ORIGINAL ARTICLE

A data decomposition and attention mechanism-based hybrid approach for electricity load forecasting

Hadi Oqaibi[1] [ID] · **Jatin Bedi[2]**

Abstract

An accurate and reliable prediction of future energy patterns is of utmost significance for the smooth operation of several related activities such as capacity or generation unit planning, transmission network optimization, better resources availability, and many more. With the availability of historical load datasets through smart grid systems, artificial intelligence and machine learning-based techniques have been extensively developed for achieving the desired objectives. However, effectively capturing strong randomness and non-linear fluctuations in the load time-series remains a critical issue that demands concrete solutions. Considering this, the current research proposes a hybrid approach amalgamating data smoothing and decomposition strategy with deep neural models for improving forecasting results. Moreover, an attention mechanism is integrated to capture relevant portions of the time series, thus achieving the desired ability to capture long-term dependencies among load demand observations. This integration enhances the prediction and generalization capabilities of the proposed model. To validate the performance benefits achieved by the proposed approach, a comparative evaluation is conducted with state-of-the-art neural-based load series prediction models. The performance assessment is carried out on a novel real-world dataset of five southern states of India, and the superiority of the proposed in capturing load time-series variations is well observed and demonstrated in terms of several performance indicators.

Keywords Attention mechanism · Energy demand estimation · Forecasting · Long short-term memory · Gated recurrent unit

Introduction

The crucial role of large-scale electricity production in improvising citizens' lifestyles, economic growth, and several aspects of society has been increasingly evident in recent decades [1]. In today's era, all development and technological progress associated with activities are increasingly linked to the availability of electricity supply. The extreme dependence of modern society on relation to the supply of electric energy has already been amply demonstrated during blackouts [2] or supply crises, as happened in 2012 in India. The power industry holds a prominent position in the Indian economy, with India currently ranking third in the world for the production of electrical energy [3]. It contributes to around 1.2% GDP of the country [4]. Increasing energy efficiency is one of the priority policy directions outlined in the energy strategy of India for a specified period. The "Energy Strategy India for the period up to 2040" assumes accommodating a significant restructuring of the energy sector to elevate it to a newer standard capable of aligning with the nation's economic growth.

Over the globe, the electricity production chain operates through a system that includes the generation, transmission, distribution, and commercialization of electricity. This system utilized several energy sources to supply electricity to the end consumers. Unlike a conventional product chain, the "product" electrical energy cannot yet be stored on a large scale. Therefore, electricity generation is always done in synchronization with the demand. In this context, the studies on estimating the demand of the electricity sector are crucial for several activities such as generation planning, operational activities, transmission optimization, resources optimization, and many more [5, 6]. For instance, estimating the demand for future years will help plan the expansion of electricity systems and long-term planning of transmission and distri-

✉ Hadi Oqaibi
   hoqaibi@kau.edu.sa

1   Information System Department King Abdulaziz University, Jeddah, Saudi Arabia

2   Department of Computer Science and Engineering Thapar Institute of Engineering and Technology, Patiala, India

bution networks. Therefore, a comprehensive understanding of demand is pivotal in ensuring a sustainable and resilient electricity supply system.

The current study focuses on developing a reliable and accurate system for estimating the futuristic demand for electricity. The energy sector faces severe, diverse, and complex challenges, while estimating the demand due to the abrupt and non-linear demand patterns inherent in the dataset. Several statistical and conventional approaches have been employed in the past to assess trends, classify patterns, and anticipate energy consumption. Typically, the futuristic load estimation approaches are divided into two categories: statistical methods and artificial intelligence methods. The conventional methods [7–9] for predicting energy usage include stochastic time-series and regression-based approaches. These strategies have been widely employed in past research and have the potential to produce improved outcomes when handling linearity issues. However, traditional techniques (like the ARIMA or exponential smoothing models) presume historical trends or patterns will persist. Consequently, these methods are unsuitable for forecasts involving significant data fluctuations and fail to capture the highly fluctuating non-linear relationship between variables efficiently [10, 11].

Compared to statistical techniques, the artificial intelligence-based method can successfully deal with non-linear fluctuations in the load data [12]. These techniques are generally subdivided into two categories, namely machine learning methods and deep learning-based load estimation approaches. Machine learning approaches [13, 14] examine non-linear relationships between input characteristics and the output data. These methods overcome the limitations of existing traditional techniques by representing complicated relationships using functional mapping [15]. However, these methods do not efficiently investigate data correlations. The strong non-linear properties have been lately handled by incorporating a higher number of feature extraction (hidden layers) in the neural network [16, 17]. Effective and reliable deep learning-based predictive approaches (long short-term memory networks: LSTMs, gated recurrent units: GRUs and recurrent neural networks: RNNs) are proposed for load estimation and several different application domains [18–23]. Although recurrent neural networks (RNN) are effective for reflecting long-term dependency, owing to gradient loss concerns as the sequence length rises (long-term dependence), RNNs face challenges in reflecting these dependencies accurately [24]. To tackle this challenge, sequential networks (long short-term memory (LSTM) networks) demonstrating improvement in the load estimation domain have been proposed [24]. The LSTM neural network models are commonly used in time-series forecasts with periodic patterns [25].

Furthermore, several advanced deep learning and hybrid techniques have been introduced in recent years to estimate energy demand patterns efficiently [23, 26]. These include gated recurrent units, convolutional neural networks, integration of data decomposition techniques with deep learning models, and many more [27, 28]. From a comprehensive analysis of research studies, it has identified that hybrid techniques are more effective at describing the variation patterns of load time-series data [5, 11]. In this direction, the present research study introduces a hybrid technique combining Gaussian smoothing, data decomposition, and an attention mechanism to forecast electricity patterns. The key research outcomes of the present study are listed below:

- The energy load data are highly complex and carries non-linear fluctuations due to several factors, such as measurement errors, unpredictable patterns, anomalies, etc. These sudden/abrupt variations in the demand data can pose challenges to developing an inefficient prediction model. To overcome this limitation, the current approach employs the Gaussian smoothing technique to improve data quality before feeding it into the learning model. The Gaussian smoothing aims to remove irregularities and inconsistencies in the data, thus contributing to the generalization reliability and accuracy of the prediction model.

- The existing data decomposition techniques, such as wavelet, Fourier transform, and mode decomposition techniques, exhibit several limitations, such as noise sensitivity, dependence on shifting algorithm, and deciding the optimal number of modes. To address these limitations, the current research integrates CEEMDAN with neural models to achieve improved results.

- Efficiently capturing historical relationships within the load time-series observations is crucial for accurate predictions. In this context, the current research further integrates an attention mechanism with data decomposition, smoothing, and neural models to extract the relevant information while reducing the impact of irrelevant noise or errors. The attention mechanism enables the model to emphasize on the extracted relevant information by giving more weightage during the model-building phase.

- Lastly, the current research provides a novel dataset describing the load estimation patterns of the southern states of the country India. The proposed approach's performance evaluation is conducted on this novel dataset using widely adopted evaluation measures. The evaluation results describe the efficacy of the proposed approach in estimating the patterns in the specific context of the southern states of India.

Literature

This section provides a comprehensive description of the significant research works in the energy demand time-series forecasting domain. The research studies are categorized into two kinds discussed below:

Traditional, machine learning, and deep neural-based forecasting models

Several physical, statistical, and data-driven approaches for estimating the energy demand have been proposed in the recent past. These methods have considered several factors such as thermal dynamics, environmental factors, behavioral information, and financial considerations for building accurate demand prediction models. On the whole, it has been found that AI has contributed significantly to this domain, as the traditional physics-driven methods have several shortcomings such as equations driven, less adaptability, less generalizability, and many more. Researchers in the past have introduced several data-driven models (including machine learning and deep learning approaches) to surpass the shortcomings of existing models.

Earlier studies in the domain investigated several well-known and popular regression models, including simple linear regression, multi-linear regression models, and auto-regressive models (AR, ARMA, ARIMA, SARIMA, SARIMAX) at the target prediction task. The performance benefits achieved have shown that the auto-regressive models work wells at understanding the hidden trends and seasonality patterns of the data. Also, some recent enhancements in the regressive models have shown promising results. In [9], the authors developed a logistic mixture vector auto-regressive model for capturing the intra-variations and inter-variations of the time-series patterns. The model has attained significant performance improvement by integrating clustering and forecasting through a probabilistic approach. Three variants of several existing methods (auto-regressive, smoothing, and neural networks) were proposed by Kychkin and Chasparis [7] in the year 2021. The experimental evaluation stated that seasonality-based regressive model performs superior to other models, thus achieving the best performance. Alotaibi et al. [29] proposed an approach to determine the demand response sizing of the renewable energy systems for remote locations. The authors employed meta-heuristic optimization strategy for achieving the reliable and accurate results. Madrid and Antonio [30] proposed a machine learning-based system to accurately estimate the short-term load. The authors suggested that the ensemble machines performs better at load forecasting than conventional models.

