selection criteria, the action space in our study is meticulously designed to encompass a diverse set of training sample combinations selected from $Candidates(t)_{52,l}$, i.e., $A(t) = \{A_0(t), A_1(t), \dots, A_n(t)\}$.

- $A_0(t)$: Selects the top m samples from $Candidates(t)_{52,l}$, gravitating towards the most similar historical instances.
- $A_1(t)$: Begins by selecting the first sample, thereafter picking every alternate one until a total of m samples are chosen.
- ...
- $A_n(t)$: Initiates with the first sample but subsequently selects every $(n+1)$th instance, accumulating up to m samples.

Each action represents a unique trade-off among the selection criteria detailed in Section 3.1. For illustration, $A_0(t)$, aims at selecting the m days that most closely mirror the current scenario, intending to formulate a precise and constrained prediction interval by leveraging patterns from historical data. While such an approach prioritizes specificity, it may falter in the face of unforeseen data anomalies. In contrast, $A_n(t)$, with its diversified sample selection, is more receptive to unexpected variations, potentially leading to wider prediction intervals. The relationship between these actions is an evolving spectrum of data specificity, starting from a narrow, history-driven focus and gradually transitioning to a broader, variance-acknowledging perspective.

Consequently, the selected action from our action space $A(t)$ serves as the final output of our training data selector, denoted as $\mathcal{D}(t)$:

$$\mathcal{D}(t) = \begin{bmatrix} d_{11}(t) & d_{12}(t) & \dots & d_{1l}(t) \\ d_{21}(t) & d_{22}(t) & \dots & d_{2l}(t) \\ . & . & \dots & . \\ d_{m1}(t) & d_{m2}(t) & \dots & d_{ml}(t) \end{bmatrix}, \tag{2}$$

where m is the number of time series in $\mathcal{D}(t)$, which is determined empirically and introduced in Section 5.

3.3.4. Reward function: $R(t)$

The essence of RL is the reward mechanism, which provides feedback to the agent based on its actions. For our purpose, the reward, denoted as $R(t)$, is computed depending on which category the outcome falls into, based on the values of $A(t)$, $F(t)$, and $I(t)$.

1. Rewarding Cases: These occur when the prediction is accurate ($A(t) = 1$), indicating a strong alignment between the forecasted and actual loads.
   - $C1$: The best scenario where prediction is accurate, and the interval size is appropriate.
     - Condition: $A(t) = 1$ and $I(t) = 1$
     - Reward: $R(t) = \beta \times \frac{L(t)}{|mid(t) - L(t)|}$
   - $C2$: Accurate prediction, but the interval size could be improved.
     - Condition: $A(t) = 1$ and $I(t) = 0$
     - Reward: $R(t) = \frac{L(t)}{|mid(t) - L(t)|}$
2. Evitable Failures: The prediction is inaccurate due to preventable reasons.
   - C3: No significant change happened, but the prediction interval still fails to capture the load.
     - Condition: $A(t) = 0$ and $F(t) = 0$.
     - Reward $R(t) = -\gamma \times \frac{|mid(t) - L(t)|}{L(t)}$
   - C4: A significant change is observed in the load, and the prediction interval is not wide enough to accommodate these fluctuations.
     - Condition: $A(t) = 0$, $F(t) = 1$, and $I(t) = 1$.
     - Reward: $R(t) = -\gamma \times \frac{|mid(t) - L(t)|}{L(t)}$
3. **Inevitable Failure:**
   - $C5$: Prediction is incorrect despite a broad prediction interval due to unexpected fluctuations.
     - Condition: $A(t) = 0$, $F(t) = 1$, and $I(t) = 0$
     - Reward: $R(t) = 0$

Here, $mid(t) = \frac{\hat{y}_l(t)+\hat{y}_h(t)}{2}$ represents the mid-point of the predicted range. Parameters β and γ are determined via empirical analysis, which we elaborate on in Section 5.

3.3.5. Algorithm

In RL, the primary objective is to optimize decision-making by estimating long-term rewards of potential actions [21,22]. Our work specifically addresses the challenge of adaptively selecting optimal training data, with consideration for the variability of prediction states and associated expected rewards. To address this, we employ the Actor–Critic method.

The Actor–Critic algorithm is recognized as a key approach within RL, melding the strengths of both policy-based and value-based methods [14]. Traditional value-based methods rely on maintaining a value function from which actions are subsequently derived. This indirect methodology can introduce computational inefficiencies and potentially delay decision-making processes [14,21,22]. In contrast, while policy-based agents provide direct action recommendations based on states, they may exhibit issues with stability, especially in complex or noisy reward scenarios [14].

The Actor–Critic method provides a solution to these challenges. By offering a direct state-to-action mapping through its policy function, it ensures rapid decisions [14]. This is essential in applications like the proposed work, where real-time responses are critical. Beyond this efficiency, the methodology's dual nature grants it both flexibility and stability. A high-level illustration of the structure of the Actor–Critic is provided in Fig. 1.

- **Actor**: Guided by the current state, this module recommends an action, shaping the agent's next move.
- **Critic**: This component evaluates the chosen action of the actor based on projected future rewards, acting as an iterative feedback mechanism to fine-tune the actor's decisions.

Within this paradigm, the actor is capable of rapidly converging to optimal policies. In parallel, the critic ensures that these policy adjustments are rooted in accurate anticipations of future rewards. The collaborative mechanism between the actor and critic accentuates the efficacy of the method, rendering it appropriate to address the challenges presented in our work.

