Membrane Technology
ISSN (online): 1873-4049

Accurate Power Load Prediction Via a Hybrid Network Based on Automatic Feature Association

Zeyang Zhao*
Integrated Circuit College, Huazhong University of Science and Technology, Wuhan, China
**Corresponding Author.*

Abstract:
Amidst the rising need for sustainable energy solutions, precise forecasting of power load has emerged as a pivotal aspect of efficient energy management. However, achieving accurate load forecasts remains challenging, requiring a delicate balance between precision and computational efficiency. This study presents a novel approach to general power load forecasting, leveraging a hybrid network based on automatic feature association to enhance the ability of power load generation and predictions. The proposed network investigates the effects of various neural network interactions on real-world power load data within the network layer, continually adjusting parameters to assess their impact. Convolution based operation as well as an integration of attention and bidirectional long short-term memory (BiLSTM) network are employed to learn comprehensive features from power load data, serving as inputs for the network layers. By integrating the two learning-based structure, the hybrid network is trained hierarchically, utilizing multi-model fusion, interaction, and infiltration strategies. Experimental findings demonstrate that the proposed approach yields competitive performance compared to traditional single-model methodologies. This highlights the effectiveness of utilizing interconnected neural networks to improve the precision of load prediction, thereby promoting more efficient energy management strategies.

Keywords: load prediction, CNN, BiLSTM, attention mechanism, neural network.

INTRODUCTION

Electricity demand forecasting entails estimating the upcoming power usage or requirement in a particular area or network. Traditional methods of load forecasting often relied on simple statistical models, which struggled to capture the dynamic nature of modern electricity systems. Many works now explore sophisticated models that can effectively handle non-linear relationships, incorporate diverse data sources (e.g., weather data, historical consumption patterns, socio-economic indicators), and adapt to changing environmental and market conditions. This interdisciplinary approach seeks to enhance model precision, resilience, and scalability, thus improving power system efficiency and sustainability. This paper focuses on achieving accurate power load predictions [1]. Presently, time series forecasting is categorized into three main methods: statistical model-based [2-5], traditional ML (Machine Learning) [6,7], deep learning [8-11], and hybrid approaches [11-13]. Statistical prediction methods utilize mathematical models to describe and address problems, offering insight into the essence and rules governing them. Commonly used models like AR (Autoregressive), ARMA, and ARIMA with Bayesian estimation excel at modeling complex temporal data [14-16]; however, they struggle with highly intricate and precise time series data, failing to precisely track its evolution. Traditional ML methods rely on empirical or prior knowledge to build predictive models, e.g., LR (Logistic Regression), DT (Decision Trees), as shown by Chase et al. [17] in their application across meteorological contexts. These methods, along with RF (Random Forest) and SVM (Support Vector Machines), are extensively applied in time series forecasting [18]. Yet, data-driven models built without considering underlying mechanisms and expert knowledge exhibit limited monitoring and diagnostic capabilities, with applicability and robustness constraints, particularly for complex industrial systems. Deep learning, conversely, uncovers hierarchical patterns and intrinsic laws in sample data, extracting valuable insights for interpreting various data types (text, images, sound). Examples include Li et al. [19] employing DNNs (Deep Neural Networks) for a comprehensive forklift robot pallet picking solution. Another study by Li et al. [20] presents a deep learning method powered by internet big data to conduct real-time load surveys effectively and conveniently. Convolutional neural network can be applied to the problem in multiple computer vision fields, and what instances are in the prediction method. For example, Mukherjee et al. [21] proposed a computer vision system driven by multi-channel convolutional neural networks for identifying the types and ripening stages of banana fruits. Tong et al. [22] proposed CNN-based load forecasting prediction, time-initial convolution network based on long-head attention, with large receptive field and strong long-time feature extraction ability.

