Short-Term Load Forecasting via a Multi-Layer Network based on Feature Weight Optimization

* Corresponding author and e-mail: Junwei Cao, jcao@tsinghua.edu.cn
\# These authors contributed to the work equally and should be regarded as co-first authors

Yan Huang[#]
Department of Automation
Tsinghua University
Beijing, China
huang_yan@tsinghua.edu.cn

Jieqiong Gan[#]
Toyota Motor Technical Research And
Service (Shanghai) Co., Ltd.
Advanced Engineering Research Div.
Shanghai, China
ganjieqiong@ttrssh.com.cn

Zhongqi Wang
Toyota Motor Technical Research And
Service (Shanghai) Co., Ltd.
Advanced Engineering Research Div.
Shanghai, China
wangzhongqi@ttrssh.com.cn

Zhaozhe Zhong
Toyota Motor Technical Research And
Service (Shanghai) Co., Ltd.
Advanced Engineering Research Div.
Shanghai, China
zhongzhaozhe@ttrssh.com.cn

Jian Liu
Toyota Motor Technical Research And
Service (Shanghai) Co., Ltd.
Advanced Engineering Research Div.
Shanghai, China
liujian@ttrssh.com.cn

Junwei Cao[*]
Beijing National Research Center of
Information Science and Technology
Tsinghua University
Beijing, China
jcao@tsinghua.edu.cn

Abstract—**Load forecasting is a critical concern for the operation of regulated power systems and electricity markets. It can guarantee the availability of electricity supply, minimize both excess and insufficient utilization of generating capacity and therefore optimize energy prices. In this paper, a novel multi-layer forecasting neural network based on feature weight optimization is developed for general reliable short-term load forecasting. It utilizes historical load data combined with some influencing factors such as seasonal and daily patterns as well as social behavior and user habits. Specifically, the proposed network involves a set of convolutional features extraction, bidirectional gate recurrent unit (BiGRU) and automatic feature weight optimization strategy, which subtly picks the nonlinear relationships and complex dependencies to facilitate the efficient generation of the close-to-true values. Experimental results demonstrate that the network achieves competitive performance with significantly higher accuracy when compared to congeneric methods, which greatly indicates its potential in the load forecasting tasks of electricity resource management.**

Keywords—load forecasting, convolution, feature extraction, BiGRU, attention optimization

I. INTRODUCTION

Amidst the escalating intricacy and volatility within the electricity market, accurate and reliable load forecasting assumes paramount significance in guaranteeing a dependable and impregnable power supply. The ability to precisely anticipate electricity demand facilitates optimal resource allocation, strategic energy planning, and economically viable grid operation. Load forecasting provides great opportunities for implementing demand response initiatives, integrating renewable energy sources, and fortifying overall grid stability [1]. Consequently, the exploring endeavors pertaining to load forecasting algorithms strive to fashion robust and accurate models adept at navigating the dynamic character of energy consumption patterns and sundry influencing factors, encompassing weather conditions, economic indicators, and social dynamics [2]. Harnessing advanced load forecasting methodologies and data analytics technologies, the ultimate goal is to enhance the efficiency and reliability of power systems, curtail operational expenses, and further foster the sustainable energy management practices.

The present methods for short-term power load forecasting encompass various approaches, spanning from that statistical techniques [3, 4], machine learning based methods [6-10], hybrid or ensemble approaches [11-13], to advanced algorithms [14, 15]. Their principal aim is to discern an optimal model by utilizing historical load and influencing data to facilitate rapid and precise forecasts of future load information across both single-step and multi-step scenarios.

Statistical based load forecasting strive to create a direct or indirect nexus between the projected values and historical data. It involves a breadth of techniques including time series analysis [4], regression theory [16], load derivative strategy [17], similar day algorithm [18], weather-based analysis [19], load profiling [20], and exponential smoothing [21]. Time series forecasting focuses on analyzing and forecasting load demand based on historical time-stamped data. Popular time series models used in load forecasting include autoregressive integrated moving average and seasonal decomposition of time series [3]. Regression theory based analysis entails establishing a mathematical relationship between load demand and pertinent explanatory variables, such as historical load data, weather conditions, and socio-economic factors. Common regressions for power load forecasting include linear regression, multiple regression and polynomial regression. Load derivative based forecasting is based on the assumption that changes in load demand can be predicted based on the rate of change of previous load readings, which involves load growth models, load ramp models and load acceleration models [22]. Similar day method tends to identify historical days with similar load demand patterns to the forecasted day and use their load data as predictors, which usually involves similar day clustering, nearest neighbor and pattern matching. Weather-based technique analyzes the impact of weather conditions on electricity demand by analyzing past weather relationships and incorporating weather variables, such as temperature, humidity, and wind speed, as predictors in load forecasting models. On the other hand, load profiling strategy aims to segment load data into different customer-specific or region-specific groups based on their load patterns and then use them to forecast future load demand based on the historical load behavior of each group [23]. Besides, exponential smoothing, including simple, double, as well as triple exponential smoothing, is commonly used in load forecasting to predict the load data by employing weighted combinations. They are beneficial in capturing trends, seasonality, and other patterns in the load data, which makes them effective for general forecasting tasks. These mentioned

statistical methods can be integrated and tailored according to the distinctive attributes and demands of the assignment at hand. Sometimes, machine learning or advanced techniques are often employed in conjunction with these techniques to augment the precision and efficacy of the power load forecasts.