3.4. Real-time load predictor

Given the collected load $L(t)$ from the smart meter and the selected training set $\mathcal{D}(t)$, our real-time load predictor is designed to forecast future power consumption. Given that the prediction horizon is set to one step ahead, the initial prediction interval provided by our quantile regression-based load prediction model is $R(t+1) = [\hat{y}_l(t+1), \hat{y}_h(t+1)]$ for the electrical load at the next step $t+1$. Here, $\hat{y}_l(t+1)$ and $\hat{y}_h(t+1)$ represent the lower and upper bounds of the interval, respectively.

As the subsequent step, we employ an adaptive prediction result adjustment approach, shown in Fig. 1. This technique learns from the prediction errors of recent steps to provide an adjustment, denoted as $e(t+1)$. This adjustment serves to further refine the forecasts, enhancing their accuracy and reliability. The final output of our forecasting strategy is $R_{adjusted}(t+1)$, a combination of the prediction result and the adjustment.

3.4.1. Quantile regression-based load prediction

In this work, we devise a quantile regression-based forecasting framework that adaptively mirrors the characteristics of the updated training data, $\mathcal{D}(t)$. This design underscores the dynamic trade-off achieved by our proposed training data selection mechanism, detailed in Section 3.1.

Quantile regression distinguishes itself as a statistical method, particularly for its heightened ability to handle outliers, offering an edge over traditional regression techniques that merely gauge the conditional median [15]. Furthermore, it is especially adept at handling data characterized by sharp peaks or heteroscedasticity, traits frequently manifested in electrical load data. While traditional regression models conduct predictive outcomes by minimizing generic loss functions, such as the mean squared error, quantile regression steers its focus towards the quantile loss function, $l_q(t)$ [8,15]:

$$l_q(t) = \begin{cases} k \cdot (L(t) - \hat{y}_i(t)), & \text{if } L(t) - \hat{y}_i(t) > 0 \\ (k-1) \cdot (L(t) - \hat{y}_i(t)), & \text{if } L(t) - \hat{y}_i(t) < 0 \end{cases} \tag{3}$$

where $\hat{y}(t)$ denotes the prediction at timestep t, $L(t)$ represents the observed load at timestep t, and k is the quantile with its value ranging from 0 to 1. Notably, the adjustment of quantile loss is intricately tied to the quantile k and the value of the prediction error. The loss is positive when the actual value is less than the predicted result and vice-versa [8,15].

In this work, after illustrating the benefits of quantile regression, we incorporated this technique within the XGBoost algorithm for two considerations. First, XGBoost stands out as a potent ensemble learning methodology, adeptly consolidating insights derived from numerous decision trees, culminating in predictions of enhanced accuracy and reliability. Second, this model has received notable recognition and approbation across the spectrum of time-series forecasting, predominantly attributable to its steadfast resilience in managing input uncertainties, its adept capability to curtail overfitting, and its laudable efficiency in learning [8]. This synthesis of quantile regression and XGBoost aims not only to harness the robust predictive prowess of the latter but also to augment it with the predictive interval capabilities provided by the former, thereby envisaging a model that is not only accurate but also adept at navigating through the intricacies and volatilities inherent in energy consumption data.

Consequently, our approach harnesses two XGBoost-augmented quantile regression models, corresponding to the quantiles k and $1-k$, to generate a prediction interval denoted as $R(t+1) = [\hat{y}_l(t+1), \hat{y}_h(t+1)]$. It is pivotal to note that there is a salient distinction between our method and standard probabilistic forecasting. While probabilistic forecasting seeks to produce a complete distribution over all possible future outcomes, our quantile regression approach specifically targets particular quantiles. Nevertheless, by modeling multiple quantiles, our method can mimic the essence of probabilistic forecasting, delineating a spectrum of potential outcomes. This nuanced approach offers precision where needed, while also capturing the overarching variability inherent in the forecast. The selection of k is determined from empirical experiments and will be discussed in Section 5.

3.4.2. Adaptive prediction result adjustment

Forecasting in electrical load prediction is challenging, primarily due to the dynamic and unpredictable nature of the data. Even with state-of-the-art forecasting techniques, there remains a margin of error that can significantly influence the effectiveness of predictions. To further mitigate this limitation, the proposed adaptive prediction result adjuster is designed to dynamically learn from the most recent forecasts, and then anticipate and compensate for the prediction errors.

Our adjustment mechanism utilizes a Support Vector Machine (SVM) model to learn from recent predictions, the meter data, and their associated prediction errors at each step. SVM is renowned for its capability to delineate intricate, non-linear relationships [7]. Furthermore, its memory efficiency makes it an apt choice for learning and predicting errors based on recent forecasts and meter readings. At step t, the training data for the proposed adjuster comprises two distinct data columns. The first encompasses the most recent h previous prediction results, i.e., $R(t), R(t-1), \ldots, R(t-h-1)$. Then, the second encompass the actual load from $t-1$ to $t-h$, i.e., $L(t-1)$, $L(t-2)$..., $L(t-h)$. These two data categories are utilized as inputs for the training process, while the errors from t to $t-h-1$ serve as the training output. The test data is composed of the most recent prediction result, $R(t+1)$, for step $t+1$, and the meter data $L(t)$. Subsequently, the SVM model predicts the error for $t+1$, denoted as $e(t+1)$.

It is imperative to underscore the temporal discrepancy between the prediction results and the actual load in the training data of SVM. Specifically, prediction results inherently lead the actual load by a one-step margin. As a result, the training data contains readings from disparate time steps. The prediction error is defined as the difference between the actual load at each step and the midpoint of the related prediction interval. For example, at step t, $e(t)$ is calculated as $e(t) = L(t) - 0.5 \times (\hat{y}_l(t) + \hat{y}_h(t))$.