LSTM networks are engineered to manage data sequences across extended intervals while preserving vital contextual information. This feature renders them especially adept at tasks such as text generation, language translation, and voice recognition, where the sequence and context of the data are pivotal. Bi-directional LSTMs, or Bi-LSTMs, augment the functionality of conventional LSTMs by analyzing data in both forward and reverse order, thereby capitalizing on both preceding and subsequent context concurrently. The dual-direction processing allows Bi-LSTMs to better capture the nuances and dependencies within the data, making them more effective for complex sequence understanding tasks. Shetty et al. [23] illustrated an advanced use of LSTM alongside Vector Autoregression (VAR) models. Their hybrid approach combines the deep learning capabilities of LSTMs with the statistical strength of VAR models to predict critical metrics in cloud computing environments. It probably offers a sturdy platform for prediction that harnesses the advantages of both approaches to enhance the precision and dependability of forecasts. Such hybrid models are indicative of the ongoing evolution in machine learning, aiming to leverage multiple techniques for enhanced performance and applicability across various prediction fields.

Gated recurrent network or gated recurrent neural network includes some gating mechanisms, mainly composed of reset gate, update gate, used for feature extraction of temporal data, which is adept at capturing prolonged temporal dependencies within time-series based datasets. Aseeri et al. [24] proposed RNN-based load forecasting prediction which had achieved excellent performance in predicting daily electricity load in enterprise level measurement data. Bidirectional gated recurrent unit (BIGRU) is a multivariate time series prediction method based on bi-way gate recurrent unit (GRU). It integrates the bidirectional structure with a gating system to proficiently encapsulate the temporal dynamics within time-series information and the interplay among multiple variables.

The introduction of attention mechanism can improve the interpretability of neural networks and is usually applied into the neural network structure to enhance the expression ability of network learning features. Xu et al. [25] designed a novel learning-based framework that introduced an attention mechanism module to improve accuracy of learning-based VO without decreasing the generalization ability. Many works tend to use attention mechanism when combining the advantages of various word vectors. The study shows that the word vector weighted based on the attention mechanism achieves good results on the same model on various tasks [26].

Existing prediction technologies not only require high complexity, but also a high prediction accuracy. Many current forecasting models struggle to achieve high accuracy, particularly in capturing sudden changes or anomalies in load patterns. Complex models may provide better accuracy but can be challenging to interpret and understand, limiting their practical utility for decision-makers. Traditional forecasting methods may struggle to capture the dynamic and non-linear relationships between various factors influencing electricity consumption. By addressing these challenges and embracing emerging technologies, the future of load forecasting holds promise for more accurate, reliable, and actionable predictions, ultimately contributing to the efficiency of energy systems.

In this paper, we provide a prediction network for automatic extraction of load feature correlation, which fully considers the characteristics of temporal correlation of temporal data and the characteristics of trans-spatial temporal data, and learns the deep logical relationships hidden in the data to achieve accurate load prediction. The key contributions can be summarized as follows.

(1) The paper provides a new hybrid method for general power load prediction tasks, which elaborately incorporates CNN structure, BiLSTM layer, attention layer to achieve automatic and accurate load prediction.

(2) The method focusses on extracting key temporal features and provides them different weights assignment for the input sequential input data to locate the

corresponding feature association to facilitate the prediction.

(3) This work conducts comprehensive evaluations to verify the proposed network, which verify that the method can achieve a higher prediction accuracy.

THE PROPOSED LOAD PREDICTION NETWORK

Problem Statement

Stable load forecasting refers to the ability of a forecasting model to consistently provide reliable and accurate predictions of electricity consumption over a specified time horizon. A stable load forecasting model should exhibit resilience against sudden changes or anomalies in load patterns, ensuring that its predictions remain consistent and trustworthy under varying conditions. Stability implies a robust performance across different scenarios, including fluctuations in weather conditions, changes in economic activities, and other influencing factors. The stability of a forecasting model is crucial for its practical utility in energy management, resource allocation, and decision-making processes within power systems, contributing to the overall reliability and efficiency of the electricity grid. In this paper, for arbitrarily input power load data $\mathbf{X}= \{x_1,x_2 \ldots x_t\}$, we need to find out a proper learning-based network to achieve the prediction, to find a network model so that any **X** can output the prediction results of Y= $\{y_1, y_2 ... y_s\}$. The objective of this study is to construct an optimized hybrid architecture that can solve the problem of accurate prediction of data with different degrees of complexity.

Data Preprocessing

Data preprocessing in load forecasting involves several steps to prepare the data for analysis and modeling. For load data in this paper, the preprocessing steps include 1) abnormal data processing; 2) zero value or missing value processing. 3) Normalization treatment. For anomaly handling, we directly complement the outliers with a linear function, and complete the substitution with the following equation, $Ax_1+Bx_2+Cx_3=1$. According to the N data values, we conduct the zero value or missing value processing, close value replacement, planning, maximum programming function to achieve normalization processing. By performing these preprocessing steps, the data is ready for use in building and evaluating load forecasting models, helping to improve the accuracy and reliability of the forecasts.

