International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Short-Term Load Forecasting using Residual Bi-directional Gated Recurrent Unit with Self-Attention in Smart Grids

Sridhar Hassan Siddappa*[1]

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract: Short-Term Load Forecasting (STLF) plays a significant role in electrical management systems which provides accurate predictions of electricity demand over short-time demands and makes effective resource allocation. STLF makes distribution system operators implement effective energy management by engaging energy consumers over the demand-response program in Smart Grids (SG). However, STLF is challenging due to load exhibits highly nonlinear patterns which result from different factors like sudden changes in consumer behavior, and complex interactions. These nonlinearities make inaccurate forecasting. To address this issue, the Residual Bi-directional Gated Recurrent Unit with Self-Attention (Rbi-GRU-SA) is proposed to accurately forecast short-term load in SG which produces enhanced forecasting performance and reliability. Initially, the data is acquired from the electric load dataset to access the proposed approach. The z-score normalization is employed to normalize the dataset's features in the pre-processing phase which maximizes stability and convergence rate. Then, the Rbi-GRU-SA is performed to forecast the electric load in SG which provides more accurate forecasting. When compared to existing approaches like Feature Engineering-Wavelet Neural Networks and Self-Adaptive Momentum Factor (FE-WNN-SAMF), FE-Adaptive Grasshopper Optimization-based Locally Weighted Support Vector Regression (FE-AGO-LWSVR), and Gaussian Process Regression (GPR), the Rbi-GRU-SA achieves better MAPE of 0.0978 and 0.1054 for North South Wales (NSW) and Victoria (VIC) respectively.

Keywords: *North South Wales, Short-Term Load Forecasting, Smart Grids*

1. Introduction

Electricity is a vital part of our daily lives and has an essential impact on activities undertaken by individuals across various fields. There is a huge demand for electricity because of the rapid growth of the population all over the world. Due to the limited abilities of conventional electric grid stations, they are being replaced with the latest digital system of the power grid called Smart Grid (SG) [1]. SG is a new type of power system which is emerged recently and is mostly utilized by power-generating companies because of its accuracy in power load forecasting [2]. Compared to traditional power grids, SG enables more reliable, efficient, intelligent, and sustainable power service by employing advanced structures [3]. Accurate load forecasts are significant for making effective decisions based on energy management and dispatch [4]. Load forecasting is split into three types: STLF, Mid-Term Load Forecasting (MTLF), and Long-Term Load Forecasting (LTLF). SLTF provides numerous horizon hours enlarged to a few days [5] [6]. MTLF generates forecasts ranging from a week to several months, extending its prediction capabilities to encompass projections for the year ahead. LTLF is vital for projecting energy demand across a multi-decadal time horizon which supports capacity expansion planning under uncertain policy, climate, and technological changes [7] [8]. The load data has characteristics of volatility, randomness, diversity, and periodicity [9].

[1] Department of Electrical and Electronics Engineering, Siddaganga Institute of Technology, Tumkur, India
ORCID ID: 0000-0002-0828-1607
** Corresponding Author Email: sri054@sit.ac.in*

Accurate load forecasting not only provides society and people with economical, sustainable, and reliable power but also provides an efficient decision-making basis for planning power market investment [10] [11]. Resident's participation in the electricity market requires clear boundaries between short-term and long-term considerations to inform purchasing decisions effectively [12] [13]. Various factors impact the efficiency of STLF: i.e., calendar factors can make significant changes in electricity load. Weather conditions like humidity and temperature bring large uncertainties and nonperiodic results. Historical load is utilized to categorize trend characteristics and strong randomness of load series [14]. DL techniques like neural networks can learn complex relationships and patterns automatically from large datasets which allows to capture complex dynamics inherent in STLF more effectively compared to traditional Machine Learning approaches [15]. However, STLF is challenging because load exhibits highly nonlinear patterns that result from different factors like sudden changes in consumer behavior and complex interactions. These nonlinearities make inaccurate forecasting. To overcome this problem, the Rbi-GRU-SA is proposed to effectively forecast short-term load in SG by combining residual connection, bi-directional information flow, and self-attention mechanism by capturing complex dependencies in data.

The main contribution of this research is as follows:

- Z-score normalization is employed to normalize the dataset features by standardizing data distribution which allows better load pattern interpretation.
- Residual connection is used to solve vanishing gradient issues, bidirectional data flow is established to capture both past and future context by feedbacking the data, and self-attention is to dynamically weight input features.
- By combining these components, the Rbi-GRU-SA effectively captures complex temporal patterns and dependencies in STLF which provides accurate forecasting.