The final output of our approach, $R_{adjusted}(t+1)$, is the sum of the prediction result and the predicted error, represented as $R_{adjusted}(t+1) = R(t+1) + e(t+1)$.

4. Experimental methodology

To validate the efficacy of our innovative online learning mechanism and the comprehensive excellence of our forecasting framework, we compare our approach with a widely used online learning training data selection method and five distinguished ML/DL-based STLF strategies.

4.1. Benchmark online learning training data selection mechanism

We use the sliding window (SW) method as our benchmark for online learning. This method is widely used due to its simplicity, efficiency, and ability to provide representative samples for training, thereby ensuring robust forecasts [7]. Consistent with our approach, it starts with the entire year's data as its initial training pool. In every iteration, the SW method selects the 24-h demand data from the same time slot as the current time, repeated for each week throughout the year. Considering that there are 52 weeks in a year, the training data comprises 52 sets of 24-h power consumption load series.

4.2. Benchmark forecasting strategies

The introduced benchmarking strategies represent the most recent efforts in STLF. We employ them because of their proven effectiveness, and strong relevance to this work.

- LSTM: This is a standalone DL model chosen for load forecasting. The LSTM model is particularly adept at processing time series data, a quality that has been validated through its consistent forecasting performance [26].
- Enhanced LSTM Models with Attention Mechanism:
    1. LSTM+Att (LSTM with Attention Mechanism): This model, inspired by [25,27], integrates an attention mechanism with the standard LSTM.
    2. Bi-LSTM+Att (Bidirectional LSTM with Attention Mechanism): This benchmark, inspired by [28,29], not only employs attention but also bidirectional learning for more comprehensive data processing.

  The attention mechanism, inspired by cognitive functionality, enhances the capabilities of sequence-to-sequence learning models. It allows these models to prioritize specific inputs during different training stages, a feature that has been evidenced to boost the performance of both LSTM and Bi-LSTM architectures [25,27–29].

Our proposed methodology is fundamentally built on a suite of quantile regression models. Thus, benchmarking our model's performance against other contemporary and widely adopted quantile regression approaches is essential to validate its relative efficacy.

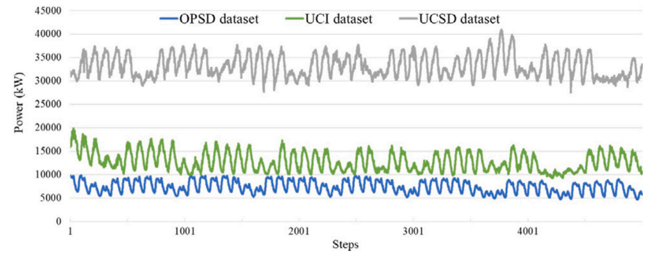

Fig. 2. The test datasets used in this study.

1. Qua+RF (Quantile Regression with Random Forest Forecasting): A model that combines quantile regression with random forest techniques, as described by [30].
2. Qua+Linear (Quantile Regression): This predictor utilizes the sklearn package [31] to implement quantile regression in a linear framework.

Therefore, we introduce five state-of-the-art STLF benchmarking methods: a standalone DL method LTM; two attention mechanism-enhanced LSTM models, LSTM+Att and Bi-LSTM+Att; and two quantile regression-based methods, Qua+RF and Qua+Linear. We comprehensively discuss the performances of each benchmarks compared with our proposed strategy.

4.3. Test datasets

To holistically evaluate the effectiveness of our approach, we employ three distinct real-world datasets, i.e., UCI, OPSD, and UCSD, to test our method and benchmark strategies. The UCI and UCSD datasets not only represent the energy consumption patterns of institutional settings, such as classrooms and laboratories, but also include on-campus residential load, capturing diverse energy dynamics of mixed-use environments [8]. The openness and the comprehensive nature of the UCSD dataset further enhance the numerical analysis. In contrast, OPSD provides energy loads from power plants across Europe. This dataset aggregates load data by country, control area, or bidding zone, covering 17 countries in Europe. Its breadth enables an analysis of load variations on a regional scale across different countries. Fig. 2 represents the three test datasets, emphasizing the diverse magnitudes, consumption trajectories, and overall complexity.

In our experiments, we adhere to a standardized training–testing protocol for all datasets, maintaining a consistent 15-minute resolution. We utilize data covering a full year, encompassing 35,040 time steps, as our training dataset. This full-year span is crucial, as it captures the entire spectrum of seasonal variations, including distinct load patterns associated with winter heating, summer cooling, and transitional periods in spring and fall. For each prediction, the timestamp of used data is advanced by one step, which ensures that our approach continuously adapts to the most immediate past conditions alongside seasonal changes. Such dynamic selection is pivotal for accurately forecasting energy demands that vary significantly throughout the year.

In the testing phase, we use a distinct segment of 5000 steps of data immediately following the training period. Executed in a sequential manner, this phase rigorously tests our method across 5000 individual forecasts. This methodological design is critical for demonstrating the capacity of our approach to accurately predict demand supporting intricate forecasting applications.

5. Experimental results

This section presents the performance results of the proposed load forecasting framework from four perspectives: prediction accuracy,

Table 1
Performance of load forecasting methods—UCI dataset.

Approach	RMSE	R^2	MAPE	MAE	SMAPE	TIC
LSTM	394.156	0.955	2.381	302.057	2.399	0.012
LSTM+Att	426.643	0.946	2.552	324.095	2.574	0.013
Bi-LSTM+Att	456.289	0.938	2.745	348.065	2.769	0.014
Qua+Linear	933.757	0.705	3.496	462.876	3.614	0.018
Qua+RF	356.248	0.967	1.173	152.858	1.191	0.006
Our method	271.404	0.981	0.710	95.731	0.699	0.004

Table 2
Performance of load forecasting methods—OPSD dataset.