Overview of the Hybrid Framework

The prediction network structure combining CNN-based operation with BiLSTM is a powerful architecture for sequence prediction tasks, including time series forecasting. In this hybrid network structure, the CNN layers are tasked with deriving layered features from the input sequence data. The convolutional layers apply filters across the input sequence, capturing patterns and local dependencies in the data. This helps in learning representations that are robust to variations and

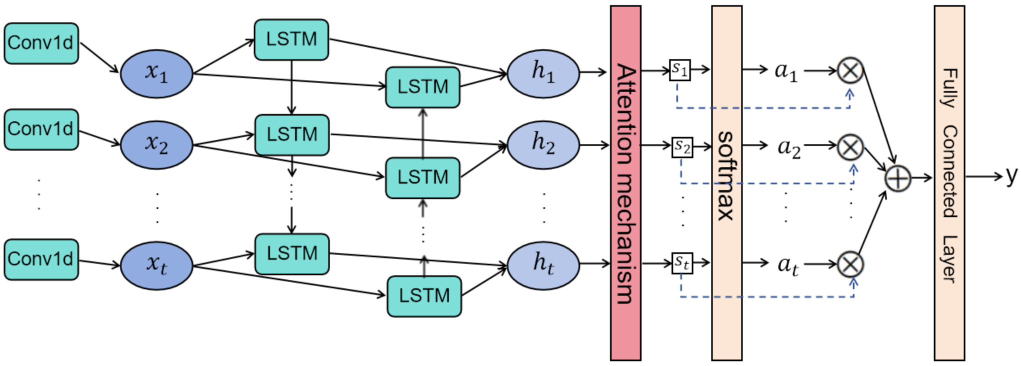


Figure 1. The framework of the proposed hybrid prediction network

noise in the input sequence. The yield of the CNN layers is subsequently channeled into the BiLSTM layers. Through concurrent handling of the input sequence in both forward and reverse orientations, BiLSTM networks can effectively capture the context and dependencies within the sequence. The combination of CNN and BiLSTM allows the network to learn both local and global features of the input sequence, enabling it to make accurate predictions while considering the temporal dynamics and dependencies present in the data. Figure 1 describes the framework of the proposed hybrid prediction network based on a clever combination of CNN feature extraction module and a BiLSTM module. The hybrid structure provides a strategy to promote the weight optimization of each related feature to obtain the final global linear model. The structure offers a powerful framework for sequence prediction tasks, providing flexibility, robustness, and the ability to capture complex patterns in sequential data.

CNN Module of the Proposed Hybrid Method

The Convolutional Neural Network (CNN) load forecasting module is a component of predictive models tailored specifically for forecasting electricity consumption or demand. CNNs are adept at extracting spatial patterns and features from input data, making them well-suited for tasks where the spatial relationships within the data are essential. In the context of load forecasting, the CNN module typically operates on input data such as historical load profiles or meteorological information. CNNs utilize convolutional layers to meticulously analyze input data, effectively identifying localized patterns and interdependencies within the sequential load data. This capability enables CNNs to extract critical features that significantly augment the model's ability to generate accurate results of next-step power load data. Within a comprehensive prediction framework, the CNN-based component often forms an integral part of an integrated system that may include other types of network structures, such as RNNs or LST and their variants. This amalgamation allows the proposed network to effectively capture both the spacial features and time-based dynamics among the input data, thereby enhancing the overall robustness and precision of the forecast. The integration of these diverse architectures facilitates a more holistic understanding of the data, leveraging the strengths of each component to produce forecasts that are not only accurate but also highly reliable. Figure 2 illustrates a simple workflow of CNN processing module, and Figures 3 and 4 illustrate the LSTM module's hidden layer and associated specifics, respectively. A distinctive feature of the CNN framework is the presence of convolutional and pooling layers, retained within the standard neural network layout. The mathematical expression for the convolution operation within the 1D convolutional layer is delineated as follows:

