Power Load Forecasting System Based on Deep Hybrid Learning Model

Shangyichen Dong
No.85 Middle School of Xi'an, Xi'an, 710000, China

Abstract. Accurate prediction of electricity demand is crucial to ensure the stability and dependability of local power grids. Numerous scholars have put forth comprehensive prediction systems; However, the majority of these models fail to capture the inherent global features present within the data. This study introduces a novel integrated prediction system that leverages the synergistic capabilities of Time Convolutional Network (TCN) and Long Short-Term Memory (LSTM) architectures, aiming to enhance the accuracy of short-term electric load forecasting. The initial step is to establish separate prediction models for LSTM and TCN, with a focus on electric load data. After combining the output results of these models, the reciprocal of the error square ratio was used as a weighting factor. By using this approach, the LSTM-TCN model for combined prediction is created. This research paper performs an exhaustive examination of the case study by employing authentic data sourced from the Australian Energy Management Authority. The study's findings support that the LSTM-TCN model outperforms both single prediction models and traditional network models in terms of performance. The results indicate that the LSTM-TCN model exhibits greater accuracy in predicting short-term energy demand.

Keywords: short-term electricity demand prediction, Long Short-Term Memory (LSTM) neural network, Time Convolutional Network (TCN), reciprocal of the squared error ratio.

1. Introduction

Short-term load forecasting (STLF) primarily pertains to the prediction of electricity demand in the immediate future, encompassing forecasts for the upcoming day or week. It provides reference for tasks such as hydroelectric dispatching, unit start-up and shutdown, as well as coordination between water and fire. It is an essential groundwork for the daily operation of the power grid. Reliable STLF is needed for the safety and stability of the power system, facilitate efficient start-up and shutdown procedures of generating units, enable effective scheduling and planning activities, and support informed trading strategies. Hence, the prediction of short-term load has emerged as a significant field of study within the operational domain of contemporary power systems [1]. Over the past few decades, numerous methodologies for load forecasting have been proposed, broadly categorized into traditional forecasting models, single network models, and hybrid network models.

Conventional models encompass techniques such as trend extrapolation, time series analysis, regression analysis, and other established methodologies. Krymova [2] proposed a trend estimation method that is globally applicable and a straightforward short-term forecasting strategy derived from this method. The proposed approach relies on minimal information pertaining to the progression of the pandemic and leverages robust seasonal trend decomposition techniques. Li [3] developed a profit and loss model for biomass energy potential by investigating the developmental trajectory of the overall population and electricity consumption within a specific research area. Ensafi [4] employed classical time series forecasting techniques, including Seasonal Autoregressive Integrated Moving Average (SARIMA) and Triple Exponential Smoothing, to predict furniture sales. Madhukumar [5] discovered that the Gaussian Process Regression (GPR) model series exhibited superior performance in load forecasting. These models, being probability-based and relying on non-parametric kernels, demonstrated the most favorable predictive capabilities. Du [6] proposed a Bayesian Optimization-based Dynamic Ensemble (BODE), which overcomes the limitations of single-model approaches and provides a dynamic ensemble forecasting combination with time-varying base patterns for STLF. Zhang [7] introduced Crossformer, a Transformer-based model that considers not only time series but also cross-dimensional dependencies among variables. These methods are easy to model and have good performance in stable load forecasting. However, they are not suitable for load sequences with strong randomness.