Approach	RMSE	R^2	MAPE	MAE	SMAPE	TIC
LSTM	197.982	0.973	2.080	155.476	2.059	0.011
LSTM+Att	190.247	0.976	2.010	148.031	2.000	0.010
Bi-LSTM+Att	199.460	0.974	2.116	155.312	2.108	0.010
Qua+Linear	501.474	0.829	4.256	320.770	4.136	0.021
Qua+RF	107.069	0.993	0.755	57.951	0.759	0.004
Our method	46.169	0.999	0.163	11.947	0.162	0.001

Table 3
Performance of load forecasting methods—UCSD dataset.

Approach	RMSE	R^2	MAPE	MAE	SMAPE	TIC
LSTM	428.076	0.966	0.918 1	302.524	0.920	0.005
LSTM+Att	453.665	0.959	1.015	335.754	1.016	0.005
Bi-LSTM+Att	468.488	0.955	1.070	354.565	1.071	0.005
Qua+Linear	785.886	0.877	1.122	378.678	1.147	0.006
Qua+RF	350.477	0.977	0.530	175.725	0.531	0.003
Our method	275.151	0.987	0.374	124.382	0.374	0.002

runtime, and robustness. Several parameters are empirically chosen or optimized to serve our forecasting process.

First, we consider the parameter $m = 7$, which denotes the number of 24-h time series selected for training. This is aligned with a weekly cycle with seven days, such that our method can effectively capture the distinct characteristics of energy usage within a complete weekly cycle. Next, $h = 96$ denotes the number of intervals in a 24-h period based on the datasets' 15-minute resolution.

Fine-tuning the quantile parameter k is crucial for balancing the accuracy of prediction outputs with their utility in real-world scenarios. In addition, α is a key indicator of how our outputs perform across practical scenarios. After extensive validation, we choose $\alpha = 0.1$ and $k = 0.7$. This careful selection is aimed at creating prediction intervals that balance accuracy and practical applicability, addressing challenges posed by our real-world datasets. Parameters β and γ are pivotal in crafting the reward function, catering to accurate forecasts and addressing prediction errors, respectively. These parameters are fine-tuned to ensure that the reward mechanism dynamically aligns with the predictor's performance, providing precise feedback for continuous improvement. Extensive calibration suggests the settings as $\beta = 10$ and $\gamma = -100$. This calibration is guided by a thorough analysis of the adaptability of the proposed method to changing load patterns, thus enhancing its accuracy and reliability across various real-world applications. All experiments in this study are conducted on a laptop with a 2.30 GHz Intel Core i9 processor, 64 GB RAM, running Windows 11.

5.1. Accuracy evaluation

The proposed accuracy evaluation in this work is conducted from two perspectives: forecasting results accuracy and the flexibility of quantile regression-based methods. We first analyze the direct forecasting accuracy of all the methods involved in this study using a diverse set of metrics. In our study, we emphasize that an effective STLF approach must reliably deliver superior results, regardless of the specific evaluation metric. To this end, six evaluation metrics are involved, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), the coefficient of determination (R^2), and Theil's Inequality Coefficient (TIC).

Next, we examine the flexibility of prediction intervals in methods that use quantile regression, focusing on the average interval sizes and their distribution. This evaluation aims to comprehensively highlight the strengths and distinctions of our method relative to benchmark strategies.

5.1.1. Forecasting accuracy: An analysis across six metrics

The performance of our proposed forecasting method is compared with other comparative benchmarks. A distinguishing characteristic of our method and the other two quantile regression-based benchmark methods is the generation of prediction intervals for each time step, as opposed to traditional point forecasts. Therefore, in the evaluation, we set the actual data points as the standard. If an interval encompasses the actual load, it is deemed accurate, and the real load is used as the prediction result for the evaluation. Conversely, if the interval does not include the real load, the boundary closer to the actual data is used as the prediction result for evaluation.

Tables 1, 2, and 3 show the 5000-step prediction errors for the UCI, OPSD, and UCSD datasets. Our framework consistently provides the smallest errors and highest R^2 scores, demonstrating superior accuracy. Qua+RF ranks second, and Qua+Linear shows the poorest performance. Attention-empowered LSTM and Bi-LSTM do not consistently outperform single LSTM, with significant discrepancies in some metrics.

To further clarify the results presented in Tables 1, 2, and 3, and to highlight the superior accuracy of our methodology across various metrics, we introduce part (a) in Figs. 3, 4, and 5. This part aims to visually present the findings shown in these tables, thereby facilitating a more intuitive understanding of the advantages our approach provides over existing forecasting methods. To ensure consistency and comparability, the values are first scaled to fit within the range [0, 1]. Subsequently, these scaled values are subtracted from 1, with the exception of the variable R^2. Namely, each index closer to the edge corresponds to more accurate results. These three figures across three datasets consistently provide a straightforward way to further demonstrate the advantages of our method compared to the five benchmarks. The results highlight the effectiveness of our RL-empowered online learning mechanism. Notably, our proposed method achieves superior performance even with a smaller amount of training data. Compared to state-of-the-art attention-based optimized DL-based methods, it underscores the efficiency and effectiveness of our online data selection mechanism in real-time load forecasting.

Given the nature of employing quantile regression for forecasting, we conduct additional discussions and result analyses in Section 5.1.2 to further evaluate the performance of our method in comparison with two other quantile-based benchmarks.