$$y_i = \sigma (y_{i-1} \otimes w_i + b_i) \tag{1}$$

where w_i denotes the weight of the convolution kernel at layer i, engaging in a convolution operation with the preceding layer's image, y_{i-1}. The result is enhanced by adding the bias vector b_i of layer i, ultimately yielding the feature map y_i through the non-linear activation function. The subsampling or pooling layer typically reduces the feature map dimensions post-convolution, according to specific rules. Its primary functions are to diminish the feature map's size and somewhat preserve the scale-invariance of the features. Consider y_i in a subsampling layer. The fully connected layers of the CNN utilize the derived features to classify and generate a probability distribution based on the input

$$y_i = subsampling(y_{i-1}) \tag{2}$$

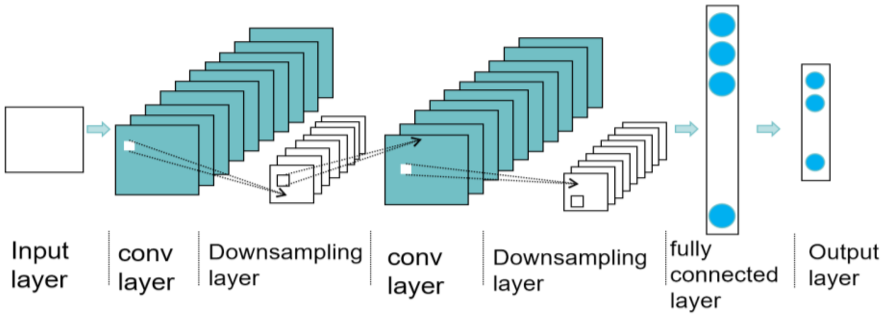


Figure 2. The workflow of CNN processing module

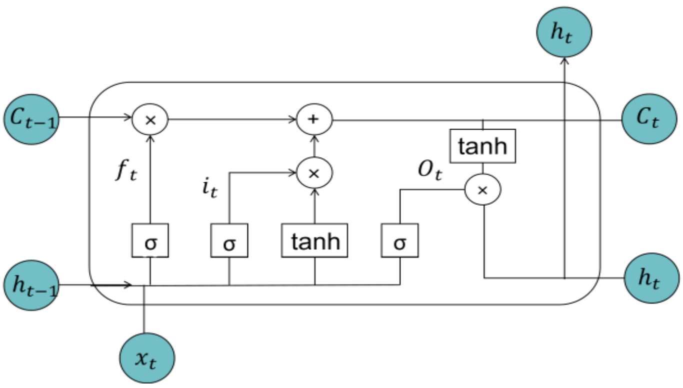


Figure 3. The hidden layer of LSTM module

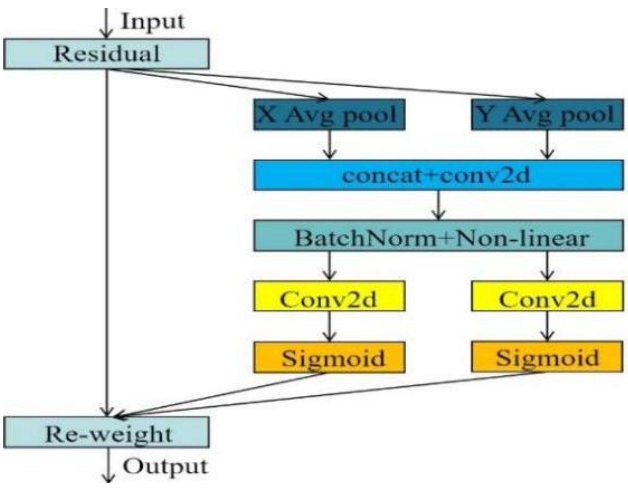


Figure 4. The details hidden layer of LSTM module

The fully connected layers of the CNN utilize the derived features to classify and generate a probability distribution based on the input.

$$G(i)=P(L=l_i|y_o,(w, b)) \tag{3}$$

The goal during the training phase for the CNN is to minimize its loss function, centering on lessening the difference between the true inputs and the estimated outputs, termed "residuals." Frequently utilized loss functions include the Mean Squared Error (MSE) and the negative log-likelihood (NLL).

$$MSE(w, b)=\frac{1}{|G|}\sum_{i=1}^{|G|}((G(i)-\widehat{G}(i))^2 \tag{4}$$

$$NLL(w, b)=-\sum_{i=1}^{|G|} log\, G(i) \tag{5}$$

To counteract overfitting, the final loss equation often incorporates an L2 norm to control the overfitting through the regularization parameter λ:

$$E(w,b) = L(W, b)+\frac{\lambda}{2}w^T w \tag{6}$$

The module for load prediction within the CNN framework elevates the prognostic prowess of models by facilitating their efficient acquisition and utilization of spatial characteristics in electrical usage data, thereby appending greater veracity and dependability to the forecasts of energy demand.

