Received: April 3, 2024. Revised: May 9, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Effective Method for Forecasting Electrical Load Data Using Temporal Convolution Attention-based Long Short-Term Memory

Sridhar Hassan Siddappa[1]*

[1]*Department of Electrical and Electronics Engineering, Siddaganga Institute of Technology, Tumkur, India*
* Corresponding author's Email: sri054@sit.ac.in

Abstract: Electric load forecasting plays a significant role in electric power systems for numerous applications, in terms of specific time horizons like Demand Side Management (DSM), grid stability, optimal operations, and Long-term strategic planning. However, inaccurate forecasts minimize the power supply safety, affecting the social and economic activities, security, and national defense. Therefore, Temporal Convolutional Attention-based Long Short-Term Memory (TCA-LSTM) is proposed for accurately forecasting electric load using Deep Learning (DL). By including an attention mechanism in an LSTM approach, the proposed technique focuses more on parameters with greater weights. Initially, the power load dataset is employed to evaluate the proposed approach. The obtained data is then pre-processed by utilizing the min-max normalization to reduce the impact of outliers. Then, the Fast Fourier Transform (FFT) technique is performed to extract the dominant amplitude-frequency. At last, the TCA-LSTM is used to accurately forecast the electric load. The proposed TCA-LSTM achieves a better Mean Square Error (MSE) of 0.002, 0.0074, and 1.5047 respectively in Godishala, Warangal, and Vijayawada, compared to the existing techniques namely, Regression Tree (RT), Principle Component Analysis with Recurrent Neural Network (PCA-RNN), and Artificial Neural Network (ANN).

Keywords: Temporal convolution attention-based long short-term memory, Deep learning, Electric load, Fast fourier transform, Vijayawada.

1. Introduction

Load Forecasting (LF) plays a significant role in power system operation and scheduling. This procedure forecasts the demand of future load by utilizing both available and historical data. There are numerous benefits of accurate LF in the power market, like achieving an equilibrium among load and supply demand, electricity market profits, and transactions in economical energy to enable appropriate decisions in future generations. It enhances the power system stability, prediction of electricity price, and so on [1]. LF is split into three types, and in that, Short-Term LF (STLF) is the primary work of dispatching the power grid, and it becomes significant in the system in electrical power [2, 3]. STLF has a pivotal role in advanced power grid context systems [4]. Using huge amounts of data produced by the infrastructure of a smart grid permits accurate energy estimation demand, which contributes to increased energy distribution management, security, and economy [5]. An inappropriate electrical STLF impacts irregular flow in power and system congestion that minimizes the protection and security of electrical power systems, leading to unbalanced generation planning [6]. Electrical LF is crucial in constructing and enhancing the efficiency of power systems as it facilitates economic and reliable planning, operation, and control of power systems [7]. Accurate load forecasts are essential for supporting energy trading in electricity markets. When accurate load forecast depends on accurate historical information, data integrity attacks can contaminate the historical information [8]. Electricity load demand forecasting is influenced by a time-series of Gross Domestic Product (GDP) [9].

The Electric Power Research Institute (EPRI) determines a smart grid that combines different forms of technology within every aspect of the electrical pipeline from generation to consumption. The purpose of technology integration is to increase markets, reduce environmental burden, strengthen service and reliability when reducing costs and increase efficiency [10]. A smart grid is a new kind of power system that is developed in recent years and is employed by power companies because of its accuracy in power load prediction [11]. Accurate forecasting of electrical load assists in formulating scheduling and power generation operational strategy which enables service providers to forecast the amount of electrical energy needed by the grid users [12]. LF not only generates essential data for policymakers to ensure effective planning capacity, but also assists in achieving optimal scheduling by utilizing economic dispatch and unit commitment [13]. Moreover, optimal dispatch in a grid-associated system is attained when the decision is to employ energy storage, while electricity cost is high during peak hours which results in minimized cost of energy and decreased peak electrical demand on the grid levels [14]. Various approaches are established to forecast electrical LF by using DL because of their ability to manage large-scale data, capture complex temporal patterns, and adapt to different input modalities [15]. However, inaccurate forecasts minimize the power supply safety and affect the social and economic activities, security, and national defense. Therefore, TCA-LSTM is proposed for accurately forecasting electric load using DL. TCA-LSTM improves the efficiency and reliability of distribution systems, making stable power supply secure against different impacts on social function and national security.