5.1.2. Flexibility assessment of prediction interval for quantile regression-empowered methods

Predictors leveraging quantile regression produce a prediction interval for each forecasting step. To further evaluate the flexibility of the results of quantile regression-based strategies, a two-dimensional assessment is proposed.

- Average interval size: The value of a prediction can be compromised by overly broad or exceedingly narrow intervals. While expansive intervals might weaken the prediction's effectiveness, excessively narrow intervals can make them seem less reliable.
- Distribution: Ensuring a stable and reasonable prediction interval size distribution across varied steps and datasets is essential for evaluating a method's applicability in real-world scenarios.

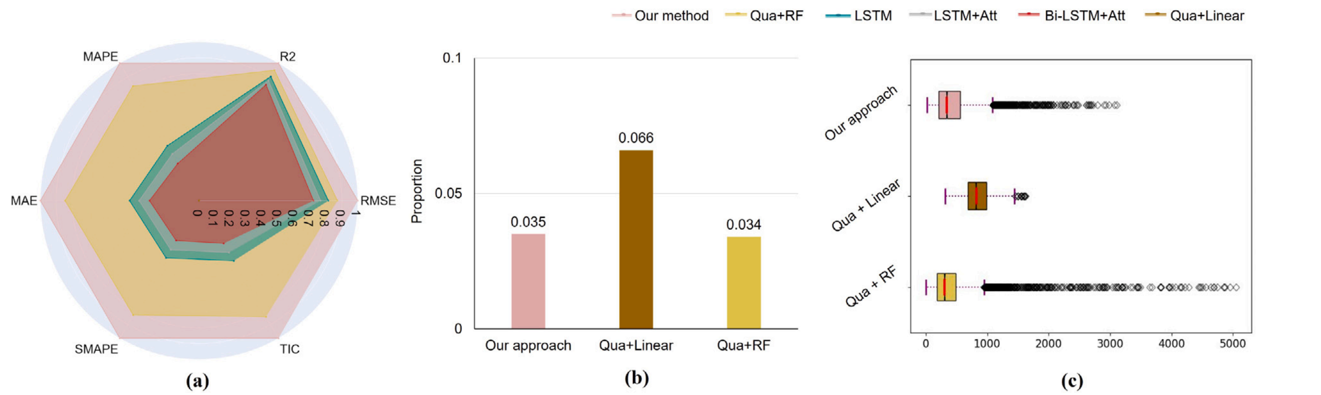

Fig. 3. Comprehensive forecasting analysis based on UCI dataset: Accuracy, Proportion, and Distribution. (a) Accuracy assessment using six metrics (b) Proportional evaluation of average interval size to test load (c) Interval size analysis: quartile distribution evaluation.

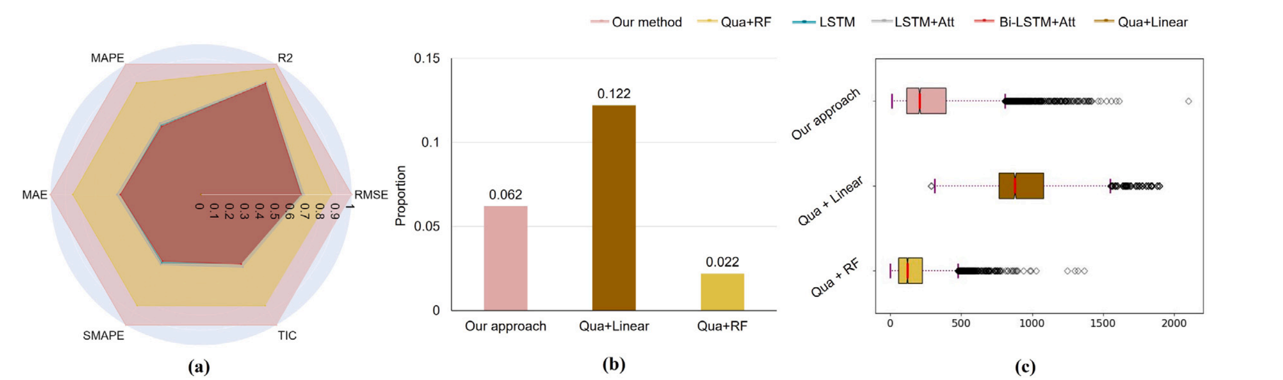

Fig. 4. Comprehensive forecasting analysis based on OPSD dataset: Accuracy, Proportion, and Distribution. (a) Accuracy assessment using six metrics (b) Proportional evaluation of average interval size to test load (c) Interval size analysis: quartile distribution evaluation.

