



Research article

Multi-directional gated recurrent unit and convolutional neural network for load and energy forecasting: A novel hybridization

Fazeel Abid¹, Muhammad Alam², Faten S. Alamri³ and Imran Siddique^{4,*}

¹ Department of Information Systems, University of Management and Technology, Lahore 54770, Pakistan

² Department of Computer Science and Informatics, London South Bank University, UK

³ Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

⁴ Department of Mathematics, University of Management and Technology, Lahore 54770, Pakistan

* **Correspondence:** Email: imransmsrazi@gmail.com.

Abstract: Energy operations and schedules are significantly impacted by load and energy forecasting systems. An effective system is a requirement for a sustainable and equitable environment. Additionally, a trustworthy forecasting management system enhances the resilience of power systems by cutting power and load-forecast flaws. However, due to the numerous inherent nonlinear properties of huge and diverse data, the classical statistical methodology cannot appropriately learn this non-linearity in data. Energy systems can appropriately evaluate data and regulate energy consumption because of advanced techniques. In comparison to machine learning, deep learning techniques have lately been used to predict energy consumption as well as to learn long-term dependencies. In this work, a fusion of novel multi-directional gated recurrent unit (MD-GRU) with convolutional neural network (CNN) using global average pooling (GAP) as hybridization is being proposed for load and energy forecasting. The spatial and temporal aspects, along with the high dimensionality of the data, are addressed by employing the capabilities of MD-GRU and CNN integration. The obtained results are compared to baseline algorithms including CNN, Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), Gated Recurrent Unit (GRU), and Bidirectional Gated Recurrent Unit (Bi-GRU). The experimental findings indicate that the proposed approach surpasses conventional approaches in terms of accuracy, Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RSME).

Keywords: load and energy forecasting; multi-directional gated recurrent unit (MD-GRU); convolutional neural networks; spatial and temporal; high dimensionality; long term dependencies

Mathematics Subject Classification: 62J02, 62J99

1. Introduction

The energy-management functions and timing have been greatly influenced by load and power prediction. An effective load forecasting and energy management system is essential to a fair and sustainable system. Many variables, including cost, load flow and power plant demand, depend on data collected by the load and power forecasting system [1,2]. Large energy storage and demand fluctuations also require a proactive balancing of load changes and energy management systems, as shown in [3,4]. A reliable load and power forecasting management system help reduce power and load forecast errors, enhance the robustness of power systems and mitigate changes in supply and demand [5,6].

Conventional statistical approaches including linear regression analysis (LR) approach and autoregressive moving average (ARMA) methodology have been previously used in pregnancy prediction methods based on massive amounts of data [7,8]. However, traditional statistical techniques cannot adequately learn these nonlinear data or match pregnancy prediction accuracy criteria due to many intrinsic nonlinear properties of big data [9,10]. Modern artificial intelligence technology enables energy systems to analyze data and manage demand properly. Machine learning methods have recently been used to predict power consumption and load patterns. These approaches lead to better control of load and power resulting in lower energy use, longer resource life and lower processing costs [11–13].

Currently, neural networks are being used more and more to tackle complex time-series prediction problems as a result of their recent success in challenges associated with computer vision and image processing. By comparing actual usage with forecasts from the forecasting system, accuracy can be improved as well as money and energy savings. Scheduling and planning of power systems can be damaged by inaccurate load forecasting which can lead to power shortages in the market and eventually increase the desire to expand existing Artificial Neural Network (ANN) designs [14,15]. By adding more hidden layers, which may handle nonlinearities in the input and output data, the extension is produced. The goal is to provide a Deep Neural Network (DNN) based power load prediction model for power usage prediction.

Initially, many neural networks used backpropagation techniques for short-term load forecasting as proposed in [16–18]. Similarly, the processing of sequence data is done in many works using a recurrent neural network proposed in [19–21]. Based on climatological data, a local and Global RNN model was developed to address the issue of long-term wind speed and energy forecasting [22,23]. However, during error backpropagation, the gradient descent problem still exists due to the enormous depth of time in the traditional Recurrent Neural Network (RNN) architecture which cannot learn the long-term historical data. Comparatively, long short-term memory (LSTM), an improved form of RNN was introduced in [24] which uses gated mechanisms to integrate short-term and long-term memory.

Latterly, Gated Recurrent Unit (GRU), based on an optimized LSTM combines the input gate and forget LSTM gate into a single update gate [25,26]. Still, the GRU model is not enough, because, the power system additionally contains high-dimensional data such as image data, spatiotemporal matrices and sequence data. These optimized models are not sufficient to handle high-dimensional data

effectively. By contrast, a CNN is suitable for processing large amounts of high-dimensional data and has been successfully applied in many research works [27,28]. The CNN may record the characteristics of the local orientation and the non-variable features of the scale when there is a significant correlation between adjacent data points [29,30].