This paper is structured as follows: Section 2 explains about literature survey. Section 3 presents a detailed explanation of the proposed methodology. Section 4 discusses the results of the proposed methodology. Section 5 summarizes the conclusion of the paper.

2. Literature Survey

The related work about electricity load forecasting is discussed in this section with different techniques along with their advantages and limitations. This analysis helps in forecasting gaps and guiding the development of a more accurate and effective forecasting approach.

Muhammad Zulfiqar et al. [16] presented a Feature Engineering-Wavelet Neural Networks and Self-Adaptive Momentum Factor (FE-WNN-SAMF) to achieve stability, fast convergence, and high accuracy in STLF. FE eliminates inappropriate data and shallow features to enable high performance of computation. SAMF integrates frequency and time domain properties of wavelet transform and modifies WNN's associating parameters. At last, SAMF was utilized to tune the WNN's control parameter by initializing random thresholds and weights. By adjusting weights and thresholds dynamically via an iterative process, optimal parameter values were effectively reached which leads to enhanced performance and convergence rate. However, wavelet transforms lack interpretability and fail to effectively generalize across different load patterns because of their complex and multi-scale nature.

M. Zulfiqar et al. [17] suggested an FE-Adaptive Grasshopper Optimization-based Locally Weighted Support Vector Regression (FE-AGO-LWSVR) for STLF in SG. The HFS was optimized by employing Recursive Feature Elimination (RFE) to solve the overfitting issue in FE. Then, the significant features were extracted utilizing Radial Basis Kernel-based Principal Component Analysis (RBF-KPCA) to remove the issue of dimensionality reduction. AGO approach tunes the LWSVR's essential parameters to efficiently avoid entrapping into local optimum which produces accurate forecasting. The productiveness and efficiency of the suggested approach were equally differentiated with its stability and convergence rate. This balance makes reliable performance when reaching effective optimal solutions which increases the overall effectiveness of the suggested approach. However, the suggested approach considers only historical short-term load and does not consider other factors that impact the model's ability to evaluate all appropriate aspects of load behavior.

Anamika Yadav et al. [18] implemented a Gaussian Process Regression (GPR) that employs a Bayesian technique to forecast electric load. GPR was a non-parametric kernel-based learning technique that can generate accurate forecasts with uncertainty in measurements. This approach enhances the accuracy of load forecasting by capturing significant non-linear relationships in data which minimizes energy waste and increases overall electrical grid efficiency. However, GPR relies on a covariance structure defined by training data which makes it sensitive to data quality and leads to inaccurate forecasting.

Sang Mun Shin et al. [19] developed a Variational Mode Decomposition-based Random Vector Functional Link Network (VMD-RVFL) for STLF. The VMD was employed for electronic load decomposition into Intrinsic Mode Functions (IMFs) which minimizes the non-stationary and non-linear behavior of electric load. To forecast and model each IMF, RVFL was utilized due to its rapid and accurate forecasting performance. The developed VMD-RVFL achieves less computational time by using VMD efficiency in decomposing signals and the simplicity of RVFL in learning relationships which produces robust and accurate forecasting. However, the VMD-RVFL model's long-term dependencies within time-series data were limited due to the decomposition process.

Ankit Kumar Srivastava et al. [20] introduced an HFS based on the Elitist Genetic Algorithm (EGA) and Random Forest (RF) technique for STLF. This approach was established utilizing the concept of an identical week and for every season on a half-hourly basis. STLF was generated by employing the M5P forecaster by using the entire input feature set and chosen input feature set. By combining EGA with RF the model efficiently chooses appropriate features, optimizing the forecasting process for enhanced accuracy while maintaining interpretability via RF's decision-making process. However, the introduced approach has an issue in managing high-dimensional data due to increased redundancy and sparsity.

In the overall analysis, it is represented that existing methods have limitations like lack of interpretability and failure to effectively generalize across different load patterns due to their complex nature, nonlinear patterns, and changing load patterns in STLF. To solve this problem, the Rbi-GRU-SA is proposed for STLF in SG by combining residual connections to reduce vanishing gradient problems, bidirectional data flow to capture past and future context, and self-attention to dynamically weight input features. By performing these operations, it collectively makes an effective capture of nonlinear load patterns and complex load dynamic behavior.

3. Proposed Methodology

The Rbi-GRU-SA is proposed to forecast STLF in SG. Initially, the data is obtained from the electric load dataset and these data are processed using z-score normalization. Then, the DWT is performed to extract the features and Rbi-GRU-SA is employed to forecast short-term load in SG. Fig. 1 indicates the overview of a block diagram of the proposed technique.