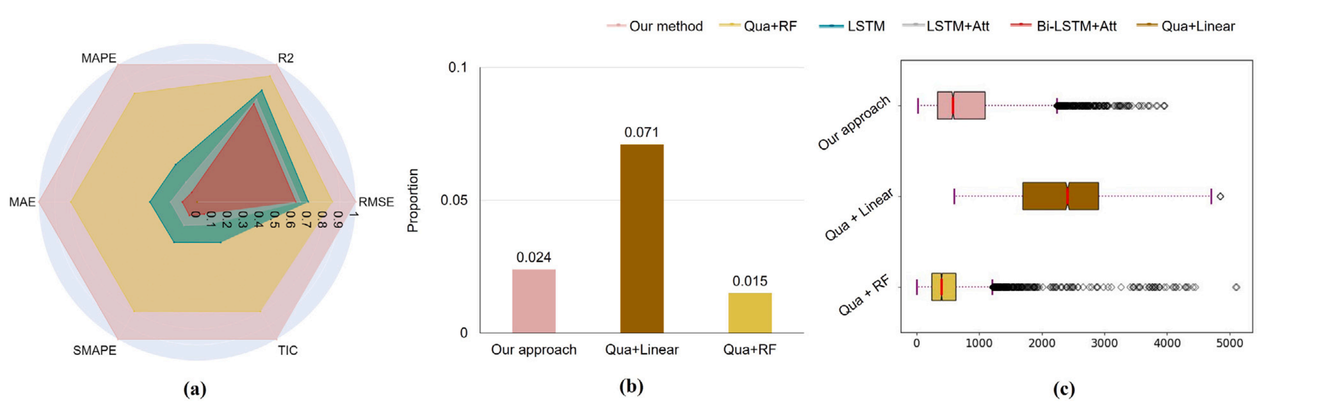

Fig. 5. Comprehensive forecasting analysis based on UCSD dataset: Accuracy, Proportion, and Distribution. (a) Accuracy assessment using six metrics (b) Proportional evaluation of average interval size to test load (c) Interval size analysis: quartile distribution evaluation.

Ideally, the predictor should maintain consistent and suitable interval sizes throughout the testing phase.

For our study, we employed a fixed quantile value of 0.7 to evaluate the prediction intervals of our method against two other quantile regression-empowered approaches.

Part (b) in Figs. 3, 4, and 5 aims to evaluate the average interval size as a proportion of the respective test dataset for each case. Our method, as observed, produces larger average interval sizes than Qua+RF but remains smaller than Qua+Linear. While Qua+RF may seem superior based solely on interval size, the quartile statistics offer a different perspective.

Part (c) in Figs. 3, 4, and 5, on the other hand, shows quartile statistics across the three datasets. This approach enriches the analysis by offering insights from a distributional perspective, thereby complementing the evaluations conducted in parts (a) and (b) within the same figures. It is evident that Qua+RF, despite its smaller average interval size, shows significant fluctuations. Especially in extreme cases, its intervals frequently exceed the third quartile (Q3), as particularly observed with the UCI and UCSD datasets. This variability limits its ability to provide reasonable intervals consistently during real-time prediction. In contrast, our method proposes better stability and consistency across all the datasets.

In this section, we compare the performance of our method with five different advanced ML/DL-based benchmarking strategies. We widely employ three different datasets to assess the accuracy of these methods, and the results consistently demonstrate that while using the smallest

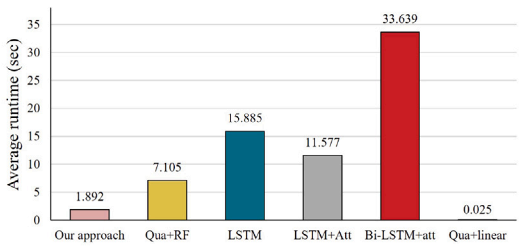

Fig. 6. Run-time performance evaluation.

amount of data, our method provides the most accurate outputs under the evaluation of six different metrics. Furthermore, considering the nature of leveraging quantile regression to conduct predictions, we further analyze the results of our method and other two quantile regression-based forecasting approaches. The results indicate that our method continuously provides reasonable prediction interval sizes across the three datasets. Its average interval size represents only 2% to 6% of the average load in each real-world power usage dataset, further emphasizing its practicality.

5.2. Runtime

Fig. 6 illustrates the average runtime per step across the three datasets. The bi-LSTM+Att strategy exhibits the most extended runtime, which aligns with expectations, considering it entails optimization over a bidirectional function. Following bi-LSTM+Att, the strategies that account for the second, third, and fourth longest runtimes are LSTM, LSTM+Att, and Qua+RF, respectively. On the contrary, our framework showcases a comparatively lower average runtime per step when juxtaposed with these benchmarking strategies. The only strategy with a shorter runtime than ours is Qua+Linear, which ranks as the second fastest among all the methodologies.

Therefore, our RL-based framework offers satisfactory runtime performance without needing specialized hardware like a GPU. Its precision in load forecasting and efficiency in computational time make it a viable solution for real-time applications where computational efficiency is crucial.

5.3. Robustness

The robustness evaluation is proposed to illustrate the robustness of the accuracy of a predictor when its target dataset changes. In this work, we point out that a practical STLF method should consistently maintain accuracy and timeliness across varying target datasets, exemplifying robustness against different forecasting scenarios. To measure the complexity of each test dataset or, in other words, its level of forecasting challenge, we use the Permutation Entropy (PE) method to evaluate the three datasets beforehand. PE values range between 0 and 1: a value closer to 1 signals a higher degree of complexity or unpredictability in the dataset, whereas a score nearing 0 indicates more order and predictability [32]. The PE results are shown in Fig. 7(a). The UCI dataset has the highest PE value, indicating the greatest challenges for forecasting, while the OPSD dataset has the lowest PE value, suggesting the lowest complexity or the most predictable pattern in the load series. These variances not only underscore the unique challenges posed by each dataset but also emphasize the broad spectrum of forecasting complexities they collectively encapsulate.