Currently, load forecasting methods based on artificial intelligence are widely used, primarily using single neural networks and support vector machines as typical methods. Hu [8] introduced the Conformal Time Convolutional Quantile Regression Network (CTCQRN), a new approach for interval prediction. This method integrates the Conformal Quantile Regression (CQR) algorithm with the Time Convolutional Network (TCN) technique, eliminating the need for distribution assumptions. Lin Jun [9] and colleagues proposed a Long Short-Term Memory (LSTM) network that incorporates a two-stage attention mechanism for short-term regional load probability forecasting. Tang [10] and co-authors presented a short-term load forecasting model that utilizes a Time Convolutional Network (TCN) with channel and time attention mechanisms (AM). This model effectively captures the nonlinear correlation between weather factors and load. Deng [11] and colleagues introduced a quantitative combined load forecasting model (QCLF) designed to handle highly random and uncertain load-related data. Ghenai [12] and co-authors developed an adaptive neural fuzzy inference system (ANFIS) specifically designed to deliver highly accurate and very short-term energy consumption predictions for educational buildings. Niu [13] and colleagues presented a novel short-term multi-energy load forecasting method that utilizes the CNN-BiGRU model. The model incorporates an attention mechanism for optimization purposes. FazlaliPisheh [14] and co-authors introduced an innovative univariate deep LSTM-based Stacked Autoencoder (DLSTM-SAE) model for short-term load forecasting. This model is enhanced with a multi-stage attention mechanism (MSAM), which includes an input attention mechanism and multiple time attention mechanisms incorporated during the training process. Yuan [15] and colleagues proposed a short-term overall daily prediction model that utilizes Variational Mode Decomposition (VMD) and two sets of multi-step strategies. This model aims to improve the accuracy of daily predictions in the short term. The complexity of power system structures has led to a growing recognition of load nonlinearity and uncertainty. Expressing the relationship between loads and their influencing factors in a single model has become challenging due to these factors' increasing complexity.

To enhance predictive performance, numerous researchers have introduced hybrid models and combination prediction models. For instance, Fan et al. [16] proposed the RF-MGF-RSM model as a hybrid approach for short-term load forecasting. This model effectively predicts electricity load by combining multiple techniques. Indeed, Hu et al. [17] proposed a fully integrated approach for load forecasting. Their method incorporates several techniques, including the use of the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Improved Grasshopper Optimization Algorithm (IGOA), and Long Short-Term Memory (LSTM) network. This integrated approach aims to improve the accuracy and adaptability of load forecasting. Certainly, Alotaibi [18] employed various techniques for short-term load forecasting. Their study utilized Deep Neural Network (DNN), Artificial Neural Network (ANN) based on Multilayer Perceptron, and Decision Tree-based prediction (DR). By employing these different models, they aimed to enhance the accuracy and reliability of short-term load forecasting. Indeed, Javed et al. [19] introduced a novel two-level Encoder-Decoder (ED) network for load forecasting. Their proposed architecture consists of two stages. The first stage utilizes a short-receptive-field based dilated causal convolution (SRDCC) network, which helps capture local dependencies and patterns in the load data. The second stage incorporates a bidirectional Long Short-Term Memory (BiLSTM) network, which enables the model to capture both past and future context information. This two-level ED network aims to improve the generalization ability and prediction accuracy of load forecasting models. In their study, Chen et al. [20] introduced a hybrid algorithm, namely CEEMDAN-IGWO-GRU (CIG), which combines complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), gated recurrent unit (GRU), and an improved grey wolf optimizer (IGWO). This hybrid model aims to enhance the performance of load forecasting by leveraging the synergistic effects of these techniques. Despite the remarkable achievements of previous hybrid models in load prediction, the DNN-ANN-DR model proposed in reference [18] falls short in capturing the intricate details of the data, thereby impacting its overall performance. This study presents an innovative prediction model that integrates the time convolutional neural network (TCN) and the long short-term memory network (LSTM). Initially, LSTM is employed to forecast the electricity load, while subsequently TCN is utilized for load prediction. Finally, a weighted combination approach is employed to merge the predictions from both models, yielding the ultimate outcome. In essence, this paper contributes by:

1. Our hybrid LSTM-TCN model exploits the strengths of both: LSTM's long-term memory for sequential data and TCN's local pattern detection for superior prediction of complex time series. This robust model increases accuracy and stability, and demonstrates improved fault tolerance and noise handling.