Accordingly, there are four categories of power load projections: long-term, medium-term, short-term and very short-term [31–34]. The main focus should be on very short-term electrical load and power forecasting, which predicts future loads and may assist power system operators set up realistic production plans, keeping supply and demand in balance, and assuring safety while minimizing resource waste and electricity prices. The conventional statistical methodology was proven to be ineffective for learning nonlinearity due to the multiple intrinsic nonlinear features of large and diverse data sets. Therefore, to predict energy consumption, machine and deep learning approaches have recently been applied to solve issues related with large dimensionality and learning long-term dependencies. A novel load and energy demand predictor using a multi-directional GRU (MD-GRU) and CNN using different pooling schemes has been proposed in this work. To improve the extraction and learning features in the chronological data, the MD-GRU is used to simulate the dynamic changes in the historical energy sequence data. The CNN module processes spatiotemporal matrices and plots them into distinct vectors by taking global values rather than max or average. The model predicts the load of power consumption 10 minutes in advance using several statistical techniques concerning predicted and actual values. These predicted values can be used to make wiser judgments and enhance the overall effectiveness of the system. The following are the main contributions of this paper:

1-A novel multi-directional GRU (MD-GRU) model is presented for the detection of dynamic changes in historical load and energy sequence data

2-This work integrates MD-GRU and CNN having global average pooling (GAP) is used to estimate the load and energy demand to capture long-term dependencies.

3-The proposed MD-GRU-CNN model is compared with the baseline models such as LSTM, GRU, Bi-LSTM, Bi-GRU, and CNN models

4-The proposed model is also assessed using other metrics including mean percentage absolute error (MAPE), root mean square error (RMSE), and conventional assessment measures.

The rest of the paper is structured as follows: Section 2 describes related work. Sections 3 and 4 contain the deep learning techniques for load and energy forecasting as well as the proposed methodology, while Section 5 contains the experimentation. Section 6 is for discussion, and Section 7 concludes the paper.

2. Related work

For the load and energy management system, there has been a major growth in the use of modern metering and sensors, which has greatly increased the information. A significant amount of data is produced, serving as a trustworthy source of information for load forecasting algorithms. There are parametric and non-parametric versions of these load forecasting models. While ANNs are among the most widely used non-parametric techniques for creating new energy demand predictions, parametric methodologies are based on analytical models [35,36]. There have been numerous electrical load forecasting models proposed in recent years using the latest deep learning techniques; some of the more popular models that researchers now employ are detailed in [37–39].

To forecast energy consumption over various time horizons, time-series approaches have been

frequently used. The most prevalent model is the autoregressive moving average (ARMA) model, which combines the moving average (MA) and autoregressive (AR) models. The AR Model relies on the linear regression of actual values depending on a database, while the MA Model makes utilization of the moving average to compute Regressions employing forecast errors for previous values [40]. Another study uses autoregressive integrated moving average (ARIMA) models to create novel forecasting systems [41]. However, there are numerous intrinsic nonlinear properties in the enormous data, and classic statistical approaches cannot effectively learn this nonlinear data even though these methods may, to some extent, estimate short-term demand. Therefore, it is exceedingly difficult for conventional statistical approaches to forecast load with any degree of accuracy.

Various machine learning (ML) techniques, such as artificial neural networks fuzzy systems [42] and support vector machines (SVM) [43] have been effectively implemented in predicting and classifying issues [44–46]. The development of a system for load and energy forecasting using SVM and ML techniques is effective in addressing the difficulties of nonlinearity [47]. An adaptive neuro-fuzzy inference system (ANFIS), was used in another project to create a short-term load and energy forecasting model [48]. The backpropagation neural network (BPNN) has also been taken into consideration for load forecasting as deep learning has advanced [49]. A hybrid model of BPNN and the multi-label algorithm based on K-nearest neighbor (KNN) and K-means was presented; however, because of the nature of feedforward neural networks, it is unable to effectively learn time-sequence data for the power system [50].

Recently, computer vision, natural language processing and speech recognition have all benefited greatly from recent advancements in neural networks, especially deep neural networks. Deep neural network applications for short-term load forecasting are currently a novel and widely discussed topic. An analysis of artificial neural networks (ANNs) for predicting short-term load is shown in [51]. Recurrent neural networks (RNN), convolutional neural networks (CNN), and their combination and long short-term memory (LSTM) are just a few of the various building blocks that have made deep neural networks highly adaptable and successful [52].