To better visualize the observations, in Fig. 7(b), we first choose the OPSD dataset, which has the lowest level of complexity, as the baseline for robustness analysis. The R^2 results are used as references for evaluation, as they range from 0 to 1 and are not affected by the scale of the datasets. It can be seen that when the target dataset changes from the baseline dataset (OPSD) to the more difficult dataset (UCSD),

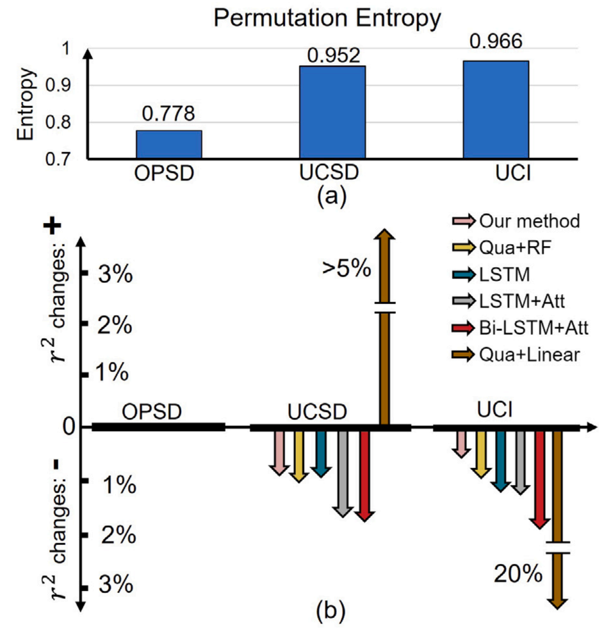

Fig. 7. Robustness evaluation—All datasets.

except for Qua+Linear, all other approaches consistently show lower R^2 values, indicating accuracy declines in terms of accuracy. Among them, Bi-LSTM+Att and LSTM+Att show the most remarkable declines. Our proposed method exhibits moderate changes, smaller than 1%. Second, when the baseline dataset changes from the UCSD to the more complex UCI dataset, all predictors show declines in accuracy, with Qua+Linear showing the biggest decline and our method providing the smallest.

Our proposed method consistently exhibits stable performance, even with varying complexity. This integration with accuracy evaluation reveals that our method not only offers precise forecasts but also demonstrates robustness against time series complexity. This makes it well-suited for real-world applications where both accuracy and robustness are vital.

6. Conclusion and discussions

In this paper, we have proposed and evaluated a reinforcement learning (RL)-enhanced online learning strategy tailored for short-term load forecasting (STLF). Notably, energy demand exhibits diverse patterns of individual user's activity over time, such as days and seasons, and also contains unpredictable fluctuation characteristics. Our method addressed this challenge through an online learning data selection mechanism, ensuring updated training data to meet four key criteria:

1. Reflect the characteristics of the current power usage;
2. Incorporate diverse scenarios to capture potential variations from regular user activity patterns, thereby enhancing load forecasting accuracy;
3. Ensure training efficiency through size-appropriate data selection;
4. Maintain a reasonable distribution of training samples to yield practical prediction interval sizes for real-world applications.

Our STLF method adeptly balances current data characteristics with related information to mitigate the impact of unexpected load fluctuations in real-time forecasting. It leverages RL for a dynamic training data selection mechanism that interacts with the dynamic prediction environment. Evaluated against five benchmark STLF strategies across three real-world datasets-ranging from institutional and residential settings within university campuses to the aggregated energy demands in regional grids across countries-our method demonstrates key advantages in accuracy, runtime, and robustness.

1. Accurate forecasting: Our approach outperforms representative benchmark STLF methods on three datasets across six evaluation metrics. It consistently provides accurate and stable prediction intervals that represent a minor fraction (2% to 6%) of the average load in tests, showing its precision and applicability.
2. Near real-time forecasting and high training-data efficiency: Relying on smaller training datasets, our method achieves superior accuracy without preprocessing and provides near real-time results with a mean runtime of approximately 2 s, without the need for advanced computing resources.
3. High robustness: Our approach exhibits exceptional robustness, delivering consistently accurate outputs across diverse datasets with varying complexity levels.

Our comprehensive analysis demonstrates the superiority of our method across key performance indicators: accuracy, runtime efficiency, and robustness. These results validate the adaptability of our approach across both small and large-scale power systems. These systems are responsible for meeting diverse energy demands spanning residential, commercial, and industrial loads and often integrating renewable energy sources. Our method's capacity to consistently provide accurate real-time forecasts can enable applications, such as real-time optimization of energy supply and management. These applications play a critical role in energy efficiency, grid reliability, and stability, contributing to smart and resilient power systems. We have made our method available as an open-source tool [18] to advance STLF and its practical applications. While promising, there are areas for future enhancement, including extending the one-step-ahead prediction horizon and automating parameter generation for real-time optimization, rather than relying on empirically determined values.

CRediT authorship contribution statement

Xinlin Wang: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Hao Wang:** Writing – review & editing, Conceptualization. **Shengping Li:** Writing – review & editing. **Haizhen Jin:** Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hao Wang reports article publishing charges was provided by Monash University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) under Grant DE230100046.