Machine learning-based techniques for load forecasting are intended to utilize neural networks with multiple hidden layers to predict load demand accurately. They leverages the ability of deep learning models to automatically learn and extract complex patterns and relationships from historical load data. Common prominent models include recurrent neural networks (RNN), convolutional neural networks (CNN), long short-term memory (LSTM), gated recurrent unit (GRU), generative adversarial network (GAN), autoencoders and variational autoencoders (VAEs), attention-based networks, and diverse iterations of their different variant [6] and pure innovative architectures [10]. These learning-based networks have shown promising results in load forecasting tasks by effectively capturing complex patterns and dependencies in historical load data. They provide accurate predictions for short-term and long-term load demand. CNN seems to excel at capturing local patterns, while LSTM is better suited for handling long-term time dependencies. If designing different feature extracting and relationship capturing modules, or incorporating other processing layer in a network architecture, may be it can potentially enhance the performance of general load forecasting tasks and provide a unique experience.

Hybrid or ensemble approaches in load forecasting mean the amalgamation of multiple forecasting techniques or models to improve the accuracy and reliability of load predictions. They amalgamate the strengths of different forecasting models and algorithms to overcome the limitations of individual methods when predicting electricity demand. Popular combinations mainly involve that weighted averaging [24], stacking ensemble [25], neural network ensemble [13], or hybrid model that both combining statistical and artificial intelligence forecasting methods [11]. This type of method poses the potential to achieve improved predictive accuracy, increased robustness and reduced overfitting when compared to using a single approach or model.

Furthermore, the extensive utilization of deep learning technology has contributed to the emergence of advanced load forecasting algorithms [26]. These methods can capture complex temporal dependencies and non-linear relationships in the data. In realistic load forecasting task, the choice of forecasting method depends on the specific application, available data, computational resources, and desired accuracy. In this paper, we incorporate the strengths of convolution process and an attention based gate recurrent unit to provide a robust multi-layer forecasting network. It adeptly extracts the spatial and temporal features from the input data and effectively learns from bidirectional sequence information that connected with an integrated attention. The goal of this paper is to ensure the method excels in providing reliable and precise forecasts for diversiform power load data.

II. THE MULTI-LAYER NETWORK AND DATABSE

A. Problem Statement

Given a set of input sequential power load data, which is denoted as **Data_in**=$\{x_{nm} \mid n=1,2,3. \ m=1,2,\ldots,+\infty\}$. The goal of this paper is to explore an appropriate model $f(*)$ that can output the accurate predictions **Data_out**=$\{O_1, O_2,\ldots,O_t\}$ of

multi-step future events or outcomes for any input. The network is expected to facilely estimate the likelihood of specific future events, states or outcomes by training the historical and referenced input data to capture the relations, patterns, trends, and correspondences within the input data. This can allow for informed decision-making and proactive measures to be taken based on the anticipated outcomes.

B. Overview of the Multi-Layer Network

The proposed multi-layer network is mainly designed upon a simplified version of convolutional neural network layer, expansion, a bidirectional gate recurrent unit layer and an integrated multi-head attention layer. Fig. 1 demonstrates the framework of the multi-layer load forecasting network, which explicitly describes the workflow from input to the output. Specially, the input data is needed preprocessed before being fed into the network. In the forecasting network, the operation of one-dimensional convolution and expansion is applied to extract the spatial features of seasonal and daily patterns from the input multi-dimensional data. We utilize convolution kernels of varying sizes to capture feature relations across different scales, and apply the *ReLU* to nonlinearly process the convolution results. The obtained feature maps will be transferred into a BiGRU layer, which is responsible for acquiring context-based representations of sequences and generating feature representations at the sequence level. These feature representations will undergo weighting and integration within the attention layer to allocate distinct attention weights to various input positions. This enables the network to focus more on sequence fragments that exert significant influence on the multi-step prediction outcomes, thereby enhancing the network's performance and generalization capabilities. Finally, a fully connected (FC) layer and *ReLU* function are constructed to transform the learned features and representations into the final output.