A novel approach with LSTM and genetic algorithm (GA) integration was put out in [53], and it produced a mean absolute percentage error with a minimal value. An innovative method for the planning and operation of energy generation called the Bi-GRU (gated recurrent unit) prediction system was created [54]. The bidirectional characteristic of GRU, in contrast to other approaches, is not enough to gather long-term sequential information from all areas of energy and load forecasting. By using positive and negative input, the bidirectional model trains the nodes to operate in forward and reverse modes. The advantage of the bidirectional ability to focus in both directions (forward and backward), as indicated is useful for extracting more information from the long-term information sequence, but it is insufficient for energy and load forecasting [55]. Multiple load and energy forecasters have been produced in which the forecaster's framework must get more intricate to produce forecasting accuracy.

3. Deep learning techniques for load and energy forecasting model

3.1. Recurrent neural network

Sequential information use is carried out via recurrent neural networks. These networks connect words with particular time steps in the sequence. [56,57]. The amount of time steps determines the

maximum input sequence length in the final result. Figure 1 depicts the recurrent neural network's architecture.

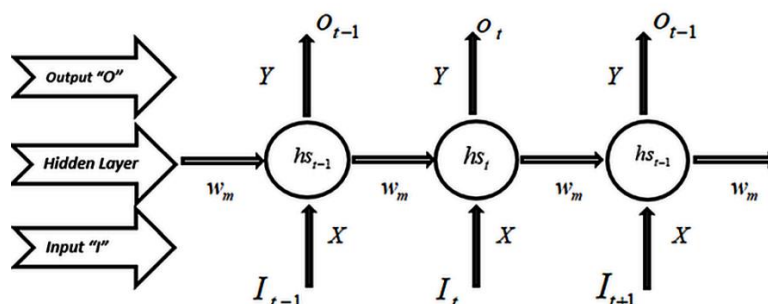


Figure 1. Recurrent neural network architecture.

The following are the sequential data as feature vectors that are provided to the recurrent neural network (RNN) as inputs after changing feeds:

$$hs_t = f(X \cdot I_t + W_m hs_{t-1} + b). \quad (1)$$

$$O_t = softmax(Y \cdot hs_t). \quad (2)$$

The outputs can be observed in the above image by the corresponding time step, however depending on the task, this may not be essential. Our objective for managing the energy and load is to concentrate on the output. A recurrent neural network can accumulate sequential information in the same situation when there is no necessity for input with the appropriate time step and hidden state. Previously, researchers discovered that by introducing LSTM, Bi-LSTM GRU, and Bi-GRU, the recurrent neural network can be improved for learning long term dependencies. However, the bidirectionality was insufficient in the case of load and energy management to achieve the desired outcomes [58].

3.2. Long short-term Memory and bidirectional module

Long-term dependency issues are addressed with the LSTM [59]. When a long short-term memory network is trained to use backpropagation with time, it can address the vanishing gradient problem [60]. Although LSTMs are considered to be an adequate methodology, still architecture only considers the past, which can sometimes result in poor performance, particularly in load and energy management where the future data is also crucial. With two hidden states, forward \overrightarrow{hs}_t and backward \overleftarrow{hs}_t , bidirectionality in LSTMs can capture both contexts: prior and upcoming.

3.3. Gated recurrent unit and bidirectional module

LSTM optimization known as the gated recurrent unit (GRU) performs better than conventional LSTMs. When compared to LSTMs, the gated recurrent unit (GRU) has fewer parameters since it consolidates the input gates and forget gates provided in [61] into only two gates. Similar to LSTMs,

but faster than Bi-LSTMs, is the bidirectional gated recurrent unit. The Bi-GRU unit can store sequences of information in both directions, such as prior and upcoming information, and have a long short-term memory for preserving long-term dependencies. Many models have been successful in terms of the informational flow relating in to both directions for load and energy management [62]. The output of a bidirectional recurrent module that processes sequences in opposite directions is, for example, the hidden state generated as:

$$[\overrightarrow{hs}_1 \parallel \overleftarrow{hs}_1, \overrightarrow{hs}_2 \parallel \overleftarrow{hs}_2, \dots, \overrightarrow{hs}_n \parallel \overleftarrow{hs}_n]. \quad (3)$$

3.4. Convolutional neural network

Convolutional neural networks were initially for image classification but a short while later, evolved into models that were used for a variety of applications. The sequential data is fed into a convolutional layer, which then generates various filters, learns several features, and applies these features successively to the various input divisions. The output is typically then aggregated to lower dimensions and routed into a linked layer [15]. Multiple features are used in this layer to extract local features. Figure 2 shows the traditional convolutional neural network architecture as it is described in [63].