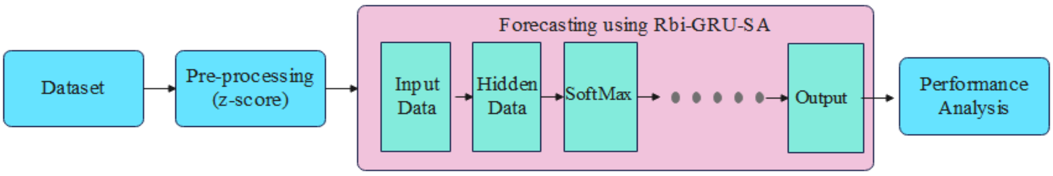

Fig. 1. Block diagram for the proposed technique

3.1. Dataset

In this research, the electric load dataset sourced from the Australian Energy Market Operator (AEMO) is employed to evaluate a proposed technique. Actual hourly load data of two Australian states are considered: NSW [21] and VIC [22] from 2018 to mid-2023 to calculate the model effectiveness for STLF. It contains electricity price and demand data recorded at 30-minute intervals. It has 5 columns which have day-ahead load demand, day-ahead temperature, real-time congestion, weather conditions, dry-bulb temperature, dynamic location marginal price, and day-ahead marginal loss. These data are preprocessed to normalize the dataset's features using z-score normalization.

3.2. Pre-processing

After obtaining the above data, the z-score normalization [23-24] is utilized to normalize the dataset's features to forecast short-term load data. It solves the problem of various scales in data by transforming features to have a mean of 0 and a standard deviation of 1 effectively in STLF. Z-score is less sensitive to outliers and generates a better distribution shape which provides clearer data interpretation. It effectively varies the distribution of data which enables it more appropriate for different scenarios of load forecasting. The z-score normalization is formulated in (1).

$$x'^{(j)} = \frac{x^j - \mu^j}{\partial^j} \quad (1)$$

Where μ represents the mean and ∂ indicates the standard deviation. Without using normalization, features with various scales may bias the learning process which leads to suboptimal model performance. Z-score is effective in managing with different scales and enhances model stability This process makes all features equally contribute to the learning process. Finally, the normalized data are fed into the feature extraction process.

3.3. Forecasting

The pre-processed data are passed as input to Rbi-GRU with self-attention for forecasting electric load data by integrating residual connections to reduce vanishing gradient problems, bidirectional data flow to capture past and future context, and self-attention to dynamically weight input features. It effectively captures nonlinear load patterns and complex load dynamic behavior. Compared to other neural networks, Rbi-GRU with self-attention can capture both hierarchical patterns and temporal dependencies efficiently in electric load data. The gating mechanism of GRU provides selective data processing which makes the network better adapt to the non-linear and dynamic nature of electric load data. This leads to improved forecasting compared to traditional RNNs due to their gating mechanism that makes the network selectively update and forget data from the previous time phase.

This gating approach contains a reset and update gate which regulates the data passage through a network which is expressed in (2) and (3).

$$u_t = \sigma(W_{hu}h_{t-1} + W_{xu}v_t + b_u) \quad (2)$$

$$r_t = \sigma(W_{hr}h_{t-1} + W_{xr}v_t + b_r) \quad (3)$$

Where u_t represents update gate at time t by utilizing data from both present inputs v_t and prior hidden state h_{t-1}. This update decision is determined by the sigmoid activation function $\sigma(x) = 1/(1 + e^{-x})$ based on a linear combination of hidden and input states with associated weight matrix W_{hu} and W_{xu} and bias term b_u. The r_t reset gate controls how much the network resets or forgets a prior hidden state h_{t-1} depending on the present input v_t. This reset decision is managed by an identical components set which has a weight matrix W_{xu} and W_{hu} and bias term b_r. These terms make the model adaptively rest and update information which is essential for temporal dependencies and enhancing forecasting performance in sequential data.