With the advancements in the neural models building domain (ANN, DNN, RNN, LSTM, and GRU), researchers worldwide have proved the prediction capabilities of these models in a variety of application domains [31–33]. Some of the eminent research studies employing these models in the load estimation domain are discussed as follows: Bedi and Toshniwal [1] developed an LSTM model-based approach to forecast electricity demand. The authors captured seasonality patterns by integrating deep learning models with clustering analysis. The prediction benefits were well demonstrated on a dataset of UT Chandigarh, India. Mohammad and Kim [34] implemented neural models (ANN and RNN) to capture dynamic and uncertain variations of the energy load data. The authors explored several combinations of hyper-parameters to develop an optimal model for prediction. Huang et al. [35] introduced a probabilistic convolutional neural network-based method to estimate futuristic load. The approach involved implementing load range discretization to generate training samples (probabilistic). Mohammed et al. [36] proposed a novel regression-based approach to estimate electricity load of schools in Saudi Arabia. The authors investigated the effect of eleven external features on the electricity load using regression model. In this direction, Xu et al. [37] developed the Bayesian neural networks-based probabilistic approach to capture epistemic and aleatoric uncertainty of the electrical load patterns. Moradzadeh et al. [38] implemented a bi-directional LSTM network to capture historical and futuristic dependencies of the load series patterns. The developed model estimates demand at a micro-grid level, considering both residential households and commercial loads. Most of the research works in energy load forecasting are based on offline learning. In this context, Fekri et al. [39] introduced online learning-based adaptive approach to extract drifting patterns from the newly collected data. An online RNN is proposed to extract and learn new dependencies present in the data. In 2022, Khan et al. [40] combined echo network and CNN for delineating the renewable energy generation and consumption patterns. The approach aimed to establish an effective equilibrium between the consumers and the production units. Yazici et al. [41] worked on demonstrating the prediction capability of CNN models at short-term load estimation. A video pixel network based on 1D-CNN is implemented to estimate the futuristic load. The proposed network has shown improved ability compared to prediction performance estimation measures.

Hybrid techniques for load forecasting

Hybrid techniques include amalgamating two or more statistical, machine learning or deep learning models for attaining better prediction accuracy. Many research studies have well investigated the performance benefits of the hybrid techniques. This section performs a deep analysis of dominant and recently introduced hybrid techniques in the energy load forecasting domain.

Mamun et al. [42] performed an in-detailed prediction performance analysis of various hybrid load demand prediction techniques, including ANN, SVM, genetic algorithms, and their variants. The authors explained the integrated analysis of several techniques by detailing the advantages, disadvantages, and difficulties associated with each approach. The energy consumption in a country may get affected by real-time pricing patterns. Considering this, a hybrid approach to forecast electricity demand has been proposed by Dai and Zhao (2020) [43]. The author combined the support vector model with feature selection and parameters optimization for the desired prediction task. Massaoudi et al. [44] proposed a novel ensemble approach combining boosting machines with extreme learning and multi-layer perceptron architecture for short-term electricity load forecasting. Five different parameters optimization strategies were employed to optimize the prediction results. The performance superiority of the ensemble approach is verified by evaluating prediction performance on two real-world datasets. Bashir et al. [45] developed a model integration prophet with LSTM. The non-linear residual patterns of the original dataset (predicted by the prophet model) were estimated by employing the LSTM model. The assessment results were obtained on an eight-year quarterly aggregated Elia grid dataset. Alsharekh et al. [46] provided a deep learning approach integrating convolutional network with sequential model for electricity load forecasting. The evaluation benefits of the approach are validated on a real-world energy dataset. Recent advancements in the electricity demand forecasting domain have shown that the combined benefits of various machine and deep learning models have helped in achieving better prediction accuracy. Convolution neural networks have been found very effective at extracting feature information from the different kinds of data. Also, the accuracy of sequential models (including RNN, LSTM, and GRU) has also been verified at the future timestamp prediction activities. In this context, Wu et al. [47] introduced an approach utilizing CNN and LSTM for short-term load forecasting. The performance evaluation of the approach demonstrated that it this approach has superior prediction accuracy than the conventional prediction methods. Bedi and Toshniwal [5] combined auto-encoder models with deep neural models for optimizing the prediction accuracy of long-term forecasting models.

The amalgamation of decomposition strategies, including fast Fourier transform, empirical mode decomposition, wavelet transform with neural models, has shown great potential in capturing the seasonal, non-linear and chaotic patterns of the load demand data. These decomposition techniques support capturing the data's inherent complex time–frequency-based featural aspects. Zhang et al. (2018) [48] proposed a hybrid approach integrating auto-regressive models with EMD algorithm and fruit fly optimization. The authors emphasized the impact of social and natural factors on energy consumption. The usefulness of the approach is validated based on the simulation experiments. In year 2022, Huang et al. [49] integrated multivariate EMD algorithm with SVR and PSO technique for the day ahead peak demand forecasting. The empirical assessment is carried on a real-time peak power demand data of Victoria and North–South Wales, Australia. In a similar context, Bedi and Toshniwal [11] proposed an approach for load forecasting by integrating the EMD method with different deep neural models. The performance results are validated on the energy consumption dataset of UT Chandigarh, India. Zang et al. [50] proposed a two-stage preprocessing (including decomposition and reconstruction mechanism), LSTM and attention procedure to forecast day-ahead electricity demand of residential individuals. Lee and Cho [13] performed a comparative assessment of several machine learning, deep learning, and hybrid methods at the time-series prediction task. From the analysis results on Korea's peal load performance dataset, the authors concluded that the hybrid models provide significant improvement than the traditional machine and deep learning models. Sekhar et al. [23] proposed a hybrid approach combining convolutional approach with Bi-LSTM architecture for short-term load forecasting. The proposed approach parameters selection is performed using the gray wolf optimization approach. Ghimire et al. [51] provided a hybrid approach utilizing LSTM architecture in amalgamation with decomposition technique. In a similar context, a load forecasting technique based on LSTM was introduced by Grandon et al. [52]. The evaluation of the approach was performed on a real-time electricity demand dataset of Ukraine. These recently introduced approaches evaluation results' on the real-time datasets validated that the hybrid approaches perform much better than the conventional prediction approaches. However, these existing techniques lack at several crucial aspects such as noise handling, abrupt variations capturing, and complexity. To resolve these shortcomings, the present work introduces a workflow implementing data decomposition, attention, and Bi-GRU architecture for electricity load forecasting.

Working methodology description

The proposed approach follows a structured methodology comprising of several phases, namely data preprocessing, data smoothing, data decomposition, and model building. An in-depth description of each phase is given as follows:

Data preprocessing (Step-II)

Time-series data are prone to noise (errors and non-standard examples), erroneous readings, missing data, inconsistencies, and redundant information that are often distributed

across multiple heterogeneous sources. These data inconsistencies are the significant factors that reduce the data quality and consequently affect the quality and reliability of results. In this context, preprocessing techniques are commonly employed to enhance data quality and extract meaningful information that can be more effectively interpreted. The two major phases involved in the data preprocessing stage are discussed as follows:

- Data cleaning: This stage aims to rectify inconsistencies in the dataset, like missing records and inconsistent values. In the past, several data cleaning techniques, such as imputations by central tendency/regression techniques/clustering, are employed to improve data quality [53]. However, in the context of the problem under consideration, the missing values imputation approach that considers the historical dependencies will be suitable. Therefore, the current research employs a missing value technique that leverages the past seven years' load time-series observations.
- Data reduction and transformation [53]: These techniques (such as data discretization, dimensionality reduction, integration, and standardization) aim to prepare/convert the data into a format that is conducive for building learning models. The current approach employs the min–max normalization technique to transform data to a common range. The mathematical representation of the min–max normalization is given as follows [53]:

$$N_{ts} = \frac{O_{ts} - min(O_{ts})}{max(O_{ts}) - min(O_{ts})}. \tag{1}$$

Here, O_{ts} and N_{ts} represent the old time series and the new scaled time series, respectively. The maximum and minimum values corresponding to the old time series are represented by $max(O_{ts})$ and $min(O_{ts})$.