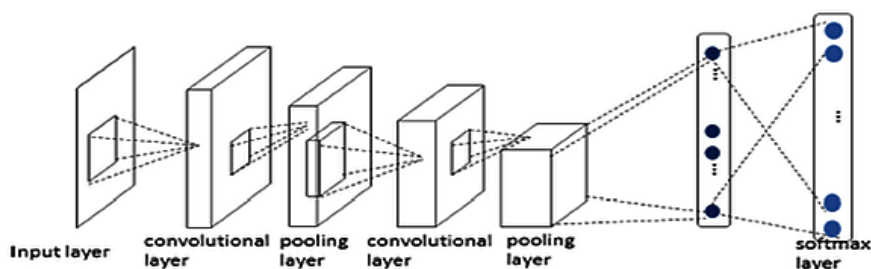


Figure 2. Traditional convolutional neural network architecture.

To choose a maximal value, the max pooling layer must first discard any other values that are not maximal. A concept of dropout layer normalization is essential to build a dense connection in both the max-pooling and convolutional layers as presented in [64].

The output of a convolutional layer is processed by a global average pooling layer, as a regularizer, replacing the fully connected layer and dropout. Further, it specifically coordinates between the mapping of features and classifications to examine the computational complexity and determine whether it is less or not. Additionally, no parameter modification is required in the global average pooling [65]. Each feature map's average value produces a unique value, which is then directly entered into the SoftMax algorithm to obtain the likelihood distribution along with confidence values for each class. The following expression illustrates the likelihood of appropriation:

$$P(i_j | k, \theta) = \frac{\exp(x_j(k, \theta))}{\sum_{1 \leq i \leq |X|} \exp(x_i(k, \theta))}, \quad 1 \leq j \leq |X|. \quad (4)$$

4. Deep learning techniques for load and energy forecasting model

For load and energy forecasting, deep learning techniques, particularly sequential models like RNN and LSTM models, are being used. Feedforward neural networks with internal memory are known as recurrent neural networks. The outcome of the current input depends on the computation of the prior inputs, as opposed to RNNs, which perform the same task for every data input. The output is produced, duplicated, and then distributed throughout the network. However, the issue of gradient vanishing and exploding prevents the typical RNN design from supporting long sequences. A modified version of recurrent neural networks makes it simpler to store past knowledge in memory like long short-term memory (LSTM) and gated recurrent units (GRUs) [66].

Categorical values require label encoding approaches since deep learning models are unable to understand them. These methods change these category values into numerical ones that are appropriate for deep-learning neural network models. Understanding how the extended input sequence is created, however, is crucial. The time period of the date, such as a certain day, month, etc., is the input data for energy and load management. As a result, it is necessary to transform this kind of data into a sequence that can be used as additional input in the architecture for load and energy forecasting.

4.1. Multidirectional gated recurrent unit

The fundamental principle of MD-GRU is to use as many recurrent connections as there are dimensions in the data to replace the one recurrent connection present in regular RNNs. The hidden layer of the network receives an external input and its activations from one step back along all dimensions at each point in the data sequence during the forward pass. This module has multiple dimensions that can individually scan all four directions in each column and row where hidden layers are calculated in the vertical and horizontal direction at every stage and added up at the end. Two output attribute maps are employed, one for the vertical direction and the other for the horizontal direction. The forward hidden states, backward hidden states, forward update gates, and backward update gates are the four sets of activations that the MD-GRU computes for each time step. The final output of the module is generated by combining these activations, which are calculated using a set of learned weights and biases. The current simulation makes use of the total of these output attribute maps as an alternative to concatenate the output feature maps as shown in Figure 3 below:

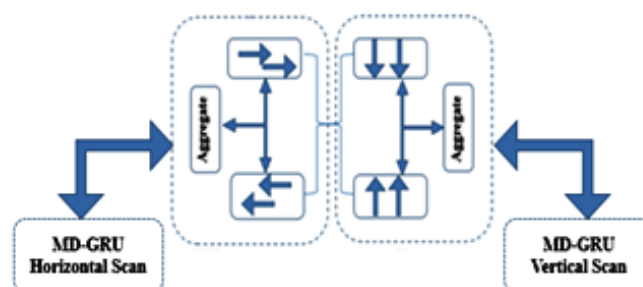


Figure 3. The architecture of multi-directional gated recurrent unit (MD-GRU).