In bi-GRU [25], 2 GRU generates a sequence of input in forward and backward directions, making a network to include both previous and future sequence context by time

t. It helps forecast electric load data which is sequential by nature and generates long-term dependencies. Outcome of h_t of Bi-GRU for present time phase t is acquired by adding forward and backward hidden state $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ which is formulated in (4).

$$h_t = [\overrightarrow{h_t}\,;\,\overleftarrow{h_t}] \quad (4)$$

The $\overrightarrow{h_t}$ is computed depending on the candidate $\widetilde{h_t}$ hidden state, prior hidden state h_{t-1}, and update the gate u_t which is expressed in (5). $\overleftarrow{h_t}$ is calculated similarly but in the opposite direction.

$$\overrightarrow{h_t} = (1 - u_t)\,\Theta\,h_{t-1} + u_t\,\Theta\,\widetilde{h_t} \quad (5)$$

Where $\widetilde{h_t}$ indicates the candidate's hidden state that is calculated depending on the reset gate r_t, and present input vector v_t which is formulated in (6).

$$\widetilde{h_t} = \tanh\left(W_{hh}(r_t\,\Theta\,h_{t-1}) + W_{xh}v_t + b_r\right) \quad (6)$$

Rbi-GRU contains a residual association with the addition of Bi-GRU layers. Residual connections make data sent directly among layers. It assists in reducing the issue of repetitive gradient multiplications during the backpropagation procedure. It generates skip connections where the next layer input is acquired by adding the prior output layer to a residual. The i^{th} layer output with residual connections are expressed in (7).

$$h_i = F(W[h_{i-1}, X] + b) + X \quad (7)$$

Where h_i indicates i^{th} layer output with residual connections, X determines residual, h_{i-1} represents the prior output layer, W illustrates the weight matrix, b refers bias vector, and F indicates the function of non-linear activation like Rectified Linear Unit (ReLU).

3.3.1. Self-attention

It is a computing mechanism utilized for capturing dependencies among various components of a sequence. In the context of forecasting, it is employed to capture significant parts of load data that contribute to binding affinity. Self-attention makes the model dynamically weigh the significance of each component by transferring each sequence component into a key, a query, and a vector value. The attention score is computed utilizing a dot product among keys, queries, and vectors to evaluate the significance of one component to another which provides the identification of appropriate patterns and features for load forecasting. Then, the attention score is normalized by employing the SoftMax function to obtain the weight of attention that is employed for calculating the weighted sum of vector value which is expressed in (8).

$$Attention\,(Q, K, V) = SoftMax\left(\frac{Q_z K_z^{\,T}}{\sqrt{d_{k_z}}}\right)V_z \quad (8)$$

Where K, Q, and V indicate the key, query, and value that is calculated from z input sequences. d_{k_z} represents the key vector's dimension with z sequence. Fig. 2 represents the structure of Rbi-GRU-SA.

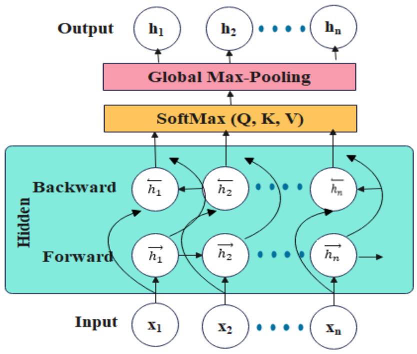

Fig. 2. Structure of Rbi-GRU-SA

The resulting weight of attention determines the significance of each component in a sequence that is later fed into a max-pooling layer. It assists in long sequences that capture dependencies among different components which is significant in electric load data. The self-attention and max-pooling operations are established to minimize the dimensionality of the output layer of self-attention. This makes the model focus on salient features of an input's sequences. Involving dropout regularization and batch normalization assist in increasing the model's generalization and avoiding overfitting. Fully Connected (FC) layers are employed to generate numerical inputs for electric load data. Then, the final output layer has a single neuron with linear activation which generates electric load forecasting. The combination of RBi-GRU with self-attention generates a robust approach to forecasting electric load data which increases forecasting performance and assists informed decision-making performance.

Table 1 shows the notation description

Table 1. Notation Description

Symbols	Description
μ	mean
∂	standard deviation
$h(k)$ and $g(k)$	low and high-pass filter
$\varphi\ (t)$	wavelet function
$\emptyset\ (t)$	scaling function
$a_{j,k}$	approximation coefficient expressed from low and high-frequency element $d_{j,k}$
v_i^{dwt}	element connected to various resolutions
n_i	sub band's number of samples
$W_{i,j}^2$	sub band for j^{th} coefficient
v_i	decomposed signal's mean square value at different values
u_t	update gate at time t
v_t	present input
h_{t-1}	prior hidden state
W_{hu} and W_{xu}	linear association of hidden and input state with associating weight matrix
b_u, b_r	bias term for update and reset gate
r_t	reset gate
$\overrightarrow{h_t}$ and $\overleftarrow{h_t}$	backward hidden state
$\widetilde{h_t}$	candidate hidden state
F	non-linear activation function
K, Q, and V	key, query, and value
d_{k_z}	key vector's dimension with z sequence