2. The combined prediction model ensures reliable power system operations, offering superior accuracy compared to standalone LSTM or TCN models. This crucially aids energy supply and societal stability, enabling informed decisions, risk reduction, and efficient energy utilization for a smoothly running power system.

3. Our hybrid model exhibits excellent generalization and predictive prowess across multiple real datasets. Extensive testing confirms its outstanding performance not only on training data but also on unseen data, showcasing strong generalization abilities, thus bolstering its reliability and real-world applicability.

2. Model Framework

In this section, a comprehensive overview is presented regarding various individual prediction models employed in this study. These models encompass the Long Short-Term Memory Network (LSTM), the Time Convolutional Network (TCN), weighted methodologies based on the reciprocal of squared error, as well as the metrics utilized for evaluating the performance of these prediction models.

2.1 Long Short-Term Memory Network (LSTM)

The Long Short-Term Memory Network (LSTM) is an enhanced version of the Recurrent Neural Network (RNN) designed to tackle the challenge of gradient explosion encountered when processing lengthy sequences. By designing effective memory and gating control mechanisms, LSTM effectively mitigates the vanishing or exploding gradient problem, enabling more effective learning and capturing of long-term relative dependencies in sequential data. LSTM, through its structural advantages, is able to learn long-range dependencies between data, thereby improving prediction accuracy. As the most widely-used and successful RNN structure, it is particularly suitable for short-term electricity load prediction.

LSTM controls the information flow in the network by introducing input gate (referred to as 'in'), forget gate (referred to as 'fn'), and output gate (referred to as 'on'). Although the LSTM model has powerful nonlinear mapping ability in processing time series data, it requires the explicit construction of relationships between features; otherwise, it is difficult to fully utilize the effective information between discontinuous features. Therefore, considering combining it with feature mining networks to improve the mining ability of load features. The three gates are shown in Equation (1):

$$\begin{aligned} i_n &= \sigma(W_i \cdot [h_n - 1, x_n] + b_i) \\ f_n &= \sigma(W_f \cdot [h_n - 1, x_n] + b_f) \\ o_n &= \sigma(W_o \cdot [h_n - 1, x_n] + b_o) \end{aligned} \tag{1}$$

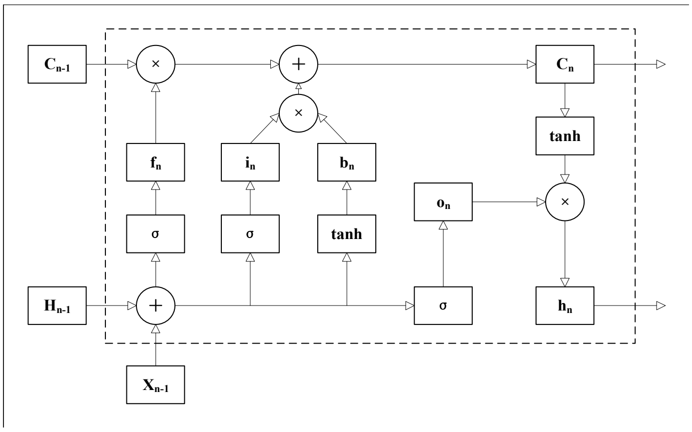

Figure 1 illustrates the structure of the LSTM memory unit, which primarily consists of three memory units: the forget gate, the input gate, and the output gate. The tanh function is considered a superior activation function.

2.2 Temporal Convolutional Network (TCN)

The Temporal Convolutional Network (TCN) is a neural network architecture used for processing time series data. Compared to traditional Convolutional Neural Networks (CNN), TCN is able to more effectively extract features from sequential data. TCN is derived from CNN by incorporating multiple residual units. By introducing residual connections, TCN can propagate information across layers, thereby enhancing its learning capability and further improving model performance.