4.2. MD-GRU-CNN module

A novel hybrid Multi-Directional MD-GRU-CNN is proposed by using the benefits of the combination of MD-GRU for the processing of time sequence data with CNN to handle high dimensionality of data. The spatial and temporal dependencies in the input data are intended to be captured by this hybrid architecture, making it particularly ideal for tasks that call for both short-term and long-term patterns.

Figure 4 illustrates the proposed hybrid multidimensional MD-GRU-CNN neural network design. With spatiotemporal matrices gathered from the power system as inputs and sequential data based on time as outputs, the proposed structure can forecast the load and energy level in a relatively very short amount of time. The MD-GRU module aims to capture long-term dependency. The MD-GRU module can learn useful information by scanning each column and row independently in different directions in the historical data for a long period of time through the memory cell, and the forget gate will forget the useless information. The inputs to the MD-GRU module are time sequence data. The MD-GRU module has a sizable number of gated recurrent units that are connected CNNs that make use of fully connected layers and global average pooling. While the MD-GRU can identify temporal patterns in the data, such as daily or weekly cycles in energy consumption or load, the CNN can identify spatial aspects of the data, like patterns in the distribution of load or energy consumption throughout an area. Finally, by calculating the mean value of each neuron in the fully connected layers, the load-predicting results may be determined as shown below in the Figure 4 below:

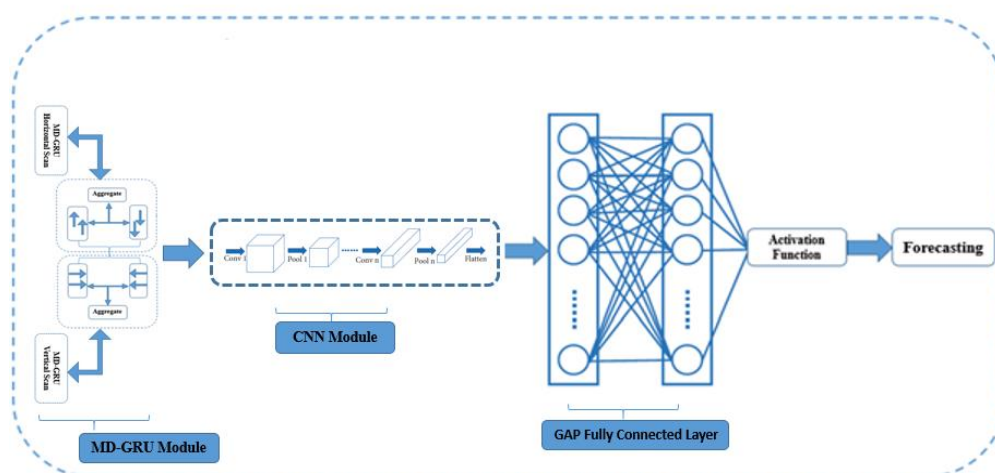


Figure 4. Proposed architecture based on hybridization of multi-directional GRU and CNN.

5. Experimentations

To evaluate our proposed methodologies, we run the system over the following datasets:

5.1. Dataset descriptions

The UCI machine learning repository is where the first dataset for the experimentation was

gathered [67]. The dataset contains 10,000 instances having 14 variables relating to the values of electricity producers, produced, price elasticity factor, the maximum value of equation root and system reliability as described refers to DS-1.

The second dataset called the Almanac of Minutely Power Dataset, contains different measurements such as (natural gas, electricity, and water) taken by a single household of 19 devices over the duration of an entire year, at a resolution of one (1) minute that is then down-sampled to a resolution of thirty (30) minutes [68] refers to DS-2.

The third dataset, referred to as the smart grid smart city was started in an initiative by Australian government [69] by collecting smart-meter data of 78,000 clients referred to as DS-3.

5.2. Metrics used for evaluating the proposed approach

Multiple metrics such as accuracy, precision, recall, and F1 Score are broadly applicable and significant for evaluating the performance of different prediction models. Additionally, the mean absolute percentage error (MAPE) and root mean square error (RMSE) are also added:

$$MAPE = \frac{1}{N} \sum_{i=1}^n \left| \frac{y(i) - \hat{y}(i)}{y(i)} \right| \times 100\%, \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^x \frac{(\hat{y}_i - y)^2}{x}} \quad (6)$$

where n refers to samples training size $y(i)$ actual and $\hat{y}(i)$ are expected values, respectively.

The average ratio between the error and actual values at one forecasting point is represented by the MAPE. The sample standard deviation of variations in predicted and actual value as RMSE. The performance of the model's prediction increases with decreasing MAPE and RMSE values.