C. Load Dataset and Data Preprocessing

This paper uses the actual measurements of power load data from a company in Shanghai with resolution of 15 minutes per point from two incoming lines to achieve the training and test of the proposed multi-layer network. According to the nonlinear characteristics of the company's power consumption, we consider the influence characteristics of multiple inputs, including seasonal and daily patterns (*working/holiday time*), social behavior and user habits (*on/off mode of incoming line*) to perform the actual multi-input and multi-out load forecasting tasks.

In this work, data preprocessing of the load forecasting tasks mainly involves several key steps. Firstly, it typically includes data cleaning, where any missing or data is identified and addressed. The multiple linear regression strategy is applied to estimate and adjust missing and inconsistent values based on other sequential data features. Secondly, the data is normalized with min-max normalization to scale the input load data within a specific range to avoid bias towards certain variables. They are normalized by the maximum value of each feature to ensure they have a similar scale, which aids in the training process. Temporal aggregation and resampling of the data is necessary to align with the desired forecasting horizon. Furthermore, the load dataset is typically split into training, validation, as well as testing sets to evaluate the performance of the forecasting network. The preprocessing ensures that the data is suitably prepared for the neural network model to effectively learn and forecast the power load data.
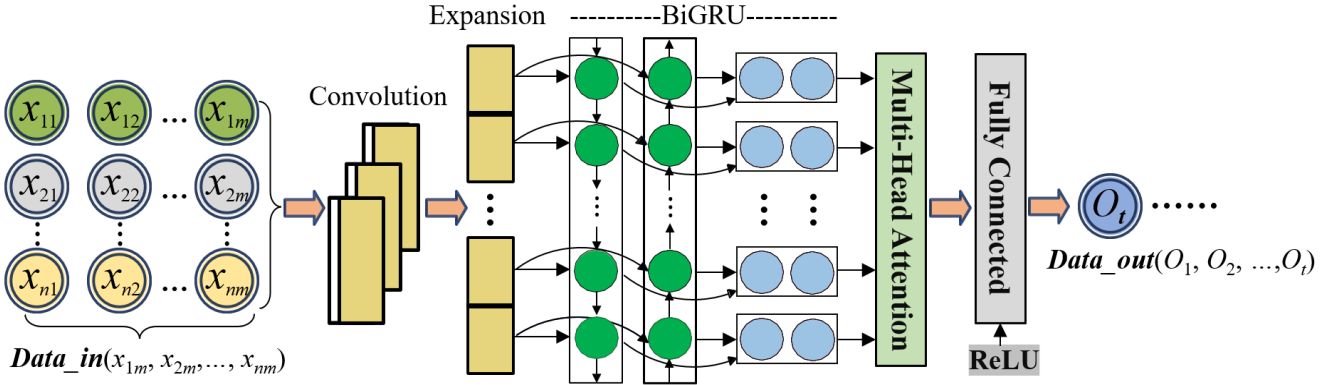
Fig. 1. Framework of the proposed multi-layer neural network for load forecasting, which depicts the process of the input load data to that forecasting output.

D. Convolution and Expansion Operation

The one-dimensional convolution operation is used to extract temporal features from time series data. The process involves sliding a one-dimensional convolutional filter over the input time series to capture local patterns at different temporal positions. It facilitates the model to identify and extract relevant features that are critical for load forecasting, such as daily or weekly patterns.

Denote the input time series data as X of length L and the convolutional filter as F with length K. The convolution operation at position i can be expressed as:

$$(X * F)_i = \sum_{k=0}^{k-1} X_{i+k} \times F_k , \qquad (1)$$

Where $(X * F)_i$ represents the result of the convolution at position i, $*$ is the convolution operation, X_{i+k} denotes the element in the input time series at position $i+k$, and F_k represents the k-th element of the filter. The result of the summation represents the output of the convolution operation at a particular position i, which captures the local relationship between elements of the input time series and the filter. The operation is central to the process of capturing relevant temporal patterns and features from time series data. After the convolution operation, an $ReLU$ activation function will be applied to introduce non-linearity. The convolution operation and linear process are defined with (4) and (5), respectively,

$$M_t = ReLU \left(\sum_{j=1}^{j} w_t * x_{nm} + b_t \right) , \qquad (2)$$

$$T_l = ReLU (C_{l-1} w_l + b_l) , \qquad (3)$$

where M_t denotes the feature map subsequent to the intricate mapping process, j represents the index value of the filter, $*$ indicates the convolution operation, and that x_{nm} is the input element of the sequential load data, w_t is the weight matrix of the convolution kernel, b_t represents the bias vector. In equation (3), C_l is the output of the l-th layer, W_l is the connected weight matrix and b_l is the bias.

The convolution is followed by an expansion operation to increase the perceptual range of the model, which aims to capture more longer-term time dependencies and broader spatial features in the input data. The designed expansion can enhance the forecasting performance of the network since it allows the model to better understand complex time series patterns and features.