Gaussian smoothing (Step-III)

Time-series data collected by means of sensors and devices are highly vulnerable to noise, outliers, or abnormal variations. The existence of such disruption or inconsistencies in the data may impact the overall analysis and prediction process. There are several means to mitigate the effect of noise in the data, i.e., by collecting larger sample space, series smoothing, sensors, or apparatus improvement or many more. In the current research, we have adopted smoothing to reduce the impact of noise on load time-series variations patterns. There exist several smoothing strategies that can be employed to improve the data, namely simple or running average smoothing, weighted average smoothing, Gaussian smoothing, and exponential smoothing [54]. The average

smoothing methods consider each timestamp observation as a weighted sum of previous observations estimated by a user-defined window. These average smoothing methods have several shortcomings such as finite width window, handling outside range values, etc. Hence, we have implemented Gaussian smoothing for removing noise present in the load series dataset by convolving each input time series with a Gaussian function at timestamp t_i is defined as (with width b representing the smoothing parameter) [23, 55]

$$gaussian\_k(t_i) = exp\left(-\frac{(t^* - t_i)^2}{2b^2}\right). \tag{2}$$

The convolution function involves utilizing the Gaussian kernel function/curve to weigh the load timestamp observations present in the input dataset. Then, we proceed by calculating the new observation using the weighted average of k data points returned by kernel function. Mathematically, it can be presented as

$$X'_t = \sum X_{t_i}.gaussian\_k(t_i), \tag{3}$$

where $X_{t(i)}$ denotes the input series, Y'_t represents the smoothed value. For an example, the input time series and the corresponding smoothened time series are demonstrated in Fig. 2. From the figure, it can be observed that the new smoothened load time series has less noise. The Gaussian smoothing filter has attuned the large dips and spikes very effectively. An important hyper-parameter while applying smoothing is to choose the value of alpha/b. Large value may lead to data loss by losing important spikes and dips of the non-linear time-series dataset. The value of this parameter has been estimated by random research technique.

Data decomposition (Step-IV)

In signal processing, decomposing a signal into multiple components is a potential technique, effectively utilized across several application domains. The specific application domains include utilizing decomposition strategies for reducing the amount of data to be transferred or stored (compression), eliminating unpredictable components, namely noise and outliers and extracting relevant information through filtering and retrieval. Several data decomposition techniques have been proposed in the past for both discrete and continuous time domains. Examples include the sine and cosine transforms, Fourier transform, wavelet decomposition, and mode decomposition (EMD) [56, 57]. From the exhaustive literature analysis, it has been identified that EMD provides more accurate results than other decomposition techniques [58], such as Fourier, wavelet decomposition, and many more.

Empirical mode decomposition (EMD) is a method of decomposing time–frequency series, whose initial proposal is to deal with signals governed by non-linear and non-stationary processes. The method was developed and introduced by Huang et al. [59] and it works by decomposing the input series into several orthogonally based functions called intrinsic mode functions (IMF). Each IMF obtained by decomposition highlights the local characteristics of the data. The lowest-frequency IMF component is the residual term, which represents the original signal's general trend or mean value. The sum of all the IMFs with the residual recompose the original series, according to the following equation:

$$I(ts) = c_1(ts) + \sum_{i=1}^{M} c_i(ts) + r(ts). \tag{4}$$

Although the EMD technique has been widely employed, the EMD lacks sufficient theoretical basis, and its susceptibility to modal aliasing causes many uncertainties in the decomposition results. Despite these limitations, the EMD technique showcases versatility in decomposing a time domain signal into a set of nearly and complete orthogonal basis functions (i.e., the IMFs), each having the same length of the original signal, but exhibiting varying frequencies. Hence, enhanced versions of EMD have been developed to address the existing shortcomings. One such method is EEMD (ensemble empirical mode decomposition) introduced by Wu et al. [60]. The core idea behind this noise-assisted data analysis approach is to fuse the actual time-series information with noise so that even if different sources collect the same process data that has different characteristics, its overall mean should be close to the real-time series. To extract the real signal from the data, a method incorporating multiple sets of white noise sequences with limited amplitude to the original sequences is introduced. Subsequently, each signal is then decomposed separately, and the mean value of the corresponding components is considered as the real component. Additionally, the introduction of noise into the signal selected for decomposition has been proposed as a strategy to the mitigate the problem of mode mixing. This solution has led to the development of new variants of the EMD, such as the ensemble empirical mode decomposition (EEMD) [60], and the complete EEMD with adaptive noise known as complementary ensemble empirical mode decomposition with assisted noise (CEEMDAN) [61]. The current study adopts CEEMDAN data decomposition technique to overcome the shortcomings of existing data decomposition techniques. The CEEMDAN method is employed due to its capability to rectify the modes mixing problem. Moreover, the method also effectively suppresses the residual noise in the IMFs, thus leading to a superior signal decomposition and reconstruction (with minimal reconstruction error). The following steps briefly summarize the procedure of decomposition by complementary ensemble empirical mode with assisted noise:

1. Add a series of white noise (adapted) in the original sequence $x(t)$. It is given as:

$$I^i(ts) = x(ts) + \omega_0 \epsilon^i(ts) \quad i \in \{1, \dots F\}. \tag{5}$$

Here, $I^i(ts)$ is the i-th time series with white noise added; ω_0 is noise figure; $\epsilon^i(ts)$ is the i-th added noise; F is integrated frequency.

2. Decompose $I^i(ts)$ using empirical mode decomposition (EMD) and compute the mean of the first IMF component $c_1^i(t)$:

$$c_1(ts) = \frac{1}{F} \sum_{i=1}^{F} c_1^i. \tag{6}$$

Remove $c_1(t)$ from the original sequence $I(td)$ to get the 1st residue:

$$r_1(ts) = I(ts) - c_1(ts). \tag{7}$$

3. Continue EMD decomposition for $r_1(ts) + w_1 E_1[\epsilon^i(ts)]$ to get the second IMF component

$$c_2(ts) = \frac{1}{F} \sum_{i=1}^{F} E_1\{r_1(ts) + w_1 E_1[\epsilon^i(ts)]\}. \tag{8}$$

Here, the j-th IMF component generated by decomposition procedure is given by $E_j(.)$

4. Iterate the subsequent procedure to compute the remaining IMF components:

$$r_k(ts) = r_{k-1}(ts) - c_k(ts), k = 2, 3, \dots, K, \tag{9}$$

$$c_{k+1}(ts) = \frac{1}{F} \sum_{i=1}^{F} E_1\{r_k(ts) + w_k E_k[\epsilon^i(ts)]\}. \tag{10}$$

Here, the total number of modes is denoted by K.
The overall procedure concludes when the residue sequence cannot be decomposed further. The final residual component is represented as:

$$R(ts) = I(ts) - \sum_{k=1}^{K} c_k(ts). \tag{11}$$

Model building (Step-V and Step-VI)

This section describes the architectural explanation of the proposed load demand forecasting strategy. The overall section comprises two sub-phases. The first phase provides a

comprehensive description, working, and suitability of the proposed attention-based bi-directional gated sequential neural model to the problem under study. In the second phase, the working of the proposed prediction model amalgamating the data decomposition, attention mechanism, and deep neural network models is summarized.

Attention-based bi-directional GRU network model

Artificial neural network (ANNs) [53] models have been widely adopted and found successful in prediction scenarios pertaining to different application domains. However, in practice, these learning models are unable to learn dependencies from the distant past. Since, in the time-series dataset, the observations of the current sequence might depend on the link between far-away sequences. Hence, capturing such remote impacts are critical for executing reliable modeling. To cover this aspect, the LSTM networks [24] have been introduced by Hochreiter Schmidhuber. These network models are capable of handling long-term contextual or historical dependencies present in the data. In the current study, the bi-directional GRU, an improved model of the LSTM, is employed to forecast the load time-series patterns. The GRU networks have proven to be efficient in identifying and representing sudden variations in the contextual data (NLP domain). Hence, the network is suitable for the current task with highly nonlinear, complex and abrupt time-series variations. Basically, as a sequential structured, GRU networks [62] efficiently forwards information across multiple time-steps, preventing the gradual loss of essential information. However, the conventional GRU network can capture dependencies in one direction (historical) only. Hence, the Bi-GRU-type networks consisting of two GRU networks in parallel is employed in the current approach: the first network processes the input sequence from right to left, while the second network processes it in reverse direction. This dual approach enhances the network's ability to capture complex temporal patterns effectively.

The GRU network operates through two gate functions [62], specifically known as the reset gate and the update gate. These gates play a crucial role in capturing both short-term and long-term dependencies within a sequence of data. Essentially, the GRU gates determine which information should be passed forward to produce an accurate output. Their significance lies in their ability to learn how to retain information from past events over extended periods (thus, addressing the vanishing gradient problem) or to discard irrelevant information, which enhances data prediction.