5.3. Baseline methods

This part begins by presenting baseline methods, which consist of LSTM, bidirectional LSTM, GRU, bidirectional GRU, and CNN experiments with various time stamps. Then, by combining Multi-directional GRU with CNN and examining several variations, such as MD-GRU separately and MD-GRU-CNN, as shown in Tables 1 and 2, we demonstrate how we expand these baseline models toward the proposed architecture.

Table 1. Baseline models experimented with different timestamps (2, 4, 6).

Models	Narrative
CNN	
LSTM	Baseline Methods
GRU	with Different Time
Bi-LSTM	Stamps
Bi-GRU	

Table 2. Proposed models experimented with different timestamps (4, 8, and 12).

Models	Narrative
MD-GRU	Proposed Methods with Different Time Stamps
MD-GRU-CNN	

5.4. Experimental setup

The purpose of the testing in this work is to show the effectiveness of the recommended procedure for diversifications in different datasets, in comparison with the tuning of the multiple parameters. During our initial testing, we checked the epoch, batch size, filter windows, MAPE, and learning rate to determine the optimum parameter. However, this work does not go into in-depth detail concerning adjusting hyperparameters.

With a neural network, we can divide the huge dataset into smaller chunks by training with small batches of several training cases. Accuracy and layers are sometimes directly associated, implying that as the number of layers increases, so does accuracy. Most of the time, this tends to increase precision; but, as the number of layers increases, accuracy can substantially decline due to the "Vanishing Gradient Problem." [38]. Sometimes the computations also rely on how far the single hidden layer is expanded. For instance, if there are not enough hidden layers, the performance will suffer. The dimensions associated with the hidden state are taken into consideration to be 128 and one layer of convolution layers is used in this work along with a field size of (3, 3, and 5).

After evaluating our model and experimenting with other window filter combinations, we found that seven (7) is the best alternative. Additionally, the batch size is set at 128, the learning rate is set at 0.02 and the epochs maintain within 10 and 40 including all datasets. Dropout, however, was fixed at 0.5 for regularization in the baseline models. Following a series of studies, it is also discovered that the activation function hyperbolic ReLu is perfect in the convolutional layer for the best performance. Lastly, MAPE 0:95 was found appropriate.

5.5. Experimental results and analysis

For experimentation, a diversity of evaluation techniques is utilized, including accuracy, precision, recall, and F1-score for both baseline and proposed models to replace the stacking of multiple convolutional layers as well as the problems associated with bidirectional architecture. Average MAPE and RMSE values were also employed to support the findings. Further, several of the most popular neural network models, including CNN, LSTM, Bi-LSTM, GRU and Bi-GRU models were selected as baseline models to assess the effectiveness and robustness of the proposed MD-GRU-CNN model.

The features of the baseline and (proposed) MD-GRU-CNN models were acquired by learning the training datasets. The model was then validated using test datasets. Tables 3–6 contain the results considering the above evaluation metrics on multiple datasets referring to DS-1, 2, and DS-3. For baseline models, the most elevated accuracy is 90.99 % achieved Bi-GRU model using DS-3 can be seen in Figure 5. The most appropriate precision and recall also observed on similar models are 91.86% and 93.18% respectively as visualized in Figures 6 and 7. Whereas the F1-Score was 92.97% using DS-1 on the Bi-GRU model as depicted in Figure 8.

Table 3. Accuracy achieved on baseline models.

Accuracy			
Models	DS-1	DS-2	DS-3
CNN	82.4	85.7	84.6
LSTM	83.8	87.1	87.9
GRU	86.99	88.12	88.71
Bi-LSTM	87.8	89.27	90
Bi-GRU	89.67	92.8	90.99

Table 4. Precision on baseline models.

Models	DS-1	DS-2	DS-3
CNN	83.5	88.7	85.6
LSTM	85.8	90.21	87.9
GRU	89.77	90.97	89.45
Bi-LSTM	90	89.87	90.78
Bi-GRU	91.27	89.76	91.86

Table 5. Recall baseline models.

Recall			
Models	DS-1	DS-2	DS-3
CNN	85.3	87.6	86.5
LSTM	88.2	92.1	89.2
GRU	87.87	91.75	90.25
Bi-LSTM	92.8	90.71	90.02
Bi-GRU	93.17	87.06	93.18

Table 6. F1-Score on baseline models.