The main contributions of this research are as follows:

- The min-max normalization is used in the pre-processing phase which is capable of converting the present data range which assists in reducing the outlier's impact and increasing the model's performance and stability.
- FFT effectively extracts the dominant amplitude frequency from electric load data that helps in determining periodic patterns like daily fluctuations. This approach is significant in time-series analysis which assists in the development of robust and accurate forecasting models.
- The TCA-LSTM is performed to forecast electric load data accurately. TCA-LSTM integrates TCN with attention to increase the LSTM performance by effectively capturing long-term dependencies and focusing on salient temporal features in input sequences and variations in the patterns of electricity consumption.

The rest of the portions of the manuscript are structured as follows: Section 2 represents a literature survey. Section 3 determines the proposed methodology for electrical LF, while Section 4 indicates the result of the study, and Section 5 contains the conclusion.

2. Literature survey

The related work about electrical LF forecasting is described along with their advantages and limitations.

Veeramsetty [16] suggested an RT to forecast the active load power. Initially, the pre-processing was performed based on missing values, normalization, outliers, and data split. This approach was employed for hour-ahead and day-ahead forecasting, and the load at a specific time of the day was forecasted depending on load at the same time. RT approach achieves less error in predicting electrical LF with the help of real-time data integration. However, RT faces challenges with non-linear relationships due to the partitioning data based on single feature thresholds, leading to inaccurate predictions.

Veeramsetty [17] integrated Principle Component Analysis (PCA) with Recurrent Neural Network RNN to enhance the ability of hourly load forecasting on a substation of electric power. The pre-processing approach was utilized to recognize outliers and to determine the skewness of data. PCA extracted the appropriate features from the provided information. RNN was employed to predict the LF effectively. This approach accurately forecasted load by a minimized input of data dimensionality which minimized the entire computational effect. However, PCA did not accurately forecast time-series data which limited RNN's capability to learn intricate dynamic loads.

Rao [18] developed an ANN for efficient load-day ahead forecasting of load demand in cluster microgrids. A zero-mean normalization approach was utilized to normalize data on temperature and load variables in the pre-processing phase. LM technique was employed to select the most appropriate features and ANN was established to forecast electrical LF. The developed approach enhanced the numerical stability and convergence rate. However, this approach faced difficulties in managing non-linear and dynamic load behavior, which affected the forecasting accuracy in intricate microgrid environments.

Ajitha [19] introduced a Recurrent Neural Network-LSTM (RNN-LSTM) to forecast electrical load for residential sector by utilizing real-time load data gathered from local utilities. Then, the gathered data was scaled and normalized by employing two scaling approaches such as standard scalar and min-max scaler. Finally, RNN-LSTM was used to predict electrical LF. This approach enhanced the forecasting accuracy with past scarce data. Nevertheless, the introduced approach faced challenges in capturing effective long-range dependencies in load patterns, particularly in scenarios with rapidly changing or highly irregular load dynamics.

Alrasheedi and Almalaq [20] presented a hybrid approach depending on Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU) and CNN-RNN to increase the results in Saudi smart grid load forecasting for enhancing problem-associated features. This accurately forecasted the complicated consumption of power to establish a reliable prediction approach and acquired knowledge of relationships among different attributes and features in Saudi smart grid. This optimal forecasting performance assisted in optimizing the network requirements and scalability. Nevertheless, inaccurate forecasts minimize power supply safety, affecting social and economic activities, security, and national defense.

From the overall analysis, it is shown that the existing techniques faced challenges with non-linear relationships due to partitioning data on single features and do not effectively capture temporal dependencies which leads to inaccurate forecasting. It minimizes the power supply safety and affects social and economic activities, security, and national defense. To overcome these issues, the TCA-LSTM is proposed to accurately forecast electrical load data.

3. Proposed methodology

The TCA-LSTM is proposed to forecast the electric load data accurately. Initially, the power load dataset is obtained from Godishala (village) and Warangal, Telangana state, India, and wind and solar factors from Vijayawada city in A.P state, India. The min-max normalization is used to convert the present data range. FFT is performed for feature extraction and A-LSTM is established to forecast electric load data. Fig. 1 indicates a block diagram for the proposed technique.