Furthermore, the GRU maintains a memory of features that contributes to generating the current state, subsequently serving as the memory for the next state in chronological order. This strengthens the correlation within the data series, leading to improved results from the extracted features. Figure 1 illustrates the schematic of the GRU network, which, unlike LSTM, does not have separate memory cells but rather utilizes connection units capable of modulating information. Each unit is equipped with an update gate (Z_t) and a reset gate (R_t), which determine the exposure of their memory content, while balancing current and past information. The update gate regulates the adaptation time. Additionally, in addition to the update and reset gates, the structure includes the candidate state ($\tilde{H}_t$) and the output block (H_t) for the time period denoted as t. The GRU network is defined by the following equation [62].

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z), \tag{12}$$

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r), \tag{13}$$

$$\tilde{H}_t = tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h), \tag{14}$$

$$H_t = (1 - Z_t) \odot H_{t-1} + Z_t \odot \tilde{H}_t;$$

$$Here \ \sigma(x) = \frac{1}{1 + e^{-x}} \ and \ tanh(x) = 1 - \frac{2}{1 + e^{-2x}}, \tag{15}$$

$$\overrightarrow{H}_t = GRU(X_t, \overrightarrow{H}_{t-1}), \tag{16}$$

$$\overleftarrow{H}_t = GRU(X_t, \overleftarrow{H}_{t-1}), \tag{17}$$

$$H_t = w_t \overrightarrow{H}_t + v_t \overleftarrow{H}_t + b_t. \tag{18}$$

Within this framework, X_t represents instances, which is a set containing the training sequence. The weights W_{xz}, W_{hz}, W_{xr}, W_{hr}, W_{xh}, W_{hh} are indicative of the connections between X_t and X_{t-1}, while b_z, b_r, and b_h signify the biases. Notably, element-wise multiplications are computed using the Hadamard product $\odot$, and the activation functions for both gates are characterized by logistic sigmoids $\sigma(\odot)$, constraining R_t and Z_t to values within the range of 0 to 1.

- **Update gate** [62]: The primary role of the update gate is to determine which past information should be retained and passed on to the future state.
- **Reset gate** [62]: The reset gate, on the other hand, decides the extent to which past information should be discarded.

The update gate is mathematically represented in eq. (11), where the input x_t and the previous state information h_{t-1} are combined with their respective weights to form the model parameters. The sigmoid activation function yields Z_t. Similarly, in the reset gate, eq. (12) demonstrates that the input x_t and previous state information $ht − 1$ are multiplied by specific weights, and the sigmoid function produces Rt. For time step t, the candidate hidden state H cap is calculated as per eq. (13). The first term in eq. (13) accounts for the contribution from input x_t and the corresponding weight, while the second term computes the product of a weight and the result
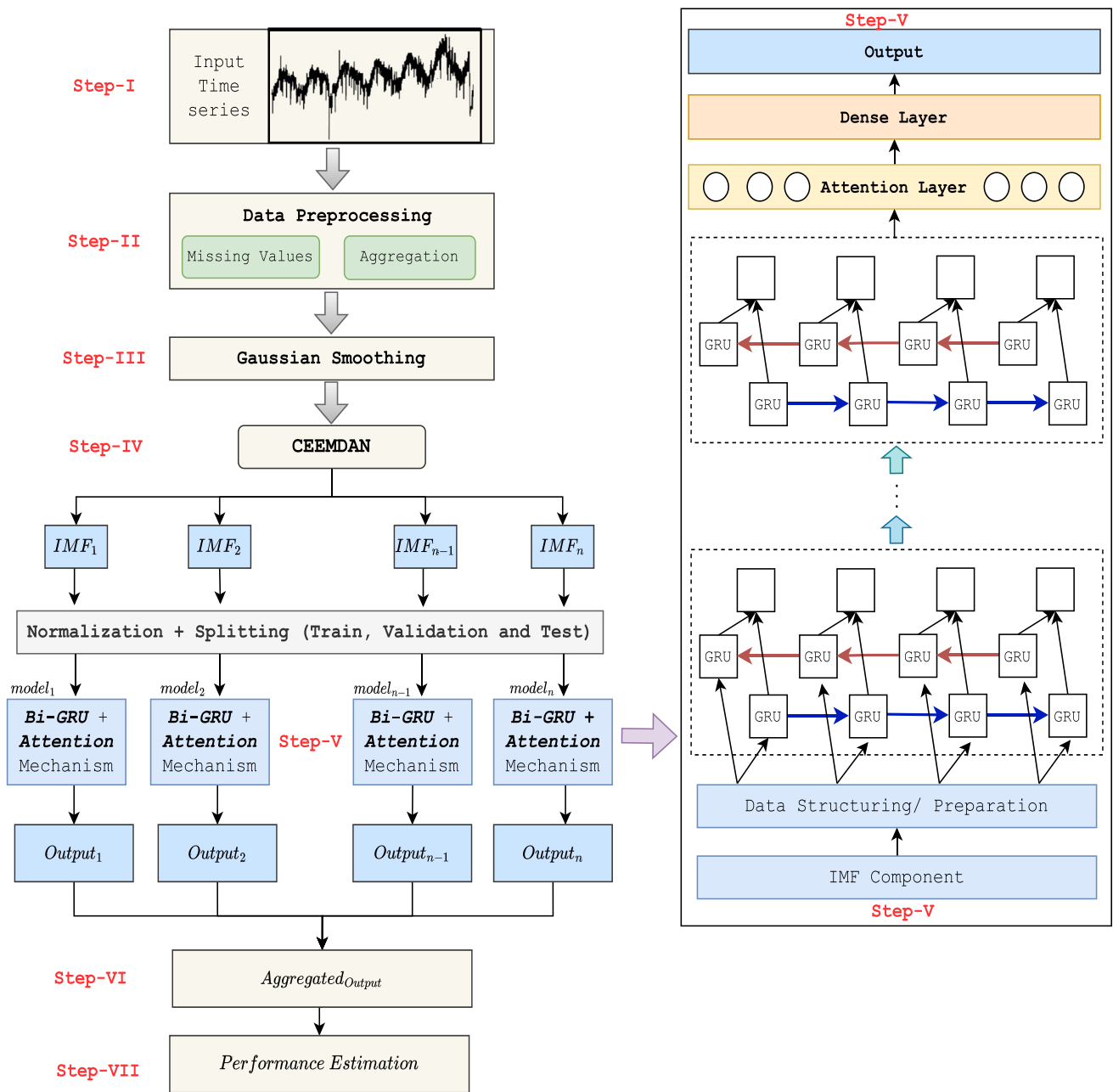
Fig. 1 Methodology of proposed approach

of the Hadamard product between R_t and H_{t-1}. Finally, the network calculates the H_t vector, which holds information about the current unit and forwards it throughout the network. The update gate plays a crucial role in implementing this process. Subsequently, each sample's vector is computed by aggregating its hidden output H_t from both directions, as described in eqs. 15, 16. This bi-directional information is incorporated into H_t, as shown in eq (17).

Sequential models for capturing long-term data dependencies are effective but computationally expensive. In recent

years, attention networks have been introduced to identify the relevant context from text data to estimate future events. Attention mechanism (AM) [63] is a cognitive psychology model that simulates human brain attention. In concept, it is highly similar to the human visual attention system. It is an effective strategy for swiftly locating relevant information from vast amounts of data, eliminating unnecessary information, and completing tasks more effectively. These methods have achieved significant accuracy in the NLP domain. In the current work, we integrate the attention mechanism with

the neural model (Bi-GRU) to extract relevant information from available long-term historical data. The aim is to identify the significant variations and fluctuations in the given load time-series data to determine future patterns efficiently. Furthermore, the inclusion of an attention mechanism to the proposed approach enables the model to capture dependencies at a deeper level compared to conventional neural models (like LSTM and Bi-GRU). As illustrated in Fig. 1, the attention mechanism computes a similarity measure between its input and the target state. A simple neural network is utilized to determine the weight of the timestamp in the suggested improvement framework. During normalization, the softmax activation function is used to make the output weights equal to one.