BiLSTM Module of the Hybrid Method

y_t indicates t Hidden state at the moment. Contextual representation of the current moment under the influence of the global information at moment 0-t. X_t : Input vector at time t. C_t: The memory of moment t. Basically, the global information at time 0-t. w_i represents weight matrix from that input gate, $W_c$represents the weight matrix in the cell states, b_i, b_c is the corresponding bias terms, and $tanh$() is the hyperbolic tangent function.[y_{t-1}, X_t] means that two vectors are connected to a longer vector, and w_i is composed of w_{ih} and w_{ix}, with the following formula: forgetting gate: control how much past C_{t-1} memory is retained. Select the information in the cell, which is depicted as follow.

$$f_t=\sigma\ (\ w_i*[y_{t-1},X_t]+b_f) \tag{7}$$

As in input gate, we use the input data from tanh ($\widetilde{C_t}$), and select a part of the memory to add this part to the new memory, to complete the memory update.

$$i_t=\sigma\ (\ w_i*[y_{t-1},X_t]+b_i) \tag{8}$$

where i_t and f_t have similar effects, both involving selective retrieval of a memory and updating to a new memory. The difference is that F_t is the historical memory selection. I_t is the current information selection, the selected information becomes a part of the memory. In the output gate, we determine the ones in the C^t the output current state. Before the output selection for C_t, C_t was first scaled by tanh transformation.

$$O_t=\sigma\ (w_o*[y_{t-1}, X_t]+b_o) \tag{9}$$

For candidate memory C_t, it is acquired by nonlinear transformations of two inputs.

$$C_t= tanh\, (w_c[y_{t-1},X_t]+b_c) \tag{10}$$

in that current unit, as input for the next time step, saving all 1-t memories.

$$C_t=F_t*C_{t-1}+I_t*\widetilde{C_t} \tag{11}$$

In that output of the hidden layer, the last layer of each LSTM time step will be output as the final result.

$$H_t= tanh\, (C_t)*O_t \tag{12}$$

Attention Mechanism

To encapsulate the focus across the dimensions of data width and height, and to encode the precise positional data, the incoming feature map is partitioned along two axes: vertical and horizontal. This process results in distinct feature maps for each dimension. The mathematical representation of this segmentation is delineated hereafter.

$$S_c^h(h)=\frac{1}{W}\sum_{0\leq i<W} x_c\,(h, i) \tag{13}$$

$$S_c^w(h)=\frac{1}{G}\sum_{0\leq j<G} x_c\,(j, w) \tag{14}$$

Afterward, the comprehensive spatial dimensions of the feature map, encompassing width and height, are aggregated from both dimensions. This compiled data is then processed through a shared 1×1 convolutional kernel, effectively compressing the feature dimensions to C/r. Post convolution, the feature F_1 undergoes batch normalization and is subjected to a Sigmoid activation function, resulting in a feature representation with dimensions 1×(W+H) ×C/r, as depicted in the following equation.

$$f=\delta(F_1([s^h, s^w]) \tag{15}$$

Thereafter, that f is processed through convolution with a 1×1 filter in alignment with its initial dimensions of height and breadth, resulting in feature maps F_h and F_w that preserve the identical channel count. Subsequent to the application of the Sigmoid activation, the attention coefficients g^h for the vertical dimension and g^w for the horizontal dimension are derived. The following mathematical expression delineates this procedure.

$$g^h=\sigma(F_h(f^h)) \tag{16}$$

$$g^w=\sigma(F_w(f^w)) \tag{17}$$

Following that aforementioned computations, a vertical attention coefficient and the horizontal attention coefficient are yielded. Subsequently, by conducting a weighted multiplication operation on the initial feature map, the ultimate feature map, imbued with attentional weights for both the vertical and horizontal axes, is produced. The pertinent equation is displayed hereinafter.

$$y_c(i,j)= x_c(i,j)\times g_c^h(i)\times g_c^w(j) \tag{18}$$

Data Training

During the training process, the network is manually set five parameters initialization (LOOK back=12, epoch=19, batch size=9, validation split=0.3, verbose=2), 60% of the data in the data set is trained, 40% is used for verification. Each time the network completes the set epochs reaches the set value or takes time to reach our threshold. When the last batch of Training data is not enough for one batch-size, the network automatically kicks off not enough batch to complete the training task. Four metrics were used to validate the training effect.