E. Bidirectional Gate Recurrent Unit Layer

Bidirectional gate recurrent unit incorporates the benefits of bidirectionality and gating mechanisms, which is designed to effectively process and model sequential data by capturing dependencies and patterns within the input sequence. The bidirectionality feature allows the structural unit to process the input sequence in both forward and backward directions, enabling it to effectively capture past and future context simultaneously. Moreover, the gating mechanisms can facilitate the network's ability to selectively retain or discard information at each time step, thereby enhancing its capacity to handle long-range dependencies and mitigate the vanishing gradient problem. For an arbitrary input sequence load data $X = \{x_{1m}, x_{2m}, \ldots, x_{tm}\}$, $X \subseteq$ **Data_in**. The forward and backward calculations of the gate recurrent unit layer can be expressed as the equation (7) and (8), respectively.

$$\vec{h}_t = GRU (\vec{h}_{t-1}, x_t) , \qquad (4)$$

$$\overleftarrow{h}_t = GRU (\overleftarrow{h}_{t-1}, x_t) , \qquad (5)$$

$$h_t = W_t \cdot \vec{h}_t + f_t \cdot \overleftarrow{h}_t + b_t , \qquad (6)$$

in which, $\vec{h}_t$ and $\overleftarrow{h}_t$ represent the forward hidden and the backward hidden state at time t, respectively. W_t, f_t and b_t represent the corresponding weights and the bias term of the hidden layer at time t, respectively.

F. Multi-Head Attention Layer

In the proposed network, multi-head attention from [27] is improved into the attention layer to enable capturing and optimizing complex relationships and dependencies. Fig. 2 describes the structural workflow of the multi-head attention layer on the input features values to generate the linear weight transformation and optimization. It usually contains an input, linear projections, attention heads, as well as concatenation and linear transformation. The input is a sequence of vectors extracted from the previous layer, which is then linearly projected into different subspaces to create queries, keys, and values for each attention head. The multi-head attention consists of multiple independent attention mechanisms, or "heads," which perform the attention operation in parallel. The aim of the layer is to enable the multi-layer network to jointly attend to different parts of the input sequence in parallel, and facilitate a more comprehensive capture of a broader range of relationships and patterns within the data.

In this paper, a scaled dot-product attention is applied and integrated in the multi-head attention structure. After the dot-product, the scale processing is conducted to avoid excessive dot-product results entering the saturated area of SoftMax. Given an input sequence of vectors $Y = [y_1, y_2, \ldots, y_n]$, Define Q to be the query matrix, which contains the queries derived from the input sequence, which denoted as the set of vectors representing what needs to be attended to. Define K to be the
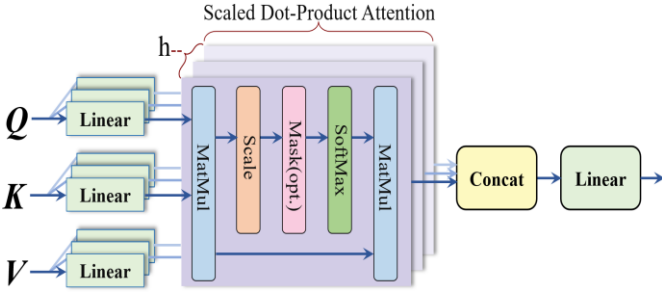
Fig. 2. The workflow of multi-head attention in the proposed neural network.

Fig. 3. The training and test workflow of the proposed forecasting network.

key matrix, containing the keys derived from the input sequence, representing the source of information with which the queries are compared. Let V represent the value matrix, containing the values derived from the input sequence, representing the information that needs to be combined or attended to. The formulation of a single attention head in the attention layer of the multi-layer network can be described as:

1) *Linear projection*. Calculate the query Q, key K, and value V matrices by linearly transforming $\mathbf{Y}$ using that $Q=\mathbf{Y}W_Q$, $K=\mathbf{Y}W_K$, $V=\mathbf{Y}W_v$, where W_Q, W_K, and W_v are the learnable linear projection matrices.

2) *Scaled dot-product attention*. For each query in Q, compute its attention scores with all keys in K, scaled by the square root of the dimension of the keys with

$$\boldsymbol{Attention}\ (Q, K, V)=\mathrm{SoftMax}(\frac{Q K^{\mathrm{T}}}{\sqrt{d_k}})V , \quad (7)$$

where d_k is the dimension of the keys.

3) *Multi-head attention*. Repeat the process 1) and 2) for multiple attention heads in parallel. When obtaining the outputs from each head, concatenate them and apply another linear transformation to produce the final multi-head attention output.

The designed multi-head attention layer can help achieve self-attention and linear weight optimization. Compared with a simple attention mechanism, it calculates the results of attention separately after multiple routing transformations of the input, and performs linear transformations again after concatenating all the results as the output. It supports parallel processing of different aspects of the input load data, which greatly enables the proposed network to capture diverse linear patterns and dependencies.