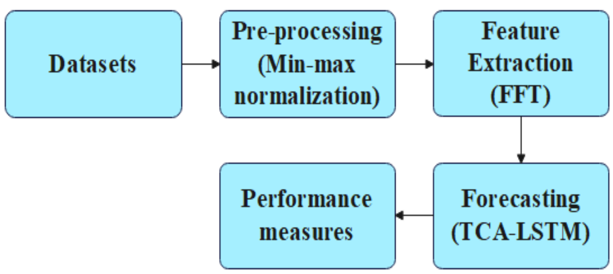

Figure. 1 Block diagram for proposed technique

3.1 Dataset

In this research, a power load dataset is established in two places for electric load forecasting in the day ahead. The data is gathered from a 33/11 kV substation distribution in Godishala (village), Telangana state, India [21] which has 8712 samples, six input, and 1 output feature. Another place is the Kakatiya University in Warangal, Telangana state, India [22] which contains the overall 2184 samples, nine input, and 1 output feature. Next, the data is gathered from wind and solar factors from Vijayawada city in A.P state, India with 44665 location ID, -16.65-degree latitude, and -80.65-degree longitude [23]. These collected data are fed into a pre-processing stage to handle missing values by imputation with median values and outliers.

3.2 Pre-processing

Once the data is gathered, the pre-processing stage is performed to transform the data values of certain datasets. There is a huge contrast among the minimum and maximum values of a dataset. Hence, normalizing the data reduces the algorithm's complexity for further processing. Normalization enables appropriate benefits for classification approaches associated with neural networks. The normalization function [24] depends on data scaling that contains the min-max technique which is capable of converting the present range of data in [-1, 1] and [0, 1] intervals. When compared to other techniques like Z-score, min-max normalization has higher capability to scale data among fixed range which generates relationship among features when preventing outliers in electrical load forecasting. The min-max normalization is expressed in Eq. (1).

$$P = \frac{(x - x_{min})(max - \min)}{(x_{max} - x_{min}) + min} \tag{1}$$

Where, max, min represents specified input variable range, x_{min} and x_{max} indicate the initial variable range of input value, and p determines converted input value. Then, the final pre-processed data are passed through a feature extraction process.

3.3 Feature extraction

The FFT is used to extract the features from normalized data by converting the time-domain signal into its frequency-domain data which makes the identification of the periodic patterns for electric load forecasting. In comparison to other conventional methods like Wavelet transform, FFT effectively captures the period patterns and components of dominant frequency from time-series data which generates the cyclic behavior identification like daily fluctuations. While data availability is too large for processing, feature extraction is employed to reduce the dimensions when maintaining the properties of data. There are different approaches to extracting features like frequency, time, and time-frequency domain, and so on that are utilized to minimize the data dimension. Here, feature extraction in the frequency domain is utilized FFT which is employed to convert the received signals into a frequency domain. Then, the first and second dominant amplitude frequencies are extracted. Two detectors are employed and two characteristics are extracted from each detector. Hence, four characteristics are established for training the neural networks which is expressed in Eq. (2).

$$Y\ (k) = \sum_{J=1}^{n} x(J) w_n^{(y-1)(k-1)} \tag{2}$$

Where, $Y\ (k) = FFT\ (X)$ and $w_n = e^{(-2\pi i)/n}$ represents one of n unity roots. Then, the extracted features are fed into the forecasting model to forecast electric load.

3.4 Forecasting

The extracted dominant frequency amplitudes are passed as input to a neural network like the TCA-LSTM approach for forecasting the electric load data. The hyperparameters chosen for TCA-LSTM are default values having a grid search with a 32-batch size, Adam optimizer, initial learning rate of 0.001 to 0.01, and dropout regularization technique. These are specifically selected to optimize the performance of the model by balancing the computational efficiency with learning capacity, improving generalization capability, and reducing overfitting for accurate electric load forecasting. It integrates LSTM cells with an attention technique that enables the model to focus dynamically on appropriate temporal features during forecasting in electric load. LSTM [25] is appropriate to process and forecast significant events with long intervals in time series. It has three control gate units namely, input gate i_t, output gate o_t, and forget gate f_t. LSTM has numerous LSTM cells that generate the data over time. To remember dependence of long-term, input and forget gates are the LSTM keys. Input gate evaluates how much data about present network state is required to be saved to internal state that is formulated in Eq. (3).

$$i_t = \sigma(U_i h_{t-1} + W_t x_t + b_i) \tag{3}$$

Where, σ represents logistic function, X_t and h_{t-1} indicates outcome of memory block at $t-1$ time, and t determines input vector time, W_i and U_i refers to the input gate's weight matrix, and b_i represents input gate's bias term. Forget gate shows how much data from past is to be eliminated which is expressed in Eq. (4).

$$f_t = \sigma(U_f h_{t-1} + W_f x_t + b_f) \tag{4}$$

Where, W_f and U_f indicate the forget gate's weight matrix and b_f illustrates the forget gate's bias term. Output gate refers to how much data internal state is required to output present moment's external state which is determined in Eq. (5).

$$o_t = \sigma(U_o h_{t-1} + W_o x_t + b_o) \tag{5}$$

Where, W_o and U_o represents output gate's weight matrix and b_o indicates output gate's bias term. Initially, the last moment's external state h_{t-1} and input vector at present moment x_t are utilized to compute 3 gates and candidate state $\widetilde{C}_t$. Then, forget gate f_t and input gate i_t are combined to update the present moment's internal state C_t. At last, internal state data is transferred to external state $h_t$depending on output gate o_t.