The one-dimensional dilated causal convolution plays a vital role in TCN's residual units. By adjusting the convolutional coefficients to control the sampling interval of the input, TCN can have a longer receptive field with fewer layers. By increasing the filter size (n) and the dilation factor (d), the Temporal Convolutional Network (TCN) enables the top layer output to encompass a broader spectrum of input information. Moreover, by concurrently applying identical filters within each layer, the overall computational efficiency of the model can be enhanced. The structure of the dilated causal convolution is depicted in Figure 2, showcasing a filter size (n) of 2 and dilation factors (d) of [1, 2, 4]. Following the inclusion of dilated convolutions, the output (yt) at time (t) can acquire information from inputs (xt-7、xt-6、...、xt).

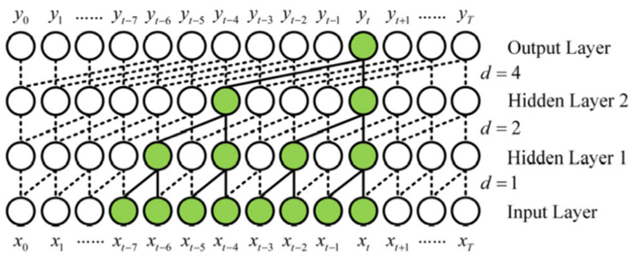

Figure 2 illustrates the structure of the expanded causal convolution, taken from reference [21].

Compared to CNN, the causal convolutions in TCN are unidirectional, preventing the model from losing historical information while avoiding the influence of future information. This ensures that TCN becomes a more strict time-constrained model.

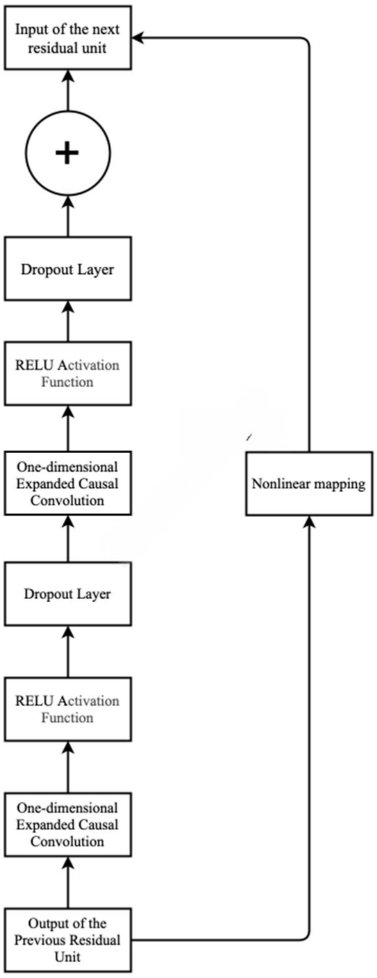

Figure 3 illustrates the working principle of TCN, where the non-linear mapping is implemented using the Sigmoid function.

2.3 Hybrid Prediction Model

Bates and Granger (1969) presented the combined prediction approach. The idea is that different prediction methods may give different results, therefore, each prediction model should be given different weights in order to obtain a better overall prediction model. Figure 4 illustrates the flowchart of the combined prediction model that integrates LSTM and TCN.

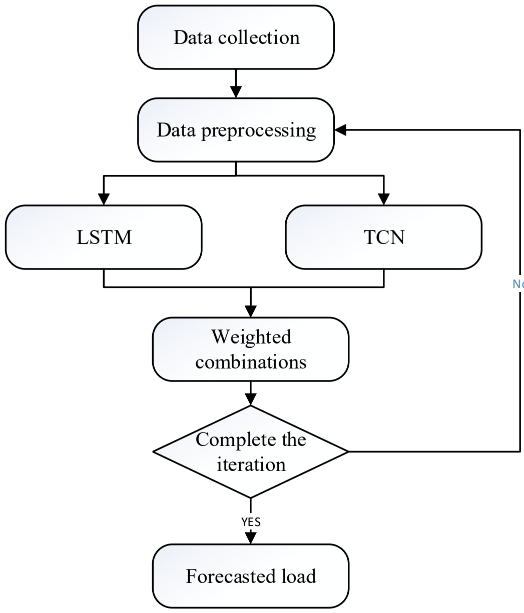

Figure 4 The flowchart of the combined prediction model, which includes LSTM model and TCN model.