4. Results

The Rbi-GRU-SA is simulated utilizing a Python 3.8 environment with 64 GB RAM, Windows 10 Operating System, 1TB memory, and i9 intel processor. The Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), R^2, and Mean Average Error (MAE) are the metrics employed to evaluate the Rbi-GRU-SA performance. The MAPE measures the prediction accuracy of the forecasting technique and it evaluates the accuracy as ratio. RMSE evaluates the average difference between a statistical technique's forecasting values and actual values. R^2 is a statistical computation that evaluates the interrelation degree and dependence among two variables. MAE is an error measure among paired observations that expresses the same phenomenon. The mathematical formula for these metrics is expressed in (9) to (12)

$$MAPE = (\frac{1}{m}\sum_{\tau-1}^{m}\frac{u_\tau^r - u_\tau^\beta}{|u_\tau^p|}) \times 100 \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{\tau=1}^{m}(\lambda_\tau - \widehat{\lambda_\tau})^2}{m}} \tag{10}$$

$$R^2 = \frac{N\sum xy - \sum x \sum y}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \tag{11}$$

$$MAE = \frac{1}{m}\sum_{\tau=1}^{m}|\lambda_\tau - \hat{\lambda}_\tau| \tag{12}$$

Where m determines the number of data points, λ_τ indicates actual values, $\hat{\lambda}_\tau$ illustrates forecasted values, u_τ^r, u_τ^β determines forecasted and mean forecasted values, N refers number of observations, $\sum x$ indicates total. of initial variable value, $\sum y$ illustrates total. of second variable value, $\sum xy$ determine the sum of the product of the initial and second variable values, and $(\sum x)^2$ and $(\sum y)^2$ refers sum of squares of the initial, and second variable values respectively.

4.1. Performance Analysis

The proposed Rbi-GRU-SA performance analysis is indicated in Tables 2 to 4. Table 2 shows the performance of feature extraction evaluation for NSW. The Principle Component Analysis (PCA), Empirical Mode Decomposition (EMD), Symbolic Aggregate Approximation (SAX), and Continuous Wavelet Transform (CWT) are the existing techniques utilized to compare with DWT. Compared to these techniques, DWT is effective because it can capture both time and frequency-domain data at the same time which generates a multi-resolution load pattern representation for forecasting. Fig. 3. represents a graphical representation of the feature extraction technique for NSW. The outcome indicates that DWT achieves a better MAPE of 0.0978 compared to other techniques.

Table 2. Performance of feature extraction evaluation for NSW

Methods	MAPE	RMSE	R^2	MAE
PCA	0.2485	257.2347	2.1587	1.4325
EMD	0.1857	203.157	2.0657	1.3597
SAX	0.1764	187.35	1.9354	1.1593
CWT	0.1286	154.6587	1.8036	0.5469
DWT	0.0978	104.8746	1.6426	0.1424

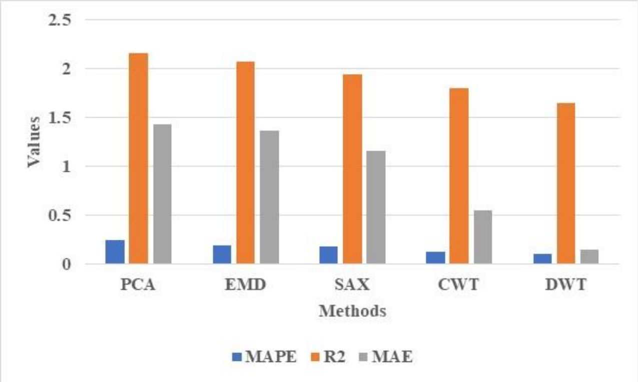

Fig. 3. Graphical representation of feature extraction analysis for NSW

Table 3 indicates the evaluation of forecasting performance for NSW. The performance of Long Short-Term Memory (LSTM), Radial Basis Function Neural Network (RBFNN), Gated Recurrent Unit (GRU), and Bi-directional-GRU (Bi-GRU) are compared with Rbi-GRU-SA. Compared to the techniques, Rbi-GRU-SA can capture the dependencies of long-range, exploit Bi-GRU data flow, and focus more on appropriate temporal features by utilizing SA. Fig. 4 determines a graphical representation of forecasting performance for NSW. The results show that Rbi-GRU-SA achieves a better MAPE of 0.0978 compared to LSTM, RBFNN, GRU, and Bi-GRU techniques.