G. Fully Connected Layer

The FC layer receives input feature from the preceding layer, and each neuron in the layer is connected to every neuron in the previous layer. The layer applies a linear transformation to the input data followed by a non-linear activation function to capture complex patterns and relationships in the data. The output of a single neuron in a fully connected layer is calculated with the equation,

$$\varphi_{out}=ReLU(\sum_{i=1}^{N} w_i\mu_i+b^T) , \quad (8)$$

where φ_{out} is the output of the neuron, N is the number of input features, w_i are the weights associated with each input feature, μ_i are the input features, b^T is the bias term. The connected FC layer finally integrates complex patterns in the input features through linear transformations and non-linear activation functions, which helps the network to capture intricate relationships and make accurate predictions.

H. Model Training

Fig. 3 depicts the training and test workflow of the muti-layer network. The dataset is partitioned into a training set
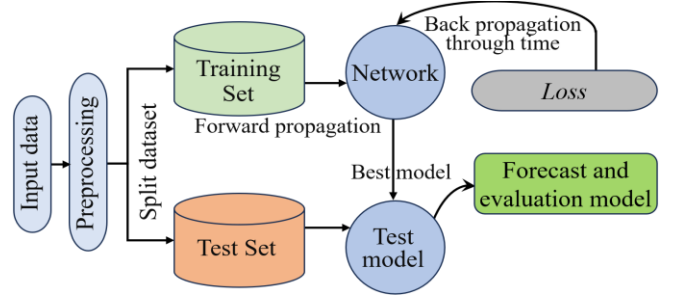
and a test set according to a predetermined ratio. The training set is utilized for training the network component, while the test set is used to evaluate the effectiveness of forecasting results. During the training process, a backpropagation mechanism is employed to compare the forecasted results with the actual results, compute the loss function, and adjust the network parameters to minimize the loss. The training dataset is typically divided into small batches for training, with the data in each batch undergoing forward and backward propagation in the network for parameter updates. The process is iteratively repeated until the predetermined number of training epochs or a stopping condition is reached. Subsequently, the trained model can be directly employed to conduct the forecasting tasks and evaluations to assess the effectiveness of forecasting results.

III. ANALYSIS OF EXPERIMENTAL RESULTS

A. Experimental Design

In order to verify the prediction accuracy of the proposed network, this paper conducts six different related networks with varying complexities for evaluation, including CNN, LSTM, GRU, LSTM-attention, BiGRU-attention, and Bi-LSTM. We use the actual dataset to perform the experimental comparisons by conducting future multi-step prediction. Multi-step forecasts in this paper means 96-steps that the networks need to forecast all the values from the next point in time to 24 o'clock on the same day at one time. The metrics of mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE), and R-squared (R^2) are applied to evaluate the quality of the comparative networks. Lower MAPE, RMSE and MAE values indicate a higher accuracy and reliability of the prediction network. A higher R^2 value implies a better fit between the predicted and actual values. The collaboration of these metrics can greatly elaborately describe the integral performance of a network.