Proposed data decomposition-based prediction strategy

In this section, a detailed explanation on the working of the proposed data decomposition integrated deep-learning model building strategy for improved forecasting efficacy is provided. The proposed strategy is fundamentally rooted in the "divide and conquer" principle, which requires the dataset to be decomposed into several different intrinsic mode components using the CEEMDAN decomposition strategy. Subsequently, each decomposed component is then used as input to an attention-based Bi-GRU model to undergo the prediction task. For better understanding and reproducibility, a comprehensive, step-by-step working of the proposed strategy is given as follows:

- **Sub-step 1:** Data decomposition—The CEEMDAN technique is implemented to decompose the load time-series dataset corresponding to each state into corresponding IMFs and residual components.
- **Sub-step 2:** The GRU model needs input data structured in the three dimensions (S, W, and F). Here, S represents the number of input timestamps, W means the sequences length, and F represents the features within each sequence. The lag parameter is applied to generate a feature input matrix for input to the bi-directional GRU models. The attention network models is designed and included to capture the dependencies within each identified mode component.
- **Sub-step 3:** Performance validation is critical for defining the reliability of the prediction models. As a result, the feature matrix obtained from sub-step 2 is divided into training, validation, and testing datasets. The training and validation sets are utilized to develop a model and obtain an unbiased estimate of its prediction accuracy.
- **Sub-step 4:** The attention-based bi-directional network models are trained and developed for each IMF and residual component (shown in Fig. 1). As desired, the relevant temporal intrinsic dependencies identified by the

attention mechanism are effectively captured by the bi-directional GRU model during this stage.
- **Sub-step** After successful training, the developed sequential models are employed to generate a forecast for the testing dataset. The target prediction outcomes are determined by aggregating the forecasting results of all IMFs (corresponding to the respective state).

Hyper-parameters tuning is crucial for building any machine or deep learning-based prediction model. The present research work employs the widely adopted grid search strategy [64] determine the optimal value of several hyper-parameters related to the proposed model, namely input window size (size of the input), out size (prediction horizon), number of dense layers, optimizer, activation function, and many more. Different permutations of hyper-parameter values initially defines the search space in the grid search strategy. The defined search space is then explored for the best set of parameters with minimal prediction errors. Finally, the prediction model with the least errors is deployed for estimating the futuristic patterns.

Experimental results and discussion

The section entails describing the experimental evaluation performance of the proposed load estimation approach. The evaluation is conducted on a novel real-world electricity demand dataset of the southern states of India. The objective here is to validate the applicability, reliability, and accurateness of the proposed approach. The overall process of comparing the proposed approach with state-of-the-art load estimation approaches is explained below:

Dataset description

The current study collects and analyzes the real-time electricity consumption dataset of five southern states: Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, and Puducherry. The southern states are known to be the key energy consumers, so targeting these states will help in finding/defining better energy management strategies for the country India. For an in-depth analysis of energy usage trends and patterns, the dataset for these states is collected over a period of seven years, starting from January 2014. The timestamp observations have been recorded on a per-day basis. Moreover, the geographical details of the five states, along with the statistical description of energy consumption usage patterns, are provided in Table 1.

Table 1 Statistical and geographical description of the southern states dataset

State	Latitude	Longitude	Min	Max	Mean	Std.
Andhra Pradesh	15.9129° N	79.7400° E	93.50	284.8	162.23	32.33
Karnataka	15.3173° N	75.7139° E	112.20	273.3	184.45	27.46
Kerala	10.1632° N	76.6413° E	38.90	89.4	65.66	6.97
Tamil Nadu	11.1271° N	78.6569° E	144.0	365.4	284.52	32.07
Puducherry	11.9416° N	79.8083° E	2.40	9.70	6.847	0.91
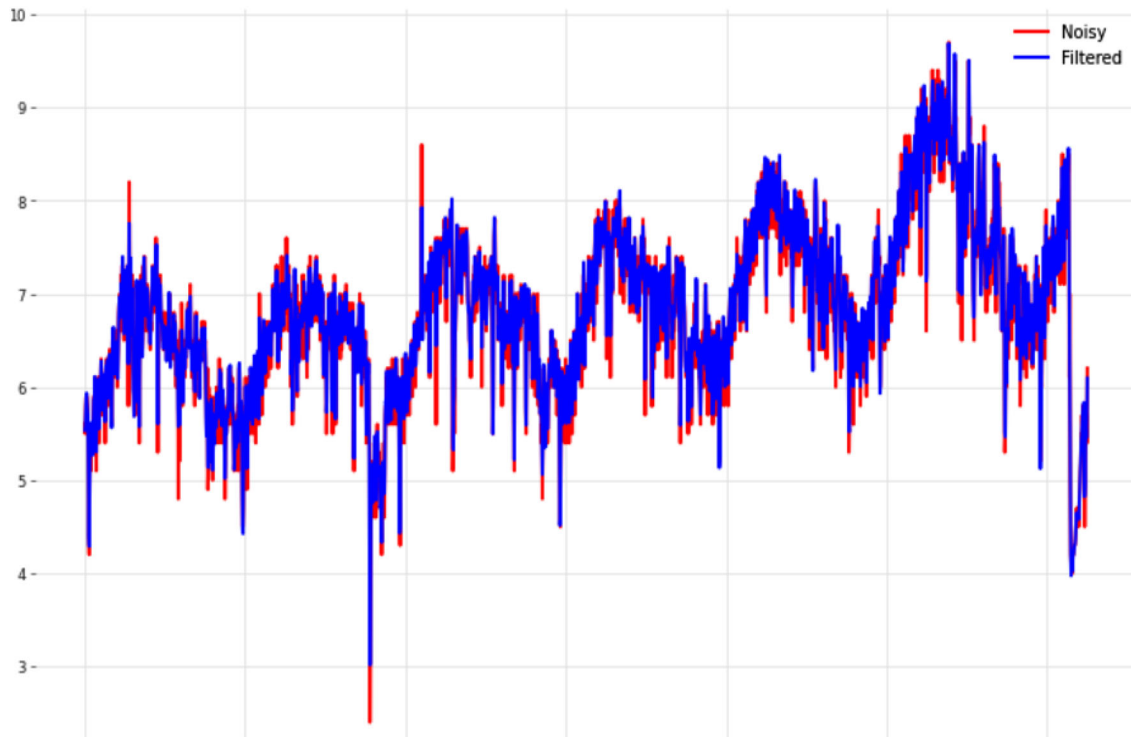
Fig. 2 Results of Gaussian smoothing

Data preprocessing and smoothing

This step entails applying the preprocessing "Data Preprocessing (Step-II)" and smoothing methods "Gaussian Smoothing (Step-III)" on the dataset collected for the five southern Indian states. There were very few missing values present in the dataset for each state. To address this, the mean values considering the next five days, corresponding to the same period of all seven years, are utilized to fill in these missing values. Since the present study considers only the historical load observations for estimating the future load, there is no need for the applicability of data reduction techniques. However, in context of data transformation, the normalization technique defined in eq. 1 is applied on the historical observations.

After data preprocessing, the time series representing each state energy usage pattern is fed to the smoothing module. The aim here is to remove the existence of any noisy and abnormal variations present in the usage profiles. The cur-

rent study applied a Gaussian kernel to perform weighted smoothing on the data. For better understanding, an example figure demonstrating the application of Gaussian smoothing on time-series data of the 'Puducherry' state is shown in Fig. 2 From the figure, it is identified that the Gaussian smoothing aims to minimize the impact of any irrelevant or sudden observation patterns that may be present in the input energy consumption patterns data. This, in turn, helps to smoothen/amplify the critical variations patterns which the learning models can easily capture.

Time-series decomposition

The next step after applying the smoothing technique is to decompose the time series corresponding to each state into several IMF components. For this, the CEEMDAN techniques discussed in "Data decomposition (Step-IV)" is employed. An example demonstration of the decomposition results achieved using CEEMDAN on the 'Andhra Pradesh'