Table 3. Evaluation of forecasting performance for NSW

Methods	MAPE	RMSE	R^2	MAE
LSTM	0.2036	234.0121	2.065	1.0354
RBFNN	0.1574	129.357	2.015	1.0158
GRU	0.1357	119.2574	1.998	0.9654
Bi-GRU	0.1035	109.1458	1.802	0.5230
Rbi-GRU-SA	0.0978	104.8746	1.6426	0.1424

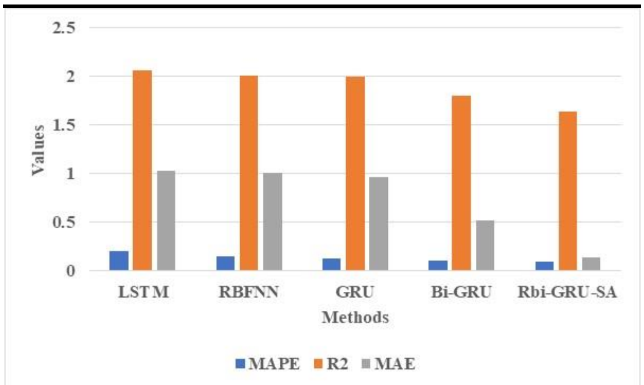

Fig. 4. Graphical representation of forecasting performance for NSW

Table 4 determines the performance of forecasting evaluation for VIC. The performance of LSTM, RBFNN, GRU, and Bi-GRU are compared with Rbi-GRU-SA. Fig. 5 indicates a graphical representation of forecasting performance for VIC. The outcome represents that Rbi-GRU-SA achieves a better MAPE of 0.1054 compared to existing techniques respectively.

Table 4. Performance of forecasting performance for VIC

Methods	MAPE	RMSE	R^2	MAE
LSTM	1.2458	265.3250	2.1584	1.3536
RBFNN	0.2036	236.4587	2.0658	1.2367
GRU	0.1936	196.3209	1.9658	1.0398
Bi-GRU	0.1563	150.0267	1.5893	1.0115
Rbi-GRU-SA	0.1054	128.6214	1.4273	0.0936

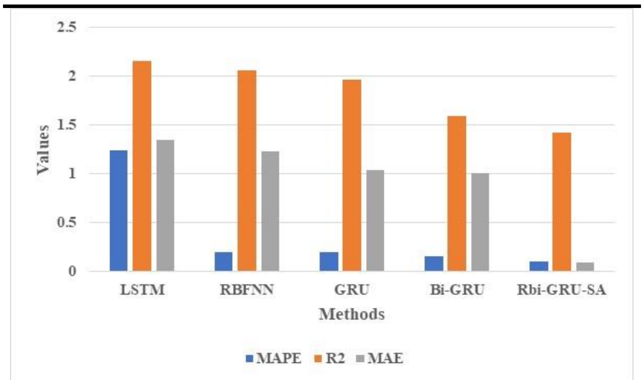

Fig. 5. Graphical representation of forecasting performance for VIC

4.2. Comparative Analysis

Table 5 indicates a comparative analysis with existing techniques for NSW and VIC. The existing approaches like FE-WNN-SAMF [16], FE-AGO-LWSVR [17], GPR [18], and VMD-RVFL [19] are employed to compare with the proposed Rbi-GRU-SA. Compared to these techniques, Rbi-GRU-SA can capture the dependencies of long-range, exploit Bi-GRU data flow, and focus more on appropriate temporal features by utilizing SA. Hence, the results show that Rbi-GRU-SA achieves better MAPE of 0.0978 and 0.1054 for NSW and VIC compared to existing techniques like FE-WNN-SAMF (1.4942, 1.4942), FE-AGO-LWSVR (0.1355, 0.1275), GPR (0.15), and VMD-RVFL (2.33, 3.87) respectively.

Table 5. Comparative analysis with existing techniques for NSW and VIC

Methods	Datasets	MAPE	RMSE	R^2
FE-WNN-SAMF [16]	NSW	1.4942	20.1891	0.92
	VIC	1.4942	22.1791	0.91
FE-AGO-	NSW	0.1355	N/A	N/A

LWSVR [17]	VIC	0.1275	N/A	N/A
GPR [18]	NSW	0.15	N/A	N/A
VMD-RVFL [19]	NSW	2.33	248.21	N/A
	VIC	3.87	310.49	N/A
Proposed Rbi-GRU-SA	NSW	0.0978	104.8746	1.6426
	VIC	0.1054	128.6214	1.4273