3.4.1. Temporal convolution attention-LSTM

LSTM efficiently solves the issue of gradient explosion. In the training process, over-fitting causes a test to fail in forecasting an electric load. Therefore, a dropout layer is added for optimizing LSTM to solve this issue in electric load forecasting due to the capability to regularize neural networks by dropping random units during training, which produce robustness against overfitting without the requirement of extra data or early stopping. LSTM sets the neural network's input vector to a fixed dimension that has a good effect while managing by a low dimension. While an input parameter's dimension is huge, it affects the model's performance because of the dimension explosion issue. To better focus on influential parameters, this research employs TCA-LSTM for forecasting electric load. The attention technique is accomplished by retaining an LSTM encoder's intermediate output for input sequence, after which model training is used for

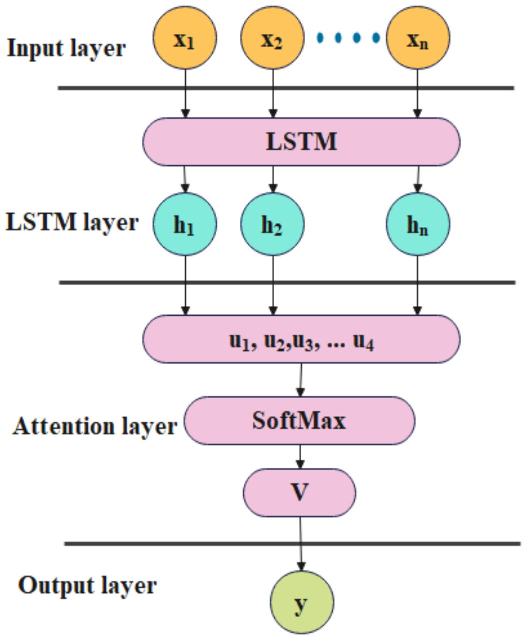

Figure. 2 Structure of Attention-based LSTM

learning these inputs selectively and relating the sequence of output with them. Most of the attention techniques depend on the Encoder-Decoder approach. The encoder procedure of the original codec approach provides an intermediate vector in the Seq2Seq that is employed to store the original sequence data. However, the vector length is fixed, and while the original input sequence is long, this vector manages limited data that restricts the model's understanding capability. The attention technique is used to break the original codec model's constraints on the fixed vectors. Fig. 2 represents the structure of attention-based LSTM.

The attention technique is primarily established in the following phase. The output $[h_1, h_2, h_3, \dots, h_n]$ in LSTM is transmitted nonlinearly to acquire $[u_1, u_2, u_3, \dots, u_n]$. In the forecasting procedure, certain operating parameters highly influence the electric load data. Hence, these parameters are required to be provided more significant weight. An attention technique will generate an attention weight matrix $[\propto_1, \propto_2, \propto_3, \dots, \propto_n]$ that indicates the significance of each intermediate state. At last, a weighted sum is established for input parameter while obtaining weight encoding vector V. An output y is acquired by decoding based on the encoding vector V which is expressed in Eqs. (5)-(7).

$$u_k = \tanh(W_k h_k + b_k) \quad (5)$$

$$\propto_k = \frac{\exp\,(u_k^T u_s)}{\sum_{k=1}^{n} \exp(u_k^T u_s)} \quad (6)$$

$$V = \sum_K^n \propto_k h_k \quad (7)$$

Where, W_k indicates weight matrics, b_k represents offset quantity, $\propto_k$ refers to the normalized attention weight, and u_s determines the randomly initialized time series attention matrix.