Fig. 3 An example demonstration of CEEMDAN method on 'Andhra Pradesh' load time series
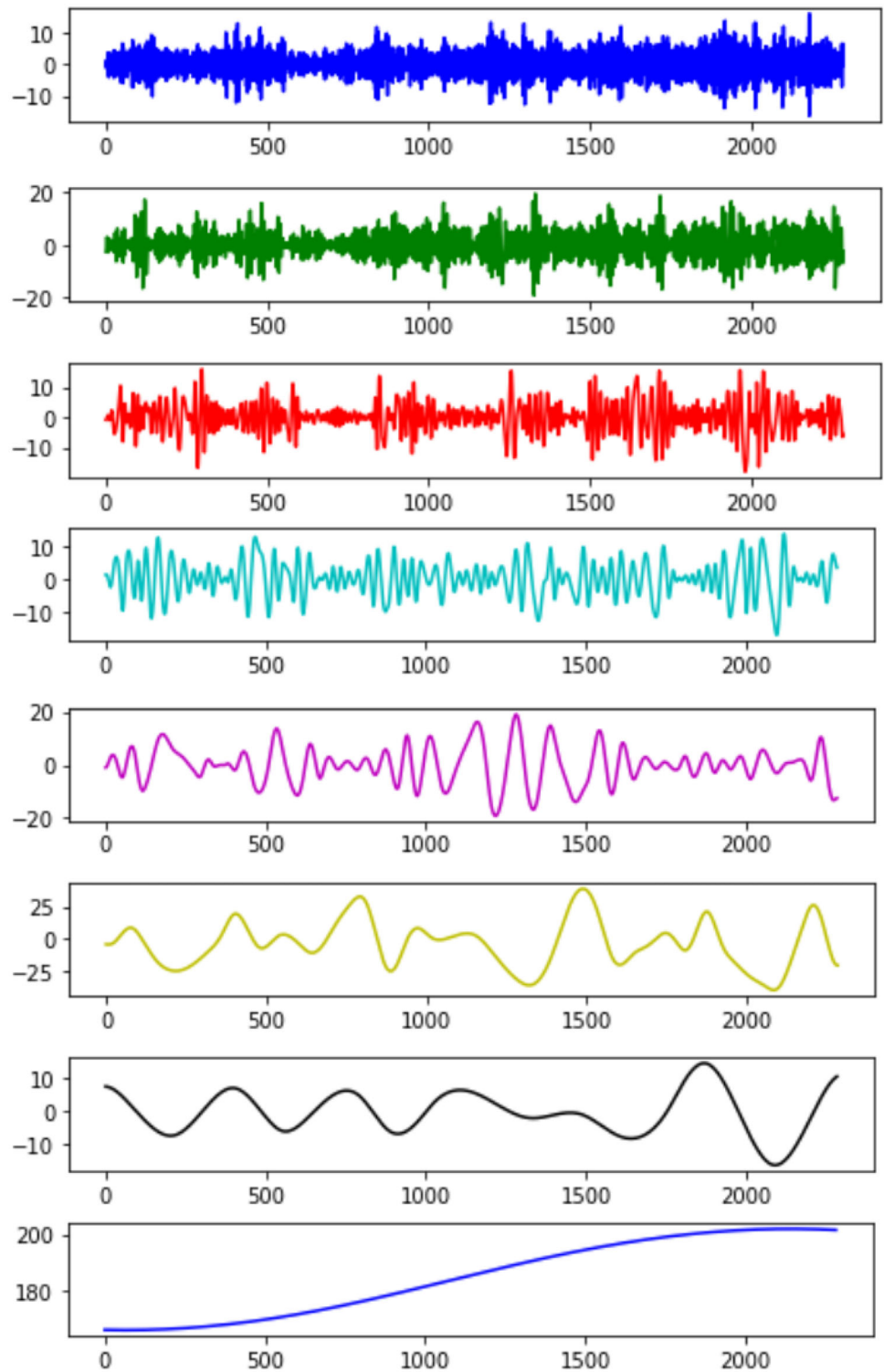
Table 2 Optimal value of hyper-parameters

Hyper-parameter name	Optimal range (for all States)
Input size	24
Prediction horizon	1
Number of Bi-LSTM layers	4 – 6
Number of neurons per layer	32 – 256
Learning rate	0.001 – 0.0001
Activation function	'ReLU'
Optimizer	'ADAM'

Table 5 Comparative results of the proposed approach with benchmark techniques (Part-III)

Pondy			
Model/Measure	RMSE	MAPE	MAE
LSTM	0.453	0.0486	0.347
GRU	0.450	0.0474	0.332
Bi-GRU	0.427	0.0472	0.331
Proposed Approach	0.247	0.0255	0.182

time series is provided in Fig. 3. The figure shows several IMFs components obtained for the input time series. Furthermore, from the figure, it is evident that the decomposition method efficiently segregates the time series into components with varying magnitude and inter IMF non-linear patterns, thus reducing the non-linearity and abrupt variations profiles in the raw input data.

Comparative evaluation of the proposed approach (Step-VII)

The IMF components resulting from the above steps are fed as an input to the attention based bi-directional gated neural models. As discussed in the subsection, this step includes building a separate model for each IMF component generated for a southern state. The process is repeated for all the states present in the input dataset. Furthermore, to validate the effectiveness and reliability of our proposed work, advanced deep neural models (LSTM, GRU, and Bi-GRU) are also developed for state-level load estimation. The comparative evaluation outcomes of these models are listed in the Tables 3, 4, and 5. The tables represent the comparative results of the

proposed approach against the benchmark approaches on the dataset of the southern states, namely Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, and Puducherry, respectively. The comparative evaluation uses three widely known performance parameters: root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE). A detailed description of these parameters is given as follows:

- **RMSE**: is a statistical measure to quantify the extent of errors between the model-predicted values from a model and the actual observations.
- **MAPE**: is a metric to quantify the error as a percentage deviation of model-predicted observations from the actual values.
- **MAE**: is a measure to signify the absolute difference between the observed and the model-predicted values.

A critical factor to consider while developing any neural-based solution to a problem is to define the most-suitable value of model hyper-parameters. In the current study context, the identified optimal values corresponding to the proposed model hyper-parameters are outlined in Table 2. The results observed for the best combination of hyper-

Table 3 Comparative results of the proposed approach with benchmark techniques (Part -I)

Andhra Pradesh				Karnataka			
Model/Measure	RMSE	MAPE	MAE	Model/Measure	RMSE	MAPE	MAE
LSTM	6.320	0.0285	4.988	LSTM	9.236	0.0385	7.392
GRU	5.694	0.0245	4.250	GRU	9.119	0.0363	7.257
Bi-GRU	5.787	0.0250	4.369	Bi-GRU	9.047	0.0340	7.194
Proposed	2.854	0.0103	2.210	Proposed	4.230	0.0200	3.018

Table 4 Comparative results of the proposed approach with benchmark techniques (Part-II)

Kerala				Tamil Nadu			
Model/Measure	RMSE	MAPE	MAE	Model/Measure	RMSE	MAPE	MAE
LSTM	2.843	0.0307	2.107	LSTM	12.575	0.0310	8.821
GRU	2.763	0.0293	1.996	GRU	12.151	0.0306	8.767
Bi-GRU	2.707	0.0288	1.832	Bi-GRU	12.063	0.0303	8.293
Proposed	2.486	0.0257	1.760	Proposed	6.875	0.0182	5.251

parameters are presented in Tables 3, 4, and 5. From analyzing the evaluation results, several critical observations are deduced:

- Comparative evaluation results of existing benchmark techniques (LSTM, GRU, and Bi-GRU) show that bidirectional gated recurrent units model performs better by capturing the dual directional load dependencies patterns present in the data.
- The overall performance comparison results shown in the Tables 3, 4, and 5 clearly indicate that the proposed approach performs accurate prediction by returning the least mean absolute prediction error compared to other existing deep learning-based solutions.
- The average prediction benefits achieved by the proposed approach are observed to be 50%. It clearly validates that the proposed approach accurately identifies and estimates the non-linear patterns present in the energy consumption dataset of all southern states. Hence, integrating Gaussian smoothing and CEEMDAN into deep neural models can be considered a viable and accurate solution to improve the prediction results of models developed for different application domains.

Visualization of prediction results

This section includes depicting the predicting outcomes of the proposed approach on the input load demand dataset belonging to different states. The aim here is to assess and demonstrate the reliability and prediction capability of the proposed approach at capturing the non-linear and chaotic variations of the energy consumption dataset. To achieve this, the predicted timestamp values for different states generated by the proposed approach are plotted against their respective actual observation values in Fig. 4. In Fig. 4, the x-axis represents the samples and the y-axis denotes the energy consumption value. The actual observations are represented using the black color, while the predicted values on the training and testing dataset are colored in 'blue' and 'purple' color, respectively. From the Fig. 4, it is observed that the proposed approach accurately quantifies/captures the energy demand variations present in the dataset of different southern states. Moreover, the approach work very well at estimating the peak and abrupt energy demand requirement. Hence, from the comprehensive analysis, it can be stated that the approach can be reliably employed to accurately forecast energy demand in the real-world scenarios.

Discussion

From the experimental evaluation of the proposed approach on five datasets corresponding to different southern states of India, it has been well validated that the proposed approach outperforms existing state-of-the-art approaches. These results are verified based on several different evaluation measures. A few critical observations are drawn from the experimental evaluation of the proposed approach. First, from the experimental evaluation, it has been observed that the models trained through the proposed approach attained saturation earlier than the conventional deep learning-based model-building strategy. The reason behind this is the inclusion of attention and smoothing mechanisms, which helped the proposed approach better learn from the critical weighted data dependencies identified through the attention module. The energy demand time-series data are highly complex, non-linear, and carries intrinsic data dependencies. In this direction, the decomposition procedure divides the complex non-linear energy demand patterns into several distinct, less-complex patterns, which has helped the proposed module easily capture the non-linear chaotic variations present in the data. This has resulted in improved accuracy of the proposed approach. Lastly, the prediction model's performance is highly impacted by noise and unknown factors introduced by human errors or malfunctioning. In this context, introducing Gaussian smoothing to the proposed approach has reduced the impact of noise present in the collected dataset.