SIMULATION AND EVALUATION

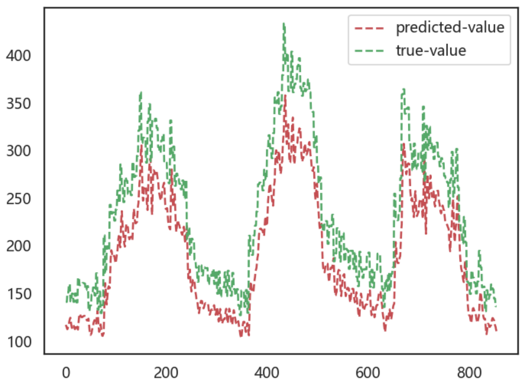
Experimental Design

To assess the efficacy of the proposed hybrid approach, within the experimental design segment, we apply the proposed network and CNN, LSTM, GRU to complete each prediction task and record the corresponding MAPE, RMSE, MAE and R2 after finishing the load prediction. A lower MAPE, RMSE, or MAE represents the more accurate predicted value and a higher R2 indicates a better fitting effect of the network.

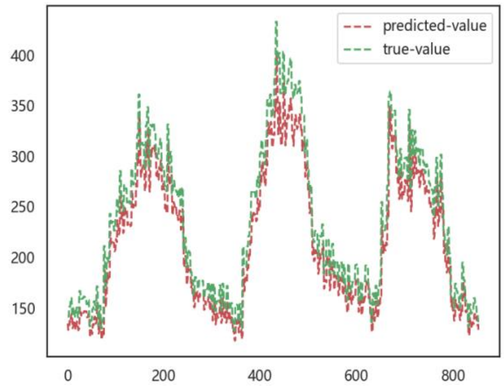
Effect Evaluation

Owing to the constraint of page numbers, Figure 5 provides a representative compared prediction result of the compared methods with real load data from a company. The horizontal axis indicates the sampling locations, whereas the vertical axis illustrates the predicted power load quantities. Our analysis reveals a close correspondence between the forecasted values generated by our methodology and the actual load values, manifesting as discernible and consistent curves. This high degree of accuracy and efficacy in load forecasting is attributable to the effective integration of multi-module structures within our approach. Upon meticulous examination of the depicted figures, it becomes apparent that our methodology consistently yields forecasted values that closely mirror the actual power load values, especially at the peak and inflection points, predictions are more accurate compared to other methods. The distinct patterns exhibited by the curves underscore the precision and reliability of our forecasting model. This graphical representation serves to reinforce the robustness and efficacy of the multi-module structures embedded within our approach. The successful amalgamation of diverse layer structures has not only contributed to enhanced objective metrics but has also translated into visually precise and dependable general load forecasting outcomes. Based on our findings, it is evident that the predicted values generated by our method are closely consistent with the actual load values, displaying a clear and unique curve upon careful examination. The good fit at the inflection point demonstrates that our method can better predict the changes in temporal data.

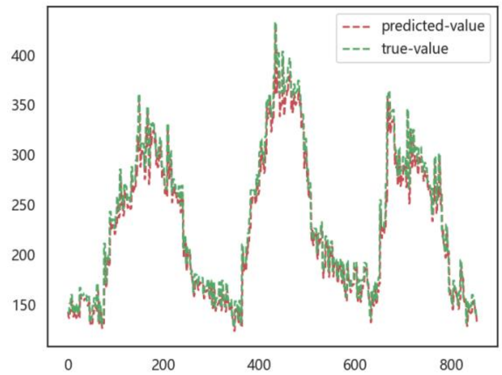
The mean objective scores for different competitive load prediction techniques are detailed in Table 1, including RMSE, MAPE, MAE, and R^2, based on a representative dataset. Our method exhibits significantly superior performance across all metrics compared to the competing methods. Notably, there is a substantial decrease in RMSE, MAPE, MAE, and an enhancement in the coefficient of determination (R^2) with the application of our forecasting method. This improvement signifies a more accurate prediction of photovoltaic output, highlighting the robustness of our approach across diverse evaluation criteria.