TCA-LSTM is a mechanism that integrates TCN with attention mechanism to increase the LSTM performance by generating the model to focus on significant temporal features within the input sequences. Temporal convolutional layers in TCN are established to capture both global and local temporal dependencies within the input sequences. The convolutional operation scans over temporal dimension that extracts features at various time scales. The attention mechanism allows the model to weight the significance of each time steps while processing sequential data, which increases the model's interpretability and captures the effective temporal dependencies in electric load forecasting. Consider X indicates input sequence to self-attention layer with $T \times D$ dimensions where T represents sequence length and D denotes feature dimension. The self-attention mechanism computes attention weights $\propto_i$ for every time step i which is expressed in Eq. (8)

$$\propto_i = softmax\left(W_q x_i^T\right). W_k x_i \quad (8)$$

Where W_q and W_k represents learnable parameters that transform the X input sequences into query and key vectors. The weighted sum of input sequences depending on attention weight is formulated in Eq. (9)

$$Z = \sum_{i=1}^{T} \propto_i . W_v x_i \quad (9)$$

Where Z determines self-attention output, W_v indicates another learnable parameter employed to transform the input sequence into value vectors. Then, the self-attention output Z is combined with LSTM hidden state h_t at each time step t which is expressed in Eq. (10)

$$H_{new} = [LSTM(X), Z] \quad (10)$$

This integrated representation includes both LSTM's sequential capability processing and self-attention mechanism ability to focus more on appropriate temporal features. It effectively captures long-term dependencies and focus on salient temporal features in input sequences and variations in the patterns of electricity consumption which increase the forecasting performance by influential

Table 1. Notation description

Symbols	Notation Description
max, min	Specified input variable range
x_{min} and x_{max}	Initial variable input value range
p	Converted input value
$Y(k) = FFT(X)$ and $w_n = e^{(-2\pi i)/n}$	One of n roots unity
σ	Logistic function
X_t and h_{t-1}	Outcome of memory block at $t-1$ time
t	Input vector time
W_i and U_i	Weight matrix of input gate
b_i	Input gate's bias term
W_f and U_f	Weight matrix of forget gate
b_f	Forget gate's bias term
W_o and U_o	Weight matrix of output gate
b_o	Output gate's bias term
h_{t-1}	External state of last moment
x_t	Input vector at present moment
$\widetilde{C}_t$	Candidate state
W_k	Weight matrics
b_k	Offset quantity
$\propto_k$	Normalized attention weight
u_s	Randomly initialized time series attention matrix
X	Input sequence to self-attention layer
T	Sequence length
D	Feature dimension
$\propto_i$	Attention weight
W_q and W_k	Learnable parameters
Z	Self-attention output
W_v	Another learnable parameter employed to transform the input sequence into value vectors

sequence factors in the prediction process. Table 1 indicates the Notation description.

4. Results

The proposed TCA-LSTM approach is simulated by using Python environment with intel i9 processor, and 128 GB RAM, 22 GB GPU, and Windows 10 operating system. The performance of TCA-LSTM is assessed in terms of error metrics Mean Average Error (MAE), MSE, Root Mean Square Error (RMSE), formulated in Eqs. (11)-(13), respectively.

$$MAE = \frac{1}{m}\sum_{\tau=1}^{m} |\lambda_\tau - \hat{\lambda}_\tau| \quad (11)$$

$$MSE = \frac{1}{m}\sum_{\tau=1}^{m} (\lambda_\tau - \widehat{\lambda_\tau})^2 \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{\tau=1}^{m}(\lambda_\tau - \widehat{\lambda_\tau})^2}{m}} \quad (13)$$

Where, m indicates the number of data points, λ_τ represents actual values, $\hat{\lambda}_\tau$ determines forecasted values, and $\overline{\lambda_\tau}$ illustrates the mean values.

4.1 Performance analysis

A qualitative and quantitative analysis of A-LSTM is presented in Tables 2 to 4. Table 2 represents the evaluation of forecasting performance in Godishala village. The Recurrent Neural Network (RNN), Deep Neural Network (DNN), CNN, and LSTM are compared with the proposed TCA-LSTM. Fig. 3 presents a graphical representation of forecasting performance in Godishala village. The obtained results show that the TCA-LSTM achieves better MSE of 0.002 when compared to RNN, DNN, CNN, and LSTM, respectively.

Table 3 shows the evaluation of forecasting performance in Warangal. The RNN, DNN, CNN, and LSTM are the existing methods employed for comparison with the TCA-LSTM approach. Fig. 4 represents the graphical representation of forecasting performance in Warangal. The acquired outcomes determine that the TCA-LSTM achieves better MSE of 0.0096 in contrast to the existing techniques.