F1-Score			
Models	DS-1	DS-2	DS-3
CNN	86.2	88.6	83.5
LSTM	90.2	93.4	88.4
GRU	89.71	92.15	91.05
Bi-LSTM	91.99	89.84	91.72
Bi-GRU	92.97	89.62	92.08

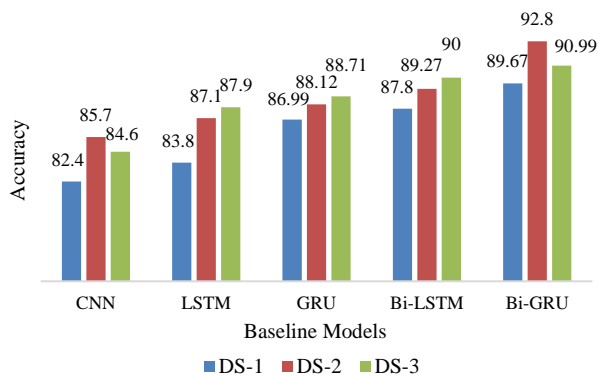


Figure 5. Accuracy on baseline models using DS-1, DS-2 and DS-3.

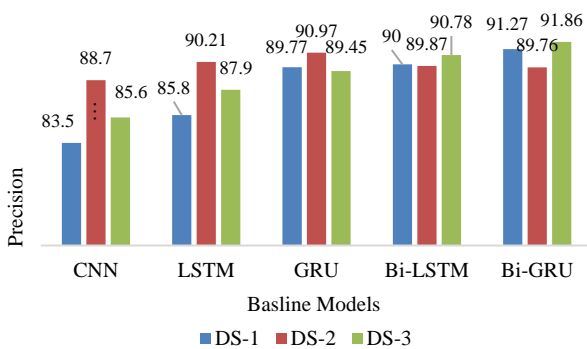


Figure 6. Precision on baseline models using DS-1, DS-2 and DS-3.

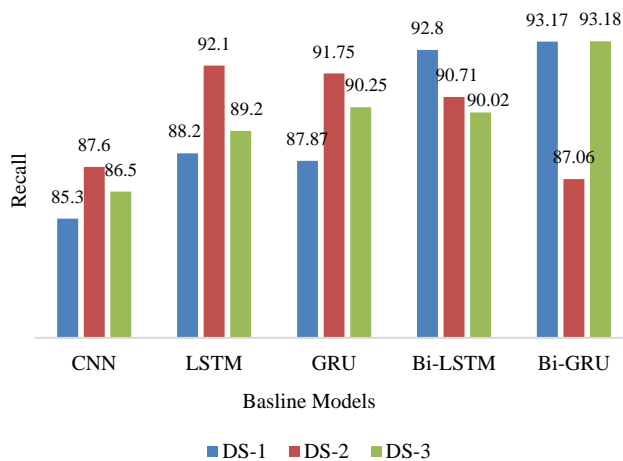


Figure 7. Recall on baseline models using DS-1, DS-2 and DS-3.

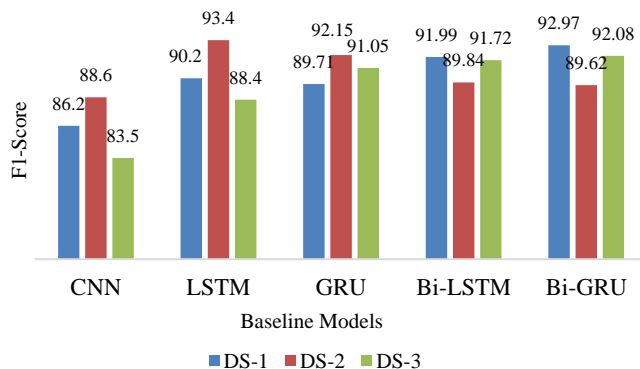


Figure 8. F1-Score on baseline models using DS-1, DS-2 and DS-3.

Similarly, the average MAPE attained on the Baseline method is also presented in Tables 7–9 using all three data sets (DS-1, DS-2, and DS-3) incorporating different time stamps (2, 4, and 6).

Table 7. Average MAPE on baseline model using DS-1.

DS-1 (Average MAPE)			
Models	2-Timestamps	4-Timestamps	6-Timestamps
CNN	40.5	39.5	41.5
LSTM	39.8	42.8	40.5
GRU	38.7	40.71	40.99
Bi-LSTM	37.8	39.1	38.99
Bi-GRU	37.9	35.4	35.01

Table 8. Average MAPE on baseline model using DS-2.