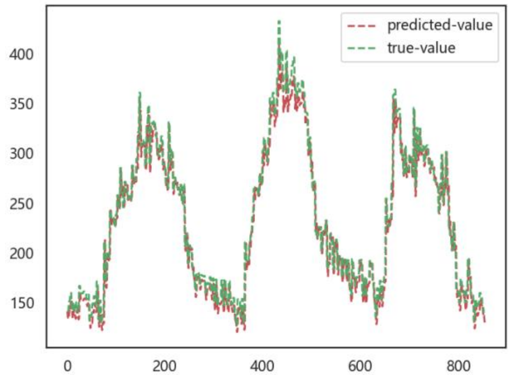
(a) Prediction sample from GRU method



(b) Prediction sample from LSTM method



(c) Prediction sample from BiLSTM method



(d) Prediction sample from the proposed method

Figure 5. A representative power load forecasting results with different compared prediction methods

Table 1. Load forecasting performance using different compared networks

	MAPE	RMSE	MAE	R^2
CNN	1.407	0.335	0.057	0.752
LSTM	1.174	0.571	0.198	0.701
BiLSTM	1.167	0.570	0.198	0.689
Ours	0.834	0.285	0.043	0.874

Moreover, the consistent outperformance of our method in comparison to alternative approaches underscores its reliability and effectiveness in navigating the complexities of load forecasting. A thorough examination of these objective values reinforce confidence in the proposed model's capability to deliver precise and reliable predictions, emphasizing its potential for practical applications within the domain of solar power prediction. In practical applications, algorithms can be deployed on power systems to accurately predict short-term power loads, providing effective decision-making for general power planning and scheduling in different kinds of energy systems.

CONCLUSION

The study provides a hybrid network based on automatic feature association, which leverages an integration of advanced network architectures to peruse high accurate power load prediction. Specifically, the method integrates a CNN layer for comprehensive feature extraction from diverse load data, a BiLSTM component to encapsulate bidirectional sequential data insights, along with an attention schema for proficient weighting distribution. The synergistic integration of these layers results in a significant enhancement of load forecasting accuracy. This combination strategy demonstrates proficiency in handling complex input load data, leading to predictions that closely approximate actual values. Comprehensive simulations results corroborate the competitive performance of the proposed network in general power load forecasting, demonstrating competitive outcomes in both objective and subjective quality metrics.

REFRENCES

[1] Dai, Y., Zhou, Q., Leng, et al. Improving the bi-lstm model with xgboost and attention mechanism: a combined approach for short-term power load prediction. Applied Soft Computing, 2022, 130(6): 109632-109644.

[2] Su-Kui, X.U., Wei, S., Zhen-Hua, et al. A statistical modeling approach to automatic evaluation of mandarin pronunciation (corpus-based data-driven speech research in asia). Journal of the Phonetic Society of Japan, 2015, 19 (1):44-52.

[3] Nemani, R., Moses, G. J., Alenezi, et al. Statistical data mining with slime mould optimization for intelligent rainfall classification. Tech Science Press, 2023, 047(010): 919-935.

[4] Sedehi, O., Papadimitriou, C., Katafygiotis, et al. Hierarchical bayesian uncertainty quantification of finite element models using modal statistical information. Mechanical Systems & Signal Processing, 2022, 179:109296.1-109296.30.

[5] Reyes, D. M. A., De Souza, R. M. C. R., De Oliveira, et al. A three-stage approach for modeling multiple time series applied to symbolic quartile data. Expert Systems with Application (Jan.), 2022, 187(Jan.):115884.1-115884.12.

[6] Maleki, S., Maleki-Zanjani, B., Kohnehrouz, et al. Machine learning tools to identify key minerals of hoagland solution for healthy kiwiberry micropropagated plant hardening. Acta Horticulturae, 2022, (1322):607-627.