This article employs a residual-based optimal weighting approach. When a single model exhibits a smaller prediction error, it will be assigned a relatively higher weight in the combined model. The calculation process of this method, represented by formula (2)(3)(4), involves the simultaneous prediction of m time points by two individual models:

$$w_i = \frac{1}{s_i \cdot \sum_{i=1}^{2} \frac{1}{s_i}} \tag{2}$$

$$s_i = \sum_{i=1}^{n} r_{ip}^2 \tag{3}$$

$$r_{ip} = a_{ip} - a_p \tag{4}$$

Here, the variables are defined as follows: a_{ip} reflects the i method final output prediction value at time-step p, a_p reflects the corresopnding actual value at time t, r_{ip} reflects the residual in i method at time-step p, s_i reflects the total squared residuals in i method at n time-step points. The weight coefficient of the i method, denoted as w_i, is calculated as the inverse of the square of the prediction error of each of the individual models.

2.4 Model Evaluation

To assess the performance of each model in this study, the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) metrics are employed. Their definitions are as follows.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{R_i - F_i}{R_r}\right| \times 100\% \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_i - F_i)^2} \tag{6}$$

In this context, the variables are defined as follows: R_i reflects the real load value, F_i reflects the predicted load value, and n reflects the number of time series points.

3. Experiments

3.1 Datasets

The dataset utilized in this research was acquired from the Australian Energy Market Operator (AEMO). The dataset consists of power load data gathered at 30-minute intervals in Tasmania from

January to March 2014. The dataset exhibits a mean value of 1044.22 kWh, accompanied by a standard deviation of 88.322. To better understand the patterns in the dataset, we selected a one-week time span and visualized the data, as shown in Figure 6.

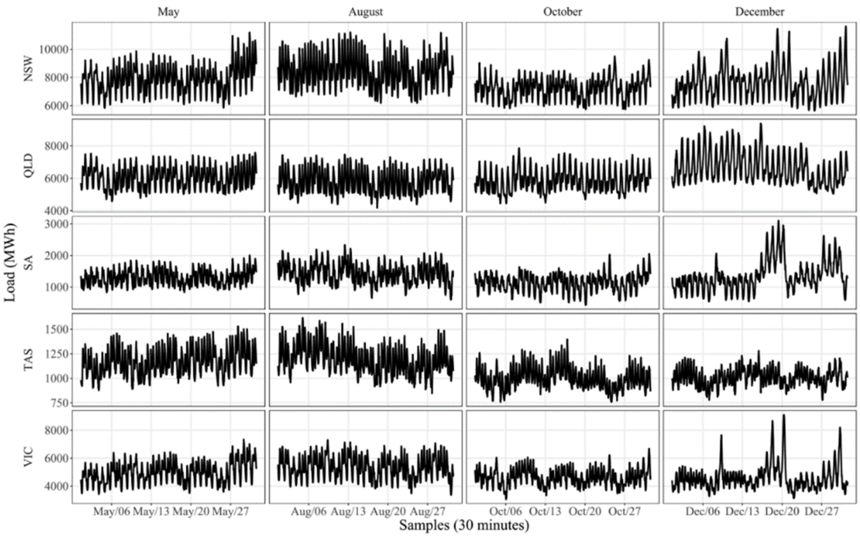

Figure 5 Description of dataset [22]: AEMO load data for each region and month (season) in 2019.