B. Subjective Evaluation

Fig. 4 describes some sampling points of the load forecasting results from the comparative networks, which is plotted by overlaying the predicted power load output values from the model with the actual measured values. It indicates that the predicted values of the seven comparison networks have good overlap with the trend of the actual values. From the sampling results, the proposed network presents a prediction result closer to the actual value, which can be further verified in Fig. 5. It presents future 12 days perdition data, whose results accurately predict the occurrence of peak and off-peak periods, as well as the timing of turning points, enabling accurate predictions of when the system will reach its maximum and minimum output. From Fig. 4, the proposed model effectively identifies the turning points, indicating the moments when the load output transitions from increasing to decreasing or vice versa. This indicates that the network is
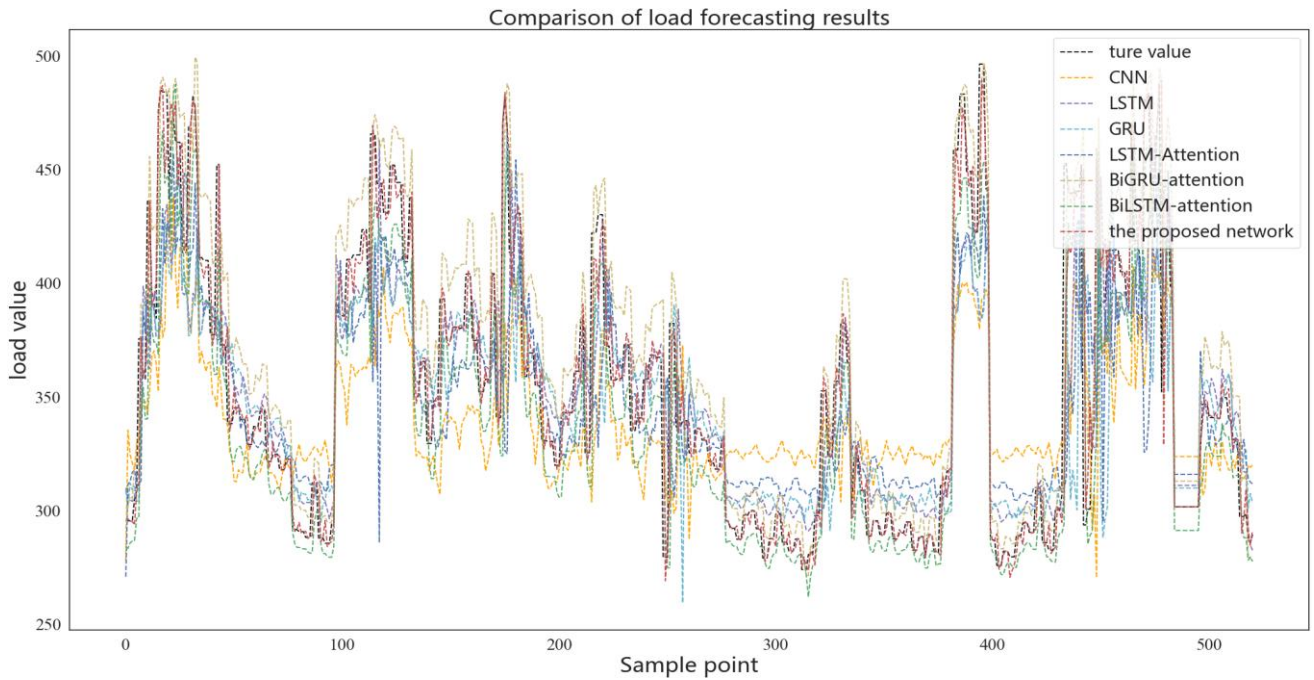
Fig. 4. Representative plotted results of 96-steps load forecasting from all the comparative networks with actual measurements from a company in Shanghai.
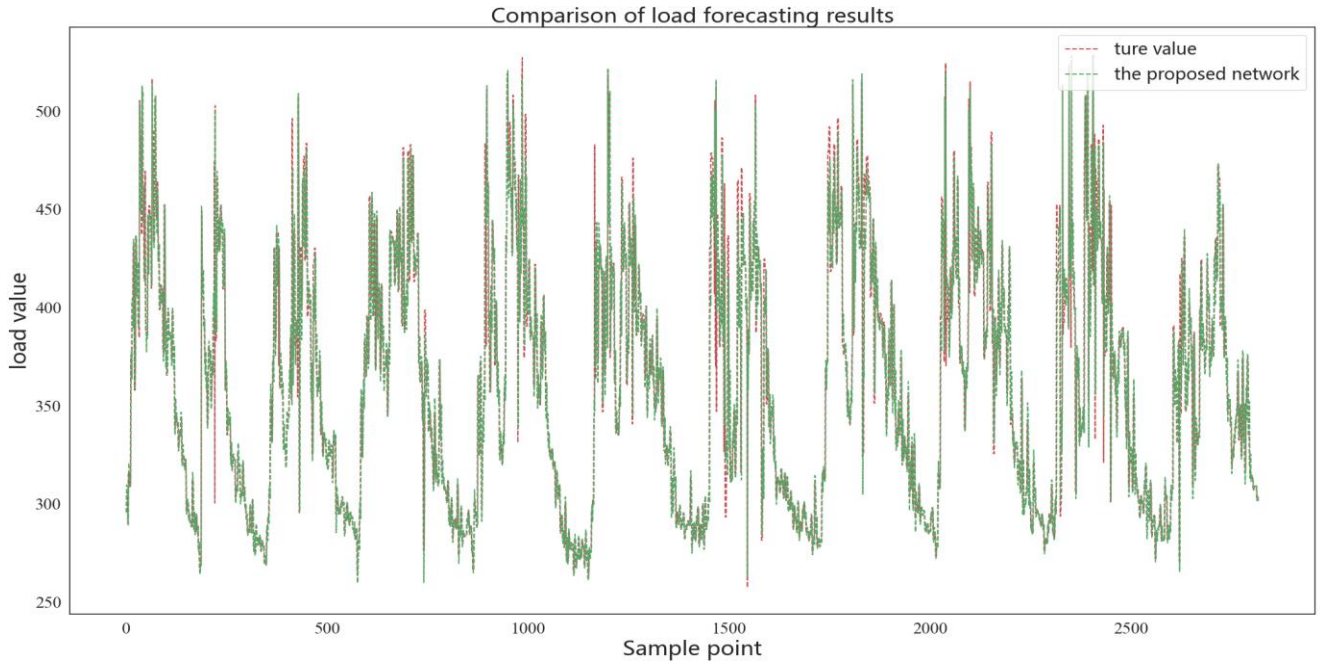

Fig. 5. Representative plotted results of 96-step load forecasting using the load data of consecutive days with the proposed multi-layer network.

able to capture the underlying patterns and variations in the power load output accurately. The alignment between the curves suggests that the network's predictions are in line with the actual power load production, thereby validating the effectiveness of the forecasting model.