Table 4 indicates the evaluation of forecasting performance in Vijayawada city. The existing techniques like RNN, DNN, CNN, and LSTM are

Table 2. Evaluation of forecasting performance in Godishala village

Methods	MSE	MAE	RMSE
RNN	0.009	0.064	0.089
DNN	0.007	0.057	0.075
CNN	0.006	0.051	0.062
LSTM	0.004	0.043	0.057
TCA-LSTM	0.002	0.035	0.052

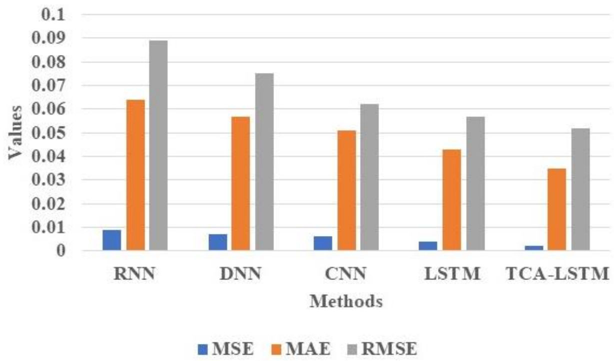

Figure. 3 Graphical representation of forecasting performance in Godishala village

Table 3. Evaluation of forecasting performance in Warangal

Methods	MSE	MAE	RMSE
RNN	0.0354	0.098	0.074
DNN	0.0268	0.087	0.064
CNN	0.0098	0.066	0.052
LSTM	0.0082	0.058	0.035
TCA-LSTM	0.0074	0.047	0.029

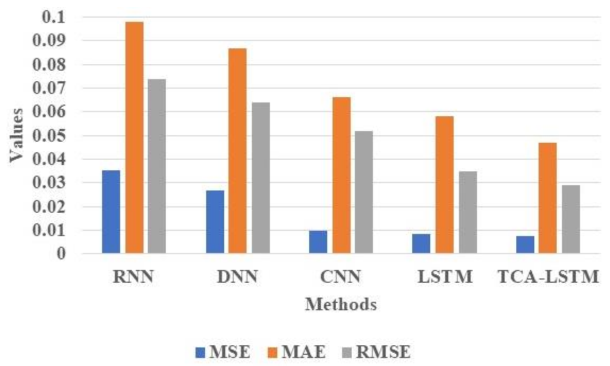

Figure. 4 Graphical representation of forecasting performance in Warangal

Table 4. Evaluation of forecasting performance in Vijayawada city

Methods	MSE	MAE	RMSE
RNN	1.6258	150.47	430.14
DNN	1.5984	147.65	429.35
CNN	1.5735	165.34	406.27
LSTM	1.4232	133.48	401.59
TCA-LSTM	1.5047	101.05	396.24

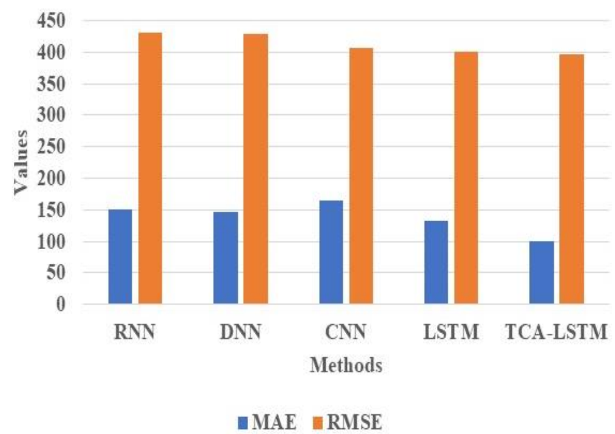

Figure. 5 Graphical representation of forecasting performance in Vijayawada city

compared with TCA-LSTM. Fig. 5 shows the graphical representation of forecasting performance in Vijayawada city. The TCA-LSTM with initial learning rate (0.001 to 0.01) focus more on appropriate temporal features which effectively captures the dependencies of long term while avoiding redundancy or data loss, and therefore generates less errors with enhanced interoperability compared to RNN, DNN, CNN, and LSTM. The obtained outcomes show that the TCA-LSTM achieves better MSE of 1.5047 when compared to the existing techniques.

4.2 Comparative analysis

The existing methods like RT [16], RDM [17], and LM-ANN [18] are used to compare with TCA-LSTM approach. Table 5 displays a comparative analysis of the existing methods in Godishala village. The obtained results show that the TCA-LSTM achieves better MSE of 0.002, MAE of 0.035, RMSE of 0.052, in contrast to RT [16] (Weekday) in Godishala village, India. Table 6 determines comparative analysis of existing technique in Warangal. When compared to PCA-RNN in RDA 1 [17], the TCA-LSTM achieves better MSE of 0.0074, RMSE of 0.029 respectively in Warangal, India. Table 7 denotes comparative analysis of existing techniques in Vijayawada City. The TCA-LSTM accomplishes superior MSE of 1.5047, MAE of 101.05, and RMSE of 396.24 in comparison to ANN [18] respectively.