By examining Figure 5, it becomes evident that the load variation exhibits periodicity, with a consistent daily trend. Specifically, the time intervals corresponding to peak and valley load curves remain relatively stable. Analysis of the box plot reveals a substantial presence of outliers in the load data during the fourth week of these three months, indicating a relatively high level of fluctuation in the power grid load during that specific week.

Table 1 Statistical Indicators of Each Dataset

Region	Month (Season)	Statistical measures				
		Minimum	Median	Average	Maximum	Std
NSW	May (Autumn)	5818.00	7867.69	7869.72	11 209.02	976.73
	August (Winter)	6176.72	8288.68	8401.73	11 221.11	1129.53
	October (Spring)	5729.21	7224.37	7203.56	9514.79	751.43
	December (Summer)	5634.88	7548.65	7631.23	11 638.99	1152.38
QLD	May (Autumn)	4578.55	6011.36	6013.01	7586.47	678.14
	August (Winter)	4173.36	5618.42	5754.84	7478.16	728.86
	October (Spring)	4442.02	5738.80	5835.85	7875.28	691.82
	December (Summer)	5044.05	6413.47	6652.01	9395.25	968.13
SA	May (Autumn)	828.51	1248.49	1285.57	2009.36	238.97
	August (Winter)	599.59	1383.53	1409.71	2334.66	317.89
	October (Spring)	431.78	1144.57	1126.31	2055.42	268.64
	December (Summer)	539.10	1263.25	1348.72	3107.74	481.98
TAS	May (Autumn)	877.24	1161.79	1169.00	1530.71	134.92
	August (Winter)	846.31	1171.83	1192.33	1618.84	146.61
	October (Spring)	757.20	988.65	1002.36	1398.31	110.90
	December (Summer)	778.20	1005.26	1009.56	1282.07	88.32
VIC	May (Autumn)	3420.13	4964.24	4966.09	7358.98	759.43
	August (Winter)	3366.14	5321.15	5327.78	7316.11	813.30
	October (Spring)	3058.20	4587.61	4622.78	6684.43	600.23
	December (Summer)	3136.49	4402.68	4597.31	9113.24	916.84

3.2 Results of the Model

For this study, the LSTM structure used 200 neurons in its hidden layer, whereas the TCN model utilized a kernel size of two for convolution and the dilation factor of 2. The original dataset was partitioned into three monthly groups, and within each group, the data sets were further split into training and testing sets in a ratio of 7:3. The testing set was utilized to compare the performance of SARIMA, DNN-ANN-DR, and the LSTM-TCN models, which are commonly employed in this study. Both SARIMA and DNN-ANN-DR models were configured with 200 hidden layers. The comparative outcomes are presented in Table 2. The best performance is indicated by bold text with a gray background, while the second best performance is indicated by a gray background only.

Table 2 Comparison of the performance of different models on various datasets

Data set	SARIMA		DNN-ANN-DR		LSTM-TCN	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Jan.	1.538	21.864	1.774	24.906	1.377	19.3693
Feb.	1.653	23.267	1.917	26.130	1.449	20.6649
Mar.	1.819	24.962	2.410	33.828	1.340	18.3162
All	1.773	24.842	1.878	26.357	**1.442**	**20.5893**

By referring to Table 1, it is evident that the LSTM-TCN model exhibits a lower MAPE value of 1.422 for the same dataset compared to the traditional SARIMA and DNN-ANN-DR models. Across various datasets, the LSTM-TCN model employed in this study consistently demonstrates a lower MAPE value compared to other traditional models.