C. Objective Evaluation

Table 1 illustrates the evaluated metrics of different forecasting networks. From the table, we can conclude that our network achieves obvious performance on the data of metrics. Specifically, there is no significant difference in the load forecasting performance between CNN, LSTM and GRU network. The proposed multi-layer network reduces 69.33% in MAPE , 76.90% in RMSE, 69.23% in MAE and gains a 22.74% improvement in R2 on average when compared with that CNN prediction network. It reduces 67.61% on average in MAPE, 74.68% in RMSE, 63.73 % in MAE and gains a

13.36% improvement in R2 over LSTM network, and reduces 66.67% on average in MAPE, 72.53% in RMSE, 59.86 % in MAE and gains a 13.75% improvement in R2 over GRU network, and it can reduce 56.60 % on average in MAPE, 66.84% in RMSE, 42.44% in MAE and gains a 7.24% improvement in R2 over LSTM-Attention, reduces 51.06% on average in MAPE, 61.97% in that RMSE, 32.17 % in MAE and gains a 5.64% improvement in R2 over that BiLSTM-Attention network, and reduces 47.73% on average in the metric of MAPE, 46.31% in RMSE,27.53 % in MAE and gains 4.97% improvement in R^2 over BiGRU-Attention network. The significant lower MAPE, RMSE, MAE values and the higher R^2 clearly establish the superior accuracy, precision, and reliability of the proposed multi-layer load forecasting network when compared to existing similar type of power load forecasting networks.

TABLE I. THE EVALUATED METRICS OF DIFFERENT METHODS.

Forecasting Method	Average Value of Each Metric			
	MAPE	RMSE	MAE	R^2
CNN	0.075	47.881	21.392	0.809
LSTM	0.071	43.695	18.146	0.876
GRU	0.069	43.662	18.398	0.873
LSTM-Attention	0.053	33.360	11.435	0.926
BiLSTM-Attention	0.047	29.086	9.704	0.940
BiGRU-Attention	0.044	20.605	9.082	0.946
The proposed network	**0.023**	**11.062**	**6.582**	**0.993**

IV. CONCLUSION

In this paper, we propose a novel multi-layer network based on feature weight optimization for general short-term load forecasting tasks. The motivation behind the method is that a convolution and expansion operation is responsible for extracting spatial features from the multi-input load data, a bidirectional gate recurrent unit layer tends to capture long-term temporal dependencies by learning and memorizing sequential context information progressively through multiple sequential steps, and an attention is employed to weight and integrate the outputs of the two layers to pursue enhanced forecasting accuracy. In comparison to some comparative forecasting models, the proposed network yields more effective performance in terms of improved accuracy and robustness in power load forecasts. The results indicate the great potential of the network on forecasting various challenging power load data.

REFERENCES

[1] Y. Zhang, H. Wen, Q. Wu, and Q. Ai, "Optimal adaptive prediction intervals for electricity load forecasting in distribution systems via reinforcement learning," IEEE Transactions on Smart Grid, vol. 14, no. 4, pp. 3259-3270, July 2023.

[2] G.Z. Yu, L. Lu, B. Tang, S.Y. Wang, and C.Y. Chung, "Ultra-Short-Term wind power subsection forecasting method based on extreme weather," IEEE Transactions on Power Systems, vol. 38, no. 6, pp. 5045-5056, November 2023.

[3] A.G. Rusina, T. Osgonbaatar, and P.V. Matrenin, "Short-term load forecasting using statistical methods for the central power system of mongolia," International Multi-Conference on Engineering, Computer and Information Sciences, Yekaterinburg, 2022, pp. 2030-2035.

[4] P. N. Mahalle and K. D. Bhambare, "Direct load forecasting using combination techniques of time series models," International Journal of Electrical Power & Energy Systems, vol. 118, pp. 105792, Jan. 2020.

[5] H. Hou, C. Liu, Q. Wang, et al., "Review of load forecasting based on artificial intelligence methodologies, models, and challenges," Electric Power Systems Research, vol. 210, pp. 1-11, September, 2022.

[6] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using q-learning technique," IEEE Transactions on Industrial Informatics, vol. 17, no. 6, pp. 4229-4237, June 2021.

[7] S. C. Pall, S. H. Cheragee, M. R. Mahamud, H. M. Mishu, and H. Biswas, "Performance analysis of machine learning and ann for load forecasting in bangladesh," International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, 2023, pp. 1-5.

[8] A. Jain and S. C. Gupta, "Peak load forecasting using machine learning algorithms," IEEE Renewable Energy and Sustainable E-Mobility Conference (RESEM), Bhopal, 2023, pp. 1-4.