4.3. Discussion

Here, the advantages of Rbi-GRU-SA and the limitations of existing techniques are discussed. The limitations of existing techniques are that FE-WNN-SAMF [16] wavelet transforms lack interpretability and fail to effectively generalize across different load patterns because of their intricate and multi-scale nature. GPR [18] relies on a covariance structure defined by training data makes it sensitive to data quality and representation which leads to inaccurate forecasting. VMD-RVFL [19] model's long-term dependencies within time-series data were limited due to the decomposition process. The proposed Rbi-GRU-SA overcomes these limitations of existing techniques. The Rbi-GRU-SA effectively captures temporal dependencies which increase the ability of the model to accurately forecast short-term load data with non-linear and dynamic load patterns. By including residual connections, the proposed technique reduces vanishing gradient problems and assists with data flow over long sequences. The self-attention mechanism is utilized to dynamically weight input features which enhances its STLF effectiveness in SG. Therefore, the Rbi-GRU-SA achieves better MAPE of 0.0978 and 0.1054 for NSW and VIC compared to FE-WNN-SAMF, FE-AGO-LWSVR, GPR, and VMD-RVFL existing techniques.

5. Conclusion

In this research, the Rbi-GRU-SA is proposed to forecast short-term load in SG. Rbi-GRU-SA is performed for STLF which captures intricate temporal dependencies and patterns by efficiently combining residual connections, bi-directional data flow, and self-attention mechanism. Residual connections are utilized to reduce vanishing gradient problems, And bi-directional data flow and self-attention mechanisms are employed to capture dynamic weight input features. Performing this process makes effective capture of non-linear load patterns and complex load behavior. Hence, compared to existing approaches like FE-WNN-SAMF, FE-AGO-LWSVR, GPR, and VMD-RVFL, the proposed Rbi-GRU-SA achieves better MAPE of 0.0978 and 0.1054 for NSW and VIC. In the future, Long-Term Load Forecasting (LTLF) will be considered for accurate load forecasting.

Author contributions

Sridhar Hassan Siddappa: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Validation, Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The author declare no conflicts of interest.

References

[1] S. U. R. Khan, I. A. Hayder, M. A. Habib, M. Ahmad, S. M. Mohsin, F. A. Khan, and K. Mustafa, "Enhanced machine-learning techniques for medium-term and short-term electric-load forecasting in smart grids," *Energies*, vol. 16, no. 1, p. 276, Dec. 2022, https://doi.org/10.3390/en16010276.

[2] 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, Oct. 2022, https://doi.org/10.3390/s22207900.

[3] A. Motwakel, E. Alabdulkreem, A. Gaddah, R. Marzouk, N. M. Salem, A. S. Zamani, A. A. Abdelmageed, and M. I. Eldesouki, "Wild horse optimization with deep learning-driven short-term load forecasting scheme for smart grids," *Sustainability*, vol. 15, no. 2, p. 1524, Jan. 2023, https://doi.org/10.3390/su15021524.

[4] Y. Liu, Z. Dong, B. Liu, Y. Xu, and Z. Ding, "FedForecast: A federated learning framework for short-term probabilistic individual load forecasting in smart grid," *Int. J. Electr. Power Energy Syst.*, vol. 152, p. 109172, Oct. 2023, https://doi.org/10.1016/j.ijepes.2023.109172.

[5] 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, Aug. 2022, https://doi.org/10.1109/ACCESS.2022.3197609.

[6] N. Alemazkoor, M. Tootkaboni, R. Nateghi, and A. Louhghalam, "Smart-meter big data for load forecasting: An alternative approach to clustering," *IEEE Access*, vol. 10, pp. 8377-8387, Jan. 2022, https://doi.org/10.1109/ACCESS.2022.3142680.

[7] M. Madhukumar, A. Sebastian, X. Liang, M. Jamil, and M. N. S. K. Shabbir, "Regression model-based short-term load forecasting for university campus load," *IEEE Access*, vol. 10, pp. 8891-8905, Jan. 2022, https://doi.org/10.1109/ACCESS.2022.3144206.

[8] T. Zhang, X. Zhang, T. K. Chau, Y. Chow, T. Fernando, and H. H. C. Iu, "Highly accurate peak and valley prediction short-term net load forecasting

approach based on decomposition for power systems with high PV penetration," *Appl. Energy*, vol. 333, p. 120641, Mar. 2023, https://doi.org/10.1016/j.apenergy.2023.120641.

[9] X. Chen, W. Chen, V. Dinavahi, Y. Liu, and J. Feng, "Short-term load forecasting and associated weather variables prediction using ResNet-LSTM based deep learning," *IEEE Access*, vol. 11, pp. 5393-5405, Jan. 2023, https://doi.org/10.1109/ACCESS.2023.3236663.