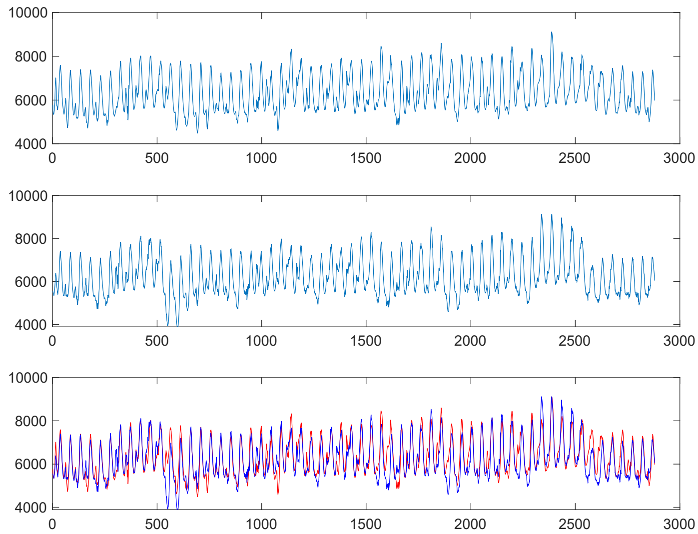

Figure 6 displays the prediction results of the LSTM model. The top plot shows the original electric load values, the middle plot shows the predicted electric load values, and the bottom plot is the visualization of the overlay between the original and predicted values.

3.3 Melting Experiment

To improve the validation of the model's performance, the experiment evaluates the predictive effects of each designed model and then compares the hybrid methods performance with them. When operating independently, the designed LSTM model is configured with an input feature dimension of 2 and 200 neurons in the hidden layer. In the standalone TCN model, the convolution kernel size is set to 2, and the convolution dilation factor is set to 2. The prediction performance of various models after 1000 training iterations is presented in Table 3.

Table 3 compares the results of independent models and the combined models on variable datasets.

Data set	LSTM		TCN		LSTM-TCN	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Jan.	2.873	40.5343	1.499	20.7559	1.377	19.3693
Feb.	2.107	28.9641	1.631	22.5112	1.449	20.6649
Mar.	2.389	32.9566	1.622	21.1821	1.340	18.3162
All	2.007	28.4452	1.515	21.0238	**1.442**	**20.5893**

Based on the findings from Table 3, it can be observed that the hybrid forecasting model, which combines LSTM and TCN as designed in this paper, demonstrates lowest RMSE value and MAPE value across diverse datasets and months. In terms of the overall dataset, the hybrid LSTM-TCN model demonstrates a 1.442% MAPE value and a 20.5893 RMSE value. The prediction accuracy has been enhanced when comparing the LSTM-TCN hybrid model with independently designed TCN and LSTM networks.

4. Conclusion

Various fields have widely applied LSTM, TCN, and other deep learning models. This study presents a novel integrated model for short-term load forecasting in the power system, utilizing a combination of LSTM and TCN architectures. The experimental results of the LSTM-TCN model, traditional models, and individual models are compared using real-world data obtained by Australian Energy Management Agency. The comparison leads to the following conclusions:

1. The combination of LSTM and TCN models with weighted reciprocal error squares in the proposed LSTM-TCN model corrects the high-error time series data in the individual models, reducing the overall errors. The experimental findings demonstrate a significant improvement in the accuracy of short-term power load prediction achieved by the LSTM-TCN model.

2. In future study, it is worth investigating the introduction of attention mechanisms into TCN. Additionally, natural factors such as temperature, carbon dioxide concentration and seasonal influence can be included in the analysis of short-term load forecasting. Moreover, by integrating specific scenarios, the applicability of this model can be extended to various industries including agriculture, commerce, and other sectors, enabling the extraction of valuable insights.

Reference

[1] Kong X, Wang Z, Xiao F, et al. Power load forecasting method based on demand response deviation correction[J]. International Journal of Electrical Power & Energy Systems, 2023, 148: 109013.

[2] Krymova E, Béjar B, Thanou D, et al. Trend estimation and short-term forecasting of COVID-19 cases and deaths worldwide[J]. Proceedings of the National Academy of Sciences, 2022, 119(32): e2112656119.

[3] Li C, Zhang Y, Ma K, et al. The effect of different prediction methods of population in villages and towns on assessing the suitability of biomass energy development[J]. Journal of Cleaner Production, 2022, 363: 132382.