[9] S. Mondal, S. Gatade, M. S. Sreekar, A. Raj, and R. K. Mogral, "A comparative study on deep learning methods for forecasting load in smart grid," IEEE International Conference on Electronics, Computing and Communication Technologies, Bangalore, 2023, pp. 1-5.

[10] G. Tziolis, A. Livera, J. Montes-Romero, S. Theocharides, G. Makrides, and G. E. Georghiou, "Direct short-term net load forecasting based on machine learning principles for solar-integrated microgrids," IEEE Access, vol. 11, pp. 102038-102049, 2023.

[11] W. Cui, C. Wan, and Y. Song, "Ensemble deep learning-based non-crossing quantile regression for nonparametric probabilistic forecasting of wind power generation," IEEE Transactions on Power Systems, vol. 38, no. 4, pp. 3163-3178, July 2023.

[12] Z. Ping, Y. Liang, S. Qiu, X. Xu, X. Cui, and Y. Han, "Sparse bayesian learning based hybrid method for load forecasting in integrated energy systems," China Automation Congress, Xiamen, 2022, pp. 3180-3185.

[13] J. Wang and J. Li, "Short-term photovoltaic power forecasting using the combined ksvd and lstm method," International Conference on Machine Learning, Cloud Computing, and Intelligent Mining (MLCCIM), Xiamen, China, 2022, pp. 134-138.

[14] N. Sergeev and P. Matrenin, "Improving accuracy of machine learning based short-term load forecasting models with correlation analysis and feature engineering," International Conference of Young Professionals in Electron Devices and Materials, Novosibirsk, 2023, pp. 1000-1004.

[15] T. Rasheed, A. R. Bhatti, M. Farhan, A. Rasool, and T. H. M. El-Fouly, "Improving the efficiency of deep learning models using supervised approach for load forecasting of electric vehicles," IEEE Access, vol. 11, pp. 91604-91619, 2023.

[16] M. Orlandi, M. Escudero-Casao, and G. Licini, "Nucleophilicity prediction via multivariate linear regression analysis," The Journal of Organic Chemistry, vol. 86, no. 4, pp. 3555-3564, 2021.

[17] Y. Tan, C. Li, Y. Huang, and J. Liu, "An Improved load derivative analysis method for short-term load forecasting," International Conference on Artificial Intelligence and Big Data (ICAIBD), Shanghai, 2021, pp. 256-261.

[18] M. Mohammadi, O. Ardakanian, and F. Karray, "A similar day algorithm for short-term load forecasting in smart grids," IEEE Transactions on Smart Grid, vol. 12, no. 3, pp. 1898-1908, May 2021.

[19] A. Claessens and D. Janssens, "Single and ensemble weather-based short-term load forecasting for belgium," Energies, vol. 13, no. 2, pp. 407, 2020.

[20] Y. Liu, Z. Zhou, C. Liu, J. Zhang, and Z. Wang, "Short-term load forecasting based on clustering and load profile shape similarity," Energies, vol. 15, no. 6, pp. 1404, 2022.

[21] E.S. Gardner and G. McKenzie, "Seasonal exponential smoothing and multiplicative Holt–Winters methods for forecasting," International Journal of Forecasting, vol. 37, no. 3, pp. 966-979, 2022.

[22] F. Qi, J. Zhang, and F. Ma, "Short-term load forecasting based on adaptive time series decomposition and load derivative analysis," Energies, vol. 14, no. 7, pp. 2020, 2023.

[23] A.T. Fard, M. Ghassemi, and H. Mohsenian-Rad, "Load profiling for short-term load forecasting using consumer energy usage statistics," IEEE Transactions on Smart Grid, vol. 11, no. 6, pp. 5206-5216, 2020.

[24] U. Vignesh and B.S. Chokkalingam, "EnC-SVMWEL: ensemble approach using cnn and svm weighted average ensemble learning for sugarcane leaf disease detection," International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, 2023, pp. 1663-1668.

[25] H. Hou, C. Liu, Q. Wang, B. Zhao, L.Q. Zhang, X.X. Wu, C.J. Xie, "Load forecasting combining phase space reconstruction and stacking ensemble learning," IEEE Transactions on Industry Applications, vol. 59, no. 2, pp. 2296-2304, March-April 2023.

[26] Y. Song, D. Tang, J. Yu, Z. Yu, and X. Li, "Short-term forecasting based on graph convolution networks and multiresolution convolution neural networks for wind power," IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1691-1702, Feb. 2023.

[27] Z. Pan, S. Zhou, H. Zou, C. Liu, M. Zang, and T. Liu, "Crmsnet: a deep learning model that uses convolution and residual multi-head self-attention block to predict rbps for rna sequence" Proteins: Structure, Function, and Genetics. vol. 9, no. 18, pp. 1032-1041, 2023.