Conclusion

In the current study, a novel hybrid approach combining several crucial components is proposed to resolve the shortcomings of the existing load time-series prediction approaches. Initially, the proposed approach comprises implementing data smoothing to reduce the impact of sudden and random fluctuations of the input dataset. Subsequently, a data decomposition strategy resolving the shortcomings of the existing decomposition techniques is employed to extract mode components from the smoothen dataset. Finally, an attention mechanism in integration with deep Bi-GRU network models is deployed to extract, learn, and estimate the dependencies patterns from the decomposed components. The attention mechanism is introduced to concentrate on the crucial information by assigning varying weights to the hidden states of the deep learning models, thus reducing loss of relevant information. To validate the accuracy of the proposed approach, the dataset corresponding to five southern states of country India is utilized. Based on the comparative evaluation, following major conclusions are drawn from the proposed work:
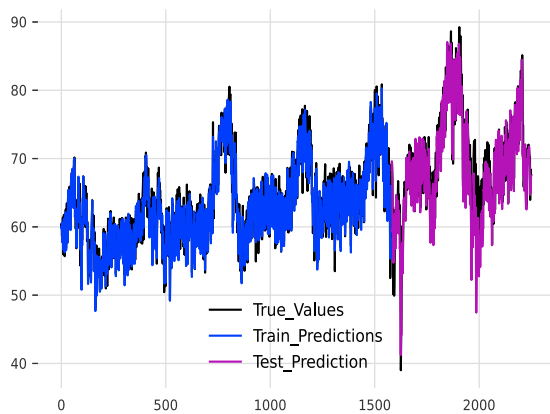
- From the prediction plots shown in Fig. 4, it is evident that the inclusion of data smoothing and decomposition strategy to the proposed approach enables effective and accurate capturing of randomness and variations in the load time-series data. This resolves a significant draw-
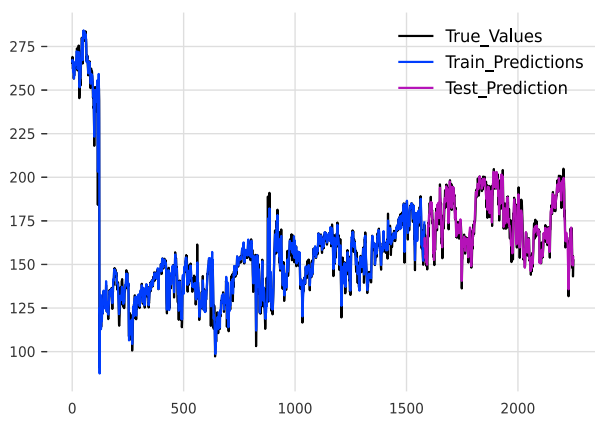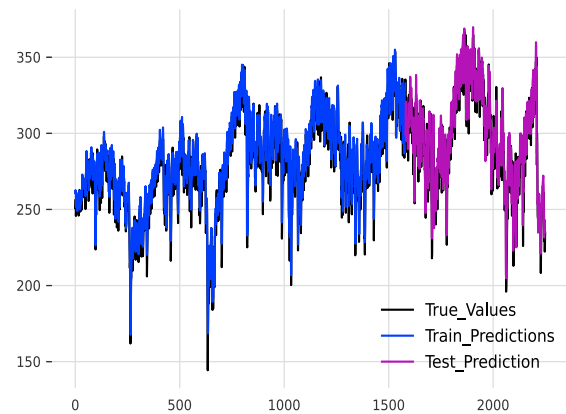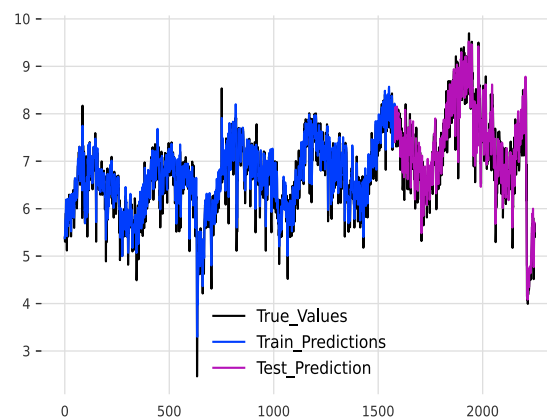
(a) Andhra Pradesh
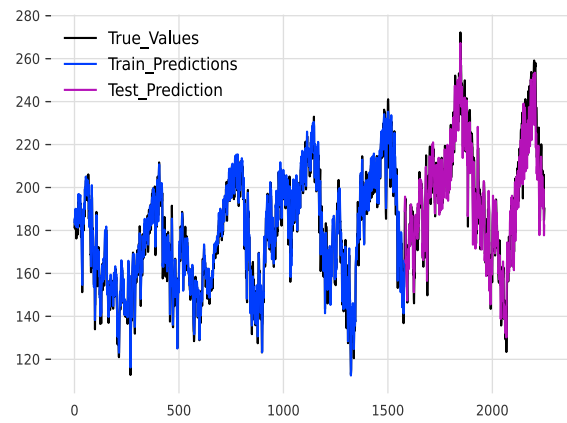
(b) Karnataka

(c) Kerala

(d) Tamil Nadu

(e) Pondy

Fig. 4 Visualization of prediction results

back of the existing research studies in the targeted domain.

- The comparative evaluation with state-of-the-art load-series prediction methods clearly demonstrates the pre-

diction accuracy, reliability and robustness of the proposed approach. From the prediction results, it is evident that the proposed approach outperforms traditional deep learning-based prediction models by providing reduction

in the prediction error. These performance benefits are achieved by the combined benefits of data decomposition and bi-direction attention mechanism employed in the proposed approach.

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interests The author(s) declare(s) that there is no conflict of interest.

References

1. Bedi J, Toshniwal D (2019) Deep learning framework to forecast electricity demand. Applied energy 238:1312–1326
2. Romero JJ (2012) Blackouts illuminate india's power problems. IEEE spectrum 49(10):11–12
3. India Energy Outlook, https://www.iea.org/reports/india-energy-outlook-2021, [Online; accessed 2023-02-07]
4. Jackson RB, Le Quéré C, Andrew R, Canadell JG, Korsbakken JI, Liu Z, Peters GP, Zheng B (2018) Global energy growth is outpacing decarbonization. Environmental Research Letters 13(12):1–8
5. Bedi J, Toshniwal D (2020) Energy load time-series forecast using decomposition and autoencoder integrated memory network. Applied Soft Computing 93:1–14
6. Khan AM, Osińska M (2023) Comparing forecasting accuracy of selected grey and time series models based on energy consumption in brazil and india. Expert Systems with Applications 212:118840. https://doi.org/10.1016/j.eswa.2022.118840
7. Kychkin AV, Chasparis GC (2021) Feature and model selection for day-ahead electricity-load forecasting in residential buildings. Energy and Buildings 249:111200. https://doi.org/10.1016/j.enbuild.2021.111200
8. Tarsitano A, Amerise IL (2017) Short-term load forecasting using a two-stage sarimax model. Energy 133:108–114
9. Jeong D, Park C, Ko YM (2021) Short-term electric load forecasting for buildings using logistic mixture vector autoregressive model with curve registration. Applied Energy 282:116249
10. Zhao Y, Zhang C, Zhang Y, Wang Z, Li J (2020) A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. Energy and Built Environment 1(2):149–164
11. Bedi J, Toshniwal D (2018) Empirical mode decomposition based deep learning for electricity demand forecasting. IEEE access 6:49144–49156
12. Manandhar P, Rafiq H, Rodriguez-Ubinas E (2023) Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods. Energy Reports 9:2757–2776
13. Lee J, Cho Y (2022) National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? Energy 239:122366. https://doi.org/10.1016/j.energy.2021.122366
14. Ribeiro AMN, P. R. X. do Carmo, P. T. Endo, P. Rosati, T. Lynn, (2022) Short-and very short-term firm-level load forecasting for warehouses: a comparison of machine learning and deep learning models. Energies 15(3):750. https://doi.org/10.3390/en15030750
15. Shahare K, Mitra A, Naware D, Keshri R, Suryawanshi H (2023) Performance analysis and comparison of various techniques for short-term load forecasting. Energy Reports 9:799–808
16. Song X, Wu N, Song S, Zhang Y, Stojanovic V (2023) Bipartite synchronization for cooperative-competitive neural networks with reaction–diffusion terms via dual event-triggered mechanism, Neurocomputing 126498 https://doi.org/10.1016/j.neucom.2023.126498
17. Zhuang Z, Tao H, Chen Y, Stojanovic V, Paszke W (2022) An optimal iterative learning control approach for linear systems with nonuniform trial lengths under input constraints, IEEE Transactions on Systems, Man, and Cybernetics: Systems . https://doi.org/10.1109/TSMC.2022.3225381
18. Song X, Wu N, Song S, Stojanovic V (2023) Switching-like event-triggered state estimation for reaction–diffusion neural networks against dos attacks, Neural Processing Letters 1–22
19. Peng Z, Song X, Song S, Stojanovic V (2023) Hysteresis quantified control for switched reaction-diffusion systems and its application. Complex & Intelligent Systems 9(6):7451–7460
20. Wieczorek M, Siłka J, Woźniak M (2020) Neural network powered covid-19 spread forecasting model. Chaos, Solitons & Fractals 140:110203
21. Siłka J, Wieczorek M, Woźniak M (2022) Recurrent neural network model for high-speed train vibration prediction from time series. Neural Computing and Applications 34(16):13305–13318
22. Dong Y, Ma X, Fu T (2021) Electrical load forecasting: A deep learning approach based on k-nearest neighbors. Applied Soft Computing 99:106900
23. Sekhar C, Dahiya R (2023) Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand. Energy 268:126660. https://doi.org/10.1016/j.energy.2023.126660
24. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation 9(8):1735–1780
25. Woźniak M, Wieczorek M, Siłka J (2023) Bilstm deep neural network model for imbalanced medical data of iot systems. Future Generation Computer Systems 141:489–499. https://doi.org/10.1016/j.future.2022.12.004
26. Bedi J (2020) Attention based mechanism for load time series forecasting: An-lstm, in: Artificial Neural Networks and Machine Learning–ICANN 2020: 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 15–18, 2020, Proceedings, Part I 29, Springer, pp. 838–849
27. Jamei M, Ali M, Malik A, Rai P, Karbasi M, Farooque AA, Yaseen ZM (2023) Designing a decomposition-based multi-phase pre-processing strategy coupled with edbi-lstm deep learning approach for sediment load forecasting. Ecological Indicators 153:110478
28. Sulaiman S, Jeyanthy PA, Devaraj D, Shihabudheen K (2022) A novel hybrid short-term electricity forecasting technique for residential loads using empirical mode decomposition and extreme learning machines. Computers & Electrical Engineering 98:107663. https://doi.org/10.1016/j.compeleceng.2021.107663