[4] Ensafi Y, Amin S H, Zhang G, et al. Time-series forecasting of seasonal items sales using machine learning–A comparative analysis[J]. International Journal of Information Management Data Insights, 2022, 2(1): 100058.

[5] Madhukumar M, Sebastian A, Liang X, et al. Regression model-based short-term load forecasting for university campus load[J]. IEEE Access, 2022, 10: 8891-8905.

[6] Du L, Gao R, Suganthan P N, et al. Bayesian optimization based dynamic ensemble for time series forecasting[J]. Information Sciences, 2022, 591: 155-175.

[7] Zhang Y, Yan J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//The Eleventh International Conference on Learning Representations. 2022.

[8] Hu J, Luo Q, Tang J, et al. Conformalized temporal convolutional quantile regression networks for wind power interval forecasting[J]. Energy, 2022, 248: 123497.

[9] Lin J, Ma J, Zhu J, et al. Short-term load forecasting based on LSTM networks considering attention mechanism[J]. International Journal of Electrical Power & Energy Systems, 2022, 137: 107818.

[10] Tang X, Chen H, Xiang W, et al. Short-term load forecasting using channel and temporal attention based temporal convolutional network[J]. Electric Power Systems Research, 2022, 205: 107761.

[11] Deng S, Chen F, Wu D, et al. Quantitative combination load forecasting model based on forecasting error optimization[J]. Computers and Electrical Engineering, 2022, 101: 108125.

[12] Ghenai C, Al-Mufti O A A, Al-Isawi O A M, et al. Short-term building electrical load forecasting using adaptive neuro-fuzzy inference system (ANFIS)[J]. Journal of Building Engineering, 2022, 52: 104323.

[13] Niu D, Yu M, Sun L, et al. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism[J]. Applied Energy, 2022, 313: 118801.

[14] Fazlipour Z, Mashhour E, Joorabian M. A deep model for short-term load forecasting applying a stacked autoencoder based on LSTM supported by a multi-stage attention mechanism[J]. Applied Energy, 2022, 327: 120063.

[15] Yuan F, Che J. An ensemble multi-step M-RMLSSVR model based on VMD and two-group strategy for day-ahead short-term load forecasting[J]. Knowledge-Based Systems, 2022, 252: 109440.

[16] Fan G F, Zhang L Z, Yu M, et al. Applications of random forest in multivariable response surface for short-term load forecasting[J]. International Journal of Electrical Power & Energy Systems, 2022, 139: 108073.

[17] Hu H, Xia X, Luo Y, et al. Development and application of an evolutionary deep learning framework of LSTM based on improved grasshopper optimization algorithm for short-term load forecasting[J]. Journal of Building Engineering, 2022, 57: 104975.

[18] Alotaibi M A. Machine Learning Approach for Short-Term Load Forecasting Using Deep Neural Network[J]. Energies, 2022, 15(17): 6261.

[19] Javed U, Ijaz K, Jawad M, et al. A novel short receptive field based dilated causal convolutional network integrated with Bidirectional LSTM for short-term load forecasting[J]. Expert Systems with Applications, 2022, 205: 117689.

[20] Chen Z, Jin T, Zheng X, et al. An innovative method-based CEEMDAN–IGWO–GRU hybrid algorithm for short-term load forecasting[J]. Electrical Engineering, 2022, 104(5): 3137-3156.

[21] Wang Y, Chen J, Chen X, et al. Short-term load forecasting for industrial customers based on TCN-LightGBM[J]. IEEE Transactions on Power Systems, 2020, 36(3): 1984-1997.

[22] Ribeiro M H D M, da Silva R G, Ribeiro G T, et al. Cooperative ensemble learning model improves electric short-term load forecasting[J]. Chaos, Solitons & Fractals, 2023, 166: 112982.