[10] L. Xiao, M. Li, and S. Zhang, "Short-term power load interval forecasting based on nonparametric Bootstrap errors sampling," *Energy Rep.*, vol. 8, pp. 6672-6686, Nov. 2022, https://doi.org/10.1016/j.egyr.2022.05.016.

[11] M. Abdel-Basset, H. Hawash, K. Sallam, S. S. Askar, and M. Abouhawwash, "STLF-Net: Two-stream deep network for short-term load forecasting in residential buildings," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4296-4311, Jul. 2022, https://doi.org/10.1016/j.jksuci.2022.04.016.

[12] R. Gonzalez, S. Ahmed, and M. Alamaniotis, "Implementing very-short-term forecasting of residential load demand using a deep neural network architecture," *Energies*, vol. 16, no. 9, p. 3636, Apr. 2023, https://doi.org/10.3390/en16093636.

[13] N. A. Mohammed and A. Al-Bazi, "An adaptive backpropagation algorithm for long-term electricity load forecasting," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 477-491, Jan. 2022, https://doi.org/10.1007/s00521-021-06384-x.

[14] W. Chen, G. Han, H. Zhu, and L. Liao, "Short-Term Load Forecasting with an Ensemble Model Using Densely Residual Block and Bi-LSTM Based on the Attention Mechanism," *Sustainability*, vol. 14, no. 24, p. 16433, Dec. 2022, https://doi.org/10.3390/su142416433.

[15] F. M. Butt, L. Hussain, S. H. M. Jafri, H. M. Alshahrani, F. N. Al-Wesabi, K. J. Lone, E. M. T. E. Din, and M. A. Duhayyim, "Intelligence based accurate medium and long term load forecasting system," *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2088452, Jun. 2022, https://doi.org/10.1080/08839514.2022.2088452.

[16] M. ZulfiqAr, M. Kamran, M. B. Rasheed, T. Alquthami, and A. H. Milyani, "A short-term load forecasting model based on self-adaptive momentum factor and wavelet neural network in smart grid," *IEEE Access*, vol. 10, pp. 77587-77602, Jul. 2022, https://doi.org/10.1109/ACCESS.2022.3192433.

[17] M. Zulfiqar, M. Kamran, M. B. Rasheed, T. Alquthami, and A. H. Milyani, "A hybrid framework for short term load forecasting with a navel feature engineering and adaptive grasshopper optimization in smart grid," *Appl. Energy*, vol. 338, p. 120829, May 2023, https://doi.org/10.1016/j.apenergy.2023.120829.

[18] 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, Mar. 2024, https://doi.org/10.1049/gtd2.12926.

[19] S. M. Shin, A. Rasheed, P. Kil-Heum, and K. C. Veluvolu, "Fast and Accurate Short-Term Load Forecasting with a Hybrid Model," *Electronics*, vol. 13, no. 6, p. 1079, Mar. 2024, https://doi.org/10.3390/electronics13061079.

[20] A. K. Srivastava, A. S. Pandey, M. A. Houran, V. Kumar, D. Kumar, S. M. Tripathi, S. Gangatharan, and R. M. Elavarasan, "A day-ahead short-term load forecasting using M5P machine learning algorithm along with elitist genetic algorithm (EGA) and random forest-based hybrid feature selection," *Energies*, vol. 16, no. 2, p. 867, Jan. 2023, https://doi.org/10.3390/en16020867.

[21] Link for NSW: https://www.kaggle.com/datasets/joebeachcapital/nsw-australia-electricity-demand-2018-2023

[22] Link for VIC: https://www.kaggle.com/datasets/joebeachcapital/electricity-demand-vic-australia-2018-2023

[23] M. Nie, P. Chen, L. Wen, J. Fan, Q. Zhang, K. Yin, and G. Dou, "Wearable Recognition System for Complex Motions Based on Hybrid Deep-Learning-Enhanced Strain Sensors," *Adv. Intell. Syst.*, vol. 5, no. 11, p. 2300222, Aug. 2023, https://doi.org/10.1002/aisy.202300222.

[24] K. N. Ravikumar, C. K. Madhusudana, H. Kumar, and K. V. Gangadharan, "Classification of gear faults in internal combustion (IC) engine gearbox using discrete wavelet transform features and K star algorithm," *Eng. Sci. Technol. Int. J.*, vol. 30, p. 101048, Jun. 2022, https://doi.org/10.1016/j.jestch.2021.08.005.

[25] M. S. Başarslan and F. Kayaalp, "MBi-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis," *J. Cloud Comput.*, vol. 12, no. 1, p. 5, Jan. 2023, https://doi.org/10.1186/s13677-022-00386-3.