Table 5. Comparative Analysis with existing methods in Godishala village

Datasets	Methods	MSE	MAE	RMSE
Power load dataset in Godishala village	RT [16]	0.004	0.041	0.067
	Proposed TCA-LSTM	0.002	0.035	0.052

Table 6. Comparative Analysis with existing methods in Warangal

Datasets	Methods	MSE	RMSE
Power load dataset in Warangal	PCA-RNN in RDM1 [17]	0.0154	0.142
	Proposed TCA-LSTM	0.0074	0.029

Table 7. Comparative Analysis with existing methods in Vijayawada

Datasets	Methods	MSE	MAE	RMSE
Solar and wind factors from Vijayawada city	ANN [18]	1.8151	131.72	426.06
	Proposed TCA-LSTM	1.5047	101.05	396.24

4.3 Discussion

Here, the advantage of TCA-LSTM and disadvantage of existing techniques are discussed. The RT [16] faces challenges with non-linear relationships due to partitioning data based on single feature thresholds which leads to inaccurate predictions. PCA [17] did not accurately forecast time-series data which limited RNN's capability to learn intricate dynamic loads. ANN [18] approach faces difficulties in managing non-linear and dynamic load behavior which affects forecasting accuracy in intricate microgrid environments. RNN-LSTM [19] faced challenges in capturing effective long-range dependencies in load patterns, particularly in scenarios with rapidly changing or highly irregular load dynamics. CNN-GRU [20] inaccurate forecasts minimize power supply safety, affecting social and economic activities, security, and national defense. The proposed TCA-LSTM overcomes these existing limitations. Temporal convolutional layers in TCN are established to capture both global and local temporal dependencies within the input sequences and attention technique enables the model to dynamically weight input sequences that enhance the forecasting accuracy. TCA-LSTM effectively captures long-range dependencies and focuses on salient temporal features in input sequences and variations in the patterns of electricity consumption which increase the forecasting performance by influential sequence factors in the prediction process. Therefore, TCA-LSTM achieves a superior MSE of 0.002, 0.0074, and 1.5047 in Godishala, Warangal, and Vijayawada compared to existing techniques like RT, PCA-RNN, and ANN.

5. Conclusion

In this research, the TCA-LSTM is proposed to forecast electric load data. By including an attention mechanism in an LSTM approach, the TCA-LSTM focuses more on parameters by greater weights. TCA-LSTM in electric load forecasting minimizes errors by streamlining dynamically on appropriate input features at time step which efficiently captures temporal dependencies via interpretability. Therefore, in contrast to the existing techniques like RT, PCA-RNN, and ANN, the proposed A-LSTM achieves a commendable MSE of 0.002, 0.0074, and 1.5047 in Godishala, Warangal, and Vijayawada cities, respectively. In the future, another self-attention-based classification technique will be considered for different datasets to enhance the efficiency of short-term load forecasting.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, the supervision and project administration, have been done by 1st author.

References

[1] A. Yadav, R. Bareth, M. Kochar, M. Pazoki, and R.A.E. Sehiemy, "Gaussian process regression-based load forecasting model", *IET Generation, Transmission & Distribution*, Vol. 18, No. 5, pp. 899-910, 2024.

[2] A. Inteha, F. Hussain, and I.A. Khan, "A data driven approach for day ahead short term load forecasting", *IEEE Access*, Vol. 10, pp. 84227-84243, 2022.

[3] C. Cai, Y. Li, Z. Su, T. Zhu, and Y. He, "Short-term electrical load forecasting based on VMD and GRU-TCN hybrid network", *Applied Sciences*, Vol. 12, No. 13, p. 6647, 2022.

[4] M. Li, H. Tian, Q. Chen, M. Zhou, and G. Li, "A hybrid prediction method for short-term load based on temporal convolutional networks and attentional mechanisms", *IET Generation, Transmission & Distribution*, Vol. 18, No. 5, pp. 885-898, 2024.

[5] A. Baul, G.C. Sarker, P. Sikder, U. Mozumder, and A. Abdelgawad, "Data-Driven Short-Term Load Forecasting for Multiple Locations: An Integrated Approach", *Big Data and Cognitive Computing*, Vol. 8, No. 2, p. 12, 2024.