References

[1] Jiang Yuqi, Gao Tianlu, Dai Yuxin, Si Ruiqi, Hao Jun, Zhang Jun, et al. Very short-term residential load forecasting based on deep-autoformer. Appl Energy 2022;328:120120.
[2] Yang Yang, Wang Zijin, Zhao Shangrui, Wu Jinran. An integrated federated learning algorithm for short-term load forecasting. Electr Power Syst Res 2023;214:108830.
[3] Wang Xinlin, Wang Hao, Bhandari Binayak, Cheng Leming. AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response. Int J Precis Eng Manuf-Green Technol 2023;1–31.
[4] Xiao Xun, Mo Huadong, Zhang Yinan, Shan Guangcun. Meta-ANN–a dynamic artificial neural network refined by meta-learning for short-term load forecasting. Energy 2022;246:123418.
[5] Lin Jun, Ma Jin, Zhu Jianguo, Cui Yu. Short-term load forecasting based on LSTM networks considering attention mechanism. Int J Electr Power Energy Syst 2022;137:107818.
[6] Yan Qin, Lu Zhiying, Liu Hong, He Xingtang, Zhang Xihai, Guo Jianlin. An improved feature-time transformer encoder-bi-LSTM for short-term forecasting of user-level integrated energy loads. Energy Build 2023;297:113396.
[7] Wang Xinlin, Ahn Sung-Hoon. Real-time prediction and anomaly detection of electrical load in a residential community. Appl Energy 2020;259:114145.
[8] Wang Xinlin, Yao Zhihao, Papaefthymiou Marios. A real-time electrical load forecasting and unsupervised anomaly detection framework. Appl Energy 2023;330:120279.
[9] Wang Xinlin, Wang Hao, Ahn Sung-Hoon. Demand-side management for off-grid solar-powered microgrids: A case study of rural electrification in tanzania. Energy 2021;224:120229.
[10] Rubasinghe Osaka, Zhang Tingze, Zhang Xinan, Choi San Shing, Chau Tat Kei, Chow Yau, et al. Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high pv penetration. Appl Energy 2023;333:120641.
[11] Yi Shiyan, Liu Haichun, Chen Tao, Zhang Jianwen, Fan Yibo. A deep LSTM-CNN based on self-attention mechanism with input data reduction for short-term load forecasting. IET Gener Transm Distrib 2023;17(7):1538–52.
[12] Wang Hong, Alattas Khalid A, Mohammadzadeh Ardashir, Sabzalian Mohammad Hosein, Aly Ayman A, Mosavi Amir. Comprehensive review of load forecasting with emphasis on intelligent computing approaches. Energy Rep 2022;8:13189–98.
[13] Grondman Ivo, Busoniu Lucian, Lopes Gabriel AD, Babuska Robert. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. IEEE Trans Syst Man Cybern C (Appl Rev) 2012;42(6):1291–307.
[14] Berenji Hamid R, Vengerov David. A convergent actor-critic-based FRL algorithm with application to power management of wireless transmitters. IEEE Trans Fuzzy Syst 2003;11(4):478–85.
[15] Zhang Wenjie, Quan Hao, Srinivasan Dipti. Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination. Energy 2018;160:810–9.
[16] Silwal Sushil, Mullican Colton, Chen Yi-An, Ghosh Avik, Dilliott John, Kleissl Jan. Open-source multi-year power generation, consumption, and storage data in a microgrid. J Renew Sustain Energy 2021;13(2):025301.
[17] Keliris Anastasis, Konstantinou Charalambos, Sazos Marios, Maniatakos Michail. Open source intelligence for energy sector cyberattacks. Critical Infrastruct Secur Resil Theor Methods Tools Technol 2019;261–81.
[18] Wang Xinlin. Github page. 2023, online https://github.com/xinlin-CSIRO/prediction_work_with_reinforcement_learning.
[19] Wang Nier, Li Zhanming. Short term power load forecasting based on BES-vmd and CNN-Bi-LSTM method with error correction. Front Energy Res 2023;10.
[20] Yue Weimin, Liu Qingrong, Ruan Yingjun, Qian Fanyue, Meng Hua. A prediction approach with mode decomposition-recombination technique for short-term load forecasting. Sustainable Cities Soc 2022;85:104034.
[21] Feng Cong, Zhang Jie. Reinforcement learning based dynamic model selection for short-term load forecasting. In: 2019 IEEE power & energy society innovative smart grid technologies conference. IEEE; 2019, p. 1–5.
[22] Park Rae-Jun, Song Kyung-Bin, Kwon Bo-Sung. Short-term load forecasting algorithm using a similar day selection method based on reinforcement learning. Energies 2020;13(10):2640.
[23] Jia Heping, Wang Xuanyuan, Zhang Xian, Liu Dunnan. Short-term load forecasting for DERs based on CNN-LSTM with attention mechanism. In: Business models and reliable operation of virtual power plants. Springer; 2023, p. 9–18.
[24] Javed Umar, Ijaz Khalid, Jawad Muhammad, Khosa Ikramullah, Ansari Ejaz Ahmad, Zaidi Khurram Shabih, Rafiq Muhammad Nadeem, Shabbir Noman. A novel short receptive field based dilated causal convolutional network integrated with bidirectional LSTM for short-term load forecasting. Expert Syst Appl 2022;205:117689.

[25] Li Ao, Xiao Fu, Zhang Chong, Fan Cheng. Attention-based interpretable neural network for building cooling load prediction. Appl Energy 2021;299:117238.
[26] Kumari Aparna, Vekaria Darshan, Gupta Rajesh, Tanwar Sudeep. Redills: Deep learning-based secure data analytic framework for smart grid systems. In: 2020 IEEE international conference on communications workshops. IEEE; 2020, p. 1–6.
[27] Lin Jun, Ma Jin, Zhu Jianguo, Cui Yu. Short-term load forecasting based on LSTM networks considering attention mechanism. Int J Electr Power Energy Syst 2022;137:107818.
[28] Wang Fu-Kwun, Amogne Zemenu Endalamaw, Chou Jia-Hong, Tseng Cheng. Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism. Energy 2022;254:124344.
[29] Wang Shouxiang, Wang Xuan, Wang Shaomin, Wang Dan. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. Int J Electr Power Energy Syst 2019;109:470–9.
[30] Meinshausen Nicolai, Ridgeway Greg. Quantile regression forests.. J Mach Learn Res 2006;7(6).
[31] Quantile regression, https://scikit-learn.org/stable/auto_examples/linear_model/plot_quantile_regression.html.
[32] Bandt Christoph, Pompe Bernd. Permutation entropy: a natural complexity measure for time series. Phys Rev Lett 2002;88(17):174102.