[7] Raza, S.A., Govindaluri, S.M., Bhutta, et al. Research themes in machine learning applications in supply chain management using bibliometric analysis tools. Benchmarking: An International Journal, 2023, 30(3):834-867.

[8] Guo, P., Rao, S., Hao, et al. Fault diagnosis of a semi-batch crystallization process through deep learning method. Computers & Chemical Engineering: An International Journal of Computer Applications in Chemical Engineering, 2022, 164, 107807, 2-16.

[9] Verma, A., Jain, P., Kumar, et al. An effective depression diagnostic system using speech signal analysis through deep learning methods. International Journal of Artificial Intelligence Tools: Architectures, Languages, Algorithms, 2023, 32(2):2340004-2340021.

[10] Xu, G., Lin, K., Li, et al. Saf-net: A spatio-temporal deep learning method for typhoon intensity prediction. Pattern Recognition Letters (Mar.), 2022, 155:121-127.

[11] Toledano-Lopez, O.G., Gonzalez, H., Madera, et al. A hybrid method based on estimation of distribution algorithms to train convolutional neural networks for text categorization. Pattern Recognition Letters, 2022, 160(Aug.):105-111.

[12] Sornalakshmi, M., Balamurali, S., Venkatesulu, et al. Hybrid method for mining rules based on enhanced apriori algorithm with sequential minimal optimization in healthcare industry (retraction of vol 34, pg 10597, 2022). Neural Computing & Applications, 2023, 35(5):4193-4193.

[13] Akyol, S. A new hybrid method based on aquila optimizer and tangent search algorithm for global optimization. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(6): 8045-8065, 2023.

[14] Jeong, D., Park, C., Ko, et al. Short-term electric load forecasting for buildings using logistic mixture vector autoregressive model with curve registration-sciencedirect. Applied Energy, 2020, 282(Jan.15 Pt. B):116249.1-116249.14.

[15] Khashei, M., & Chahkoutahi, F. A comprehensive low-risk and cost parallel hybrid method for electricity load forecasting. Computers & Industrial Engineering, 2021, 155(3): 107182.1.

[16] Sales, L. O. F., Alencar, A. P., Ho, et al. The berg generalized autoregressive moving average model for count time series. Computers & Industrial Engineering, 2022, 168: 108104-108117.

[17] Chase, R. J., Harrison, D. R., Burke, et al. A machine learning tutorial for operational meteorology. part i: traditional machine learning. Weather and Forecasting, 2022, 37(8):1509-1529.

[18] Anmala, J., & Turuganti, V. Comparison of the performance of decision tree (dt) algorithms and extreme learning machine (elm) model in the prediction of water quality of the upper green river watershed. Water environment research: a research publication of the Water Environment Federation, 2021, 93(11), 2360-2373.1

[19] Li, Y., Ding, G., Li, C., et al. A systematic strategy of pallet identification and picking based on deep learning techniques. Industrial Robot, 2023, 50(2):353-365.

[20] Li, Y., & Chen, J. Structural live load surveys by deep learning. Smart Structures and Systems, 2022, 30(2):145-157.

[21] Mukherjee, G., Chatterjee, A., Tudu, et al. A multi-channel convolutional neural network driven computer vision system towards identification of species and maturity stage of banana fruits: case studies with martaman and singapuri banana. Int. J. Comput. Intell. Stud., 2022, 11(1): 1-23.

[22] Tong, C., Zhang, L., Li, et al. Temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting. IET Generation, Transmission & Distribution (8), 2022, 16:1680-1696.

[23] Shetty, J., Cottur, K., Shobha, et al. A weighted ensemble of var and lstm for multivariate forecasting of cloud resource usage. Journal of Advances in Information Technology, 2023, 14(2):264-270.

[24] Aseeri, A. O. Effective rnn-based forecasting methodology design for improving short-term power load forecasts: application to large-scale power-grid time series. Journal of Computational Science, 2023, 68(Apr.):1.1-1.13

[25] Xu, Y., Qin, H., Huang, et al. An end-to-end deep context gate convolutional visual odometry system based on lightweight attention mechanism. Industrial Robot, 2022, 49(1):47-53.

[26] Zheng, Q., Chen, Z., Liu, et al. Msranet: learning discriminative embeddings for speaker verification via channel and spatial attention mechanism in alterable scenarios. Expert Systems with Applications, 2023, 217(May): 119511.1-119511.10.