[6] K. Ijaz, Z. Hussain, J. Ahmad, S.F. Ali, M. Adnan, and I. Khosa, "A novel temporal feature selection based LSTM model for electrical short-term load forecasting", *IEEE Access*, Vol. 10, pp. 82596-82613, 2022.

[7] M. Abumohsen, A.Y. Owda, and M. Owda, "Electrical load forecasting using LSTM, GRU, and RNN algorithms", *Energies*, Vol. 16, No. 5, p. 2283, 2023.

[8] J. Luo, T. Hong, Z. Gao, and S.C. Fang, "A robust support vector regression model for electric load forecasting", *International Journal of Forecasting*, Vol. 39, No. 2, pp. 1005-1020, 2023.

[9] N.A. Mohammed, and A. Al-Bazi, "An adaptive backpropagation algorithm for long-term electricity load forecasting", *Neural Computing*

Received: April 3, 2024. Revised: May 9, 2024.

and Applications, Vol. 34, No. 1, pp. 477-491, 2022.

[10] A. L'Heureux, K. Grolinger, and M.A. Capretz, "Transformer-based model for electrical load forecasting", *Energies*, Vol. 15, No. 14, p. 4993, 2022.

[11] L. Sun, H. Qin, K. Przystupa, M. Majka, and O. Kochan, "Individualized short-term electric load forecasting using data-driven meta-heuristic method based on LSTM network", *Sensors*, Vol. 22, No. 20, p. 7900, 2022.

[12] A. Moradzadeh, M. Mohammadpourfard, C. Konstantinou, I. Genc, T Kim, and B. Mohammadi-Ivatloo, "Electric load forecasting under False Data Injection Attacks using d2022.eep learning", *Energy Reports*, Vol. 8, pp. 9933-9945.

[13] T. Bashir, C. Haoyong, M.F. Tahir, and Z. Liqiang, "Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN", *Energy Reports*, Vol. 8, pp. 1678-1686, 2022.

[14] A. Shaqour, T. Ono, A. Hagishima, and H. Farzaneh, "Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building", *Energy and AI*, Vol. 8, p. 100141, 2022.

[15] P.C. Huy, N.Q. Minh, N.D. Tien, and T.T.Q. Anh, "Short-term electricity load forecasting based on temporal fusion transformer model", *IEEE Access*, Vol. 10, pp. 106296-106304, 2022.

[16] V. Veeramsetty, M. Sai Pavan Kumar, and S.R. Salkuti, "Platform-Independent Web Application for Short-Term Electric Power Load Forecasting on 33/11 kV Substation Using Regression Tree", *Computers*, Vol. 11, No. 8, p. 119, 2022.

[17] V. Veeramsetty, D.R. Chandra, F. Grimaccia, and M. Mussetta, "Short term electric power load forecasting using principal component analysis and recurrent neural networks", *Forecasting*, Vol. 4, No. 1, pp. 149-164, 2022.

[18] S.N.V.B. Rao, V.P.K. Yellapragada, K. Padma, D.J. Pradeep, C.P. Reddy, M. Amir, and S.S. Refaat, "Day-ahead load demand forecasting in urban community cluster microgrids using machine learning methods", *Energies*, Vol. 15, No. 17, p. 6124, 2022.

[19] A. Ajitha, M. Goel, M. Assudani, S. Radhika, and S. Goel, "Design and development of Residential Sector Load Prediction model during COVID-19 Pandemic using LSTM based RNN", *Electric Power Systems Research*, Vol. 212, p. 108635, 2022.

[20] A. Alrasheedi, and A. Almalaq, "Hybrid Deep Learning Applied on Saudi Smart Grids for Short-Term Load Forecasting", *Mathematics*, Vol. 10, No. 15, p. 2666, 2022.

[21] Dataset link for Godishala village: https://data.mendeley.com/datasets/tj54nv46hj/1

[22] Dataset link for Warangal: https://data.mendeley.com/datasets/ycfwwyyx7d/2

[23] Dataset link for Vijayawada: https://power.larc.nasa.gov

[24] M. Shantal, Z. Othman, and A.A. Bakar, "A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization", *Symmetry*, Vol. 15, No. 12, p. 2185, 2023.

[25] M. Koklu, I. Cinar, and Y.S. Taspinar, "CNN-based bi-directional and directional long-short term memory network for determination of face mask", *Biomedical signal processing and control*, Vol. 71B, p. 103216, 2022.