29. Alotaibi MA, Eltamaly AM (2021) A smart strategy for sizing of hybrid renewable energy system to supply remote loads in saudi arabia. Energies 14(21):7069

30. Aguilar Madrid E, Antonio N (2021) Short-term electricity load forecasting with machine learning. Information 12(2):50

31. Liu G, Yin Z, Jia Y, Xie Y (2017) Passenger flow estimation based on convolutional neural network in public transportation system. Knowledge-Based Systems 123:102–115

32. Abosamra G, Oqaibi H (2021) Using residual networks and cosine distance-based k-nn algorithm to recognize on-line signatures. IEEE Access 9:54962–54977

33. Chauhan J, Bedi J (2023) Effvit-covid: A dual-path network for covid-19 percentage estimation. Expert Systems with Applications 213:118939. https://doi.org/10.1016/j.eswa.2022.118939

34. Mohammad F, Kim Y-C (2020) Energy load forecasting model based on deep neural networks for smart grids. International Journal of System Assurance Engineering and Management 11(4):824–834

35. Huang Q, Li J, Zhu M (2020) An improved convolutional neural network with load range discretization for probabilistic load forecasting. Energy 203:117902. https://doi.org/10.1016/j.energy.2020.117902

36. Mohammed A, Alshibani A, Alshamrani O, Hassanain M (2021) A regression-based model for estimating the energy consumption of school facilities in saudi arabia. Energy and Buildings 237:110809

37. Xu L, Hu M, Fan C (2022) Probabilistic electrical load forecasting for buildings using bayesian deep neural networks. Journal of Building Engineering 46:103853. https://doi.org/10.1016/j.jobe.2021.103853

38. Moradzadeh A, Moayyed H, Zakeri S, Mohammadi-Ivatloo B, Aguiar AP (2021) Deep learning-assisted short-term load forecasting for sustainable management of energy in microgrid. Inventions 6(1):1–15

39. Fekri MN, Patel H, Grolinger K, Sharma V (2021) Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network. Applied Energy 282:116177

40. Khan ZA, Hussain T, Baik SW (2022) Boosting energy harvesting via deep learning-based renewable power generation prediction, Journal of King Saud University-Science 101815 https://doi.org/10.1016/j.jksus.2021.101815

41. Yazici I, Beyca OF, Delen D (2022) Deep-learning-based short-term electricity load forecasting: A real case application. Engineering Applications of Artificial Intelligence 109:104645. https://doi.org/10.1016/j.engappai.2021.104645

42. Mamun A Al, Sohel M, Mohammad N, Sunny MSH, Dipta DR, Hossain E (2020) A comprehensive review of the load forecasting techniques using single and hybrid predictive models, IEEE Access 8 134911–134939

43. Dai Y, Zhao P (2020) A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. Applied Energy 279:115332. https://doi.org/10.1016/j.apenergy.2020.115332

44. Massaoudi M, Refaat SS, Chihi I, Trabelsi M, Oueslati FS, Abu-Rub H (2021) A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting. Energy 214:118874. https://doi.org/10.1016/j.energy.2020.118874

45. Bashir T, Haoyong C, Tahir MF, Liqiang Z (2022) Short term electricity load forecasting using hybrid prophet-lstm model optimized by bpnn. Energy Reports 8:1678–1686

46. Alsharekh MF, Habib S, Dewi DA, Albattah W, Islam M, Albahli S (2022) Improving the efficiency of multistep short-term electricity load forecasting via r-cnn with ml-lstm. Sensors 22(18):6913

47. Wu L, Kong C, Hao X, Chen W (2020) A short-term load forecasting method based on gru-cnn hybrid neural network model, Mathematical Problems in Engineering 2020

48. Zhang J, Wei Y-M, Li D, Tan Z, Zhou J (2018) Short term electricity load forecasting using a hybrid model. Energy 158:774–781

49. Huang Y, Hasan N, Deng C, Bao Y (2022) Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting. Energy 239:122245. https://doi.org/10.1016/j.energy.2021.122245

50. Zang H, Xu R, Cheng L, Ding T, Liu L, Wei Z, Sun G (2021) Residential load forecasting based on lstm fusing self-attention mechanism with pooling. Energy 229:120682

51. Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S (2023) Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach. Energy Conversion and Management 297:117707. https://doi.org/10.1016/j.enconman.2023.117707

52. Grandon TG, Schwenzer J, Steens T, Breuing J (2023) Electricity demand forecasting with hybrid statistical and machine learning algorithms: Case study of ukraine, arXiv preprint arXiv:2304.05174 . https://doi.org/10.48550/arXiv.2304.05174

53. Tan P-N, Steinbach M, Kumar V (2016) Introduction to data mining, Pearson Education India

54. Mills TC (2019) Applied time series analysis: A practical guide to modeling and forecasting, Academic press

55. Chung MK (2020) Gaussian kernel smoothing, arXiv preprint arXiv:2007.09539 . https://doi.org/10.48550/arXiv.2007.09539

56. Fan G-F, Peng L-L, Hong W-C (2022) Short-term load forecasting based on empirical wavelet transform and random forest. Electrical Engineering 104(6):4433–4449

57. Mounir N, Ouadi H, Jrhilifa I (2023) Short-term electric load forecasting using an emd-bi-lstm approach for smart grid energy management system. Energy and Buildings 288:113022. https://doi.org/10.1016/j.enbuild.2023.113022

58. Liu D, Wang H (2023) Time series analysis model for forecasting unsteady electric load in buildings. Energy and Built Environment. https://doi.org/10.1016/j.enbenv.2023.07.003

59. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences 454 (1971) , 903–995

60. Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in adaptive data analysis 1(01):1–41

61. Colominas MA, Schlotthauer G, Torres ME (2014) Improved complete ensemble emd: A suitable tool for biomedical signal processing. Biomedical Signal Processing and Control 14:19–29

62. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555

63. Niu Z, Zhong G, Yu H (2021) A review on the attention mechanism of deep learning. Neurocomputing 452:48–62

64. Liashchynskyi P, Liashchynskyi P (2019) Grid search, random search, genetic algorithm: a big comparison for nas, arXiv preprint arXiv:1912.06059