DS-2 (Average MAPE)			
Models	2-Timestamps	4-Timestamps	6-Timestamps
CNN	39.6	41.8	41.53
LSTM	38.41	40.3	38.55
GRU	38.34	39.91	37.56
Bi-LSTM	37.04	38.78	36.92
Bi-GRU	35.2	36.8	35.79

Table 9. Average MAPE on baseline model using DS-3.

DS-3 (Average MAPE)			
Models	2-Timestamps	4-Timestamps	6-Timestamps
CNN	38.6	42.4	41.03
LSTM	36.97	39.47	40.89
GRU	37.14	39.47	37.58
Bi-LSTM	37.04	38.14	39.87
Bi-GRU	36.2	35.49	35.17

Similarly, in baseline models, the least observed MAPE value is 35.01 on Bi-GRU using DS-1, 35.4 on 4 timestamps using DS-3, and 35.2 on DS-2 using 2 timestamps can be seen in Figures 9–11. By considering accuracy, precision, as well as recall and F1-score, Tables 10–13 demonstrate the effectiveness of the proposed models. The proposed model outperforms by achieving reliable 95.76% accuracy on MD-GRU and 97.85% using DS-3 compared to baseline models depicted in Figure 12.

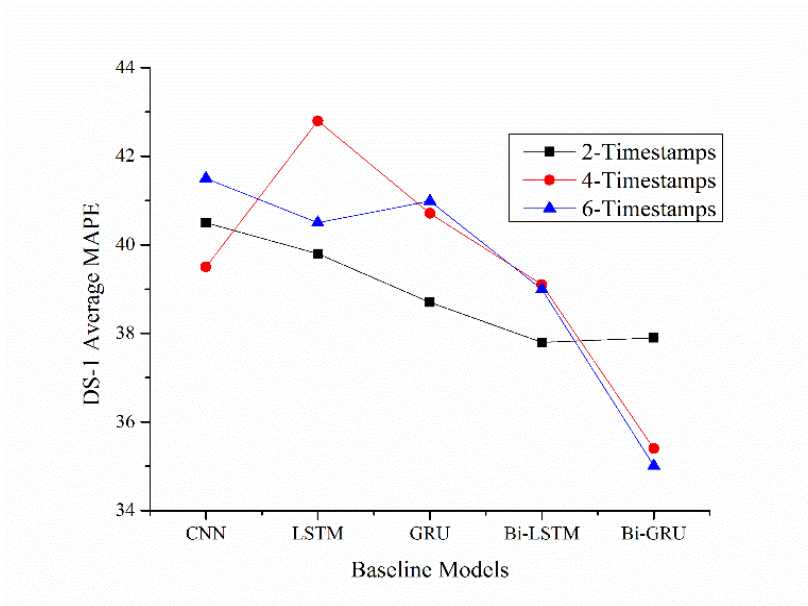


Figure 9. Average MAPE on baseline model using DS-1 with different timestamps.

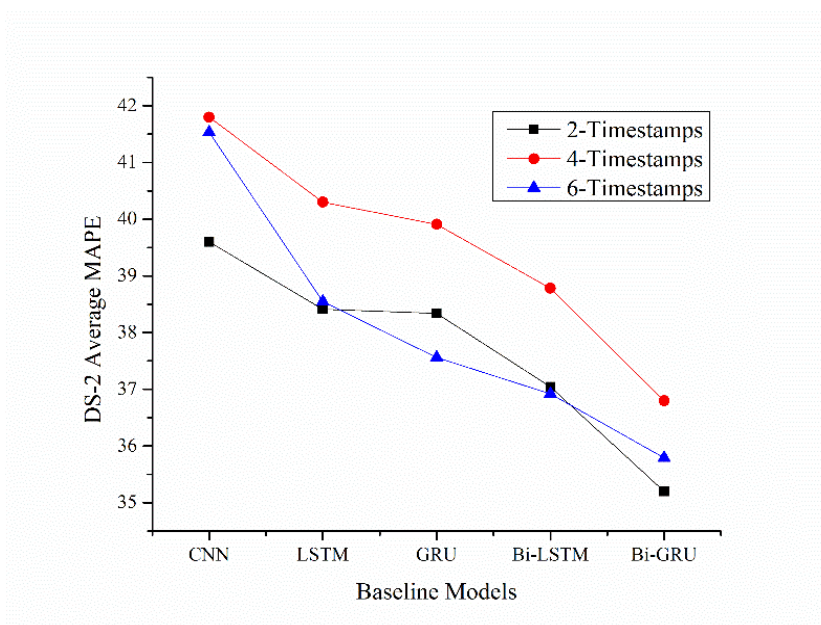


Figure 10. Average MAPE on baseline model using DS-2 with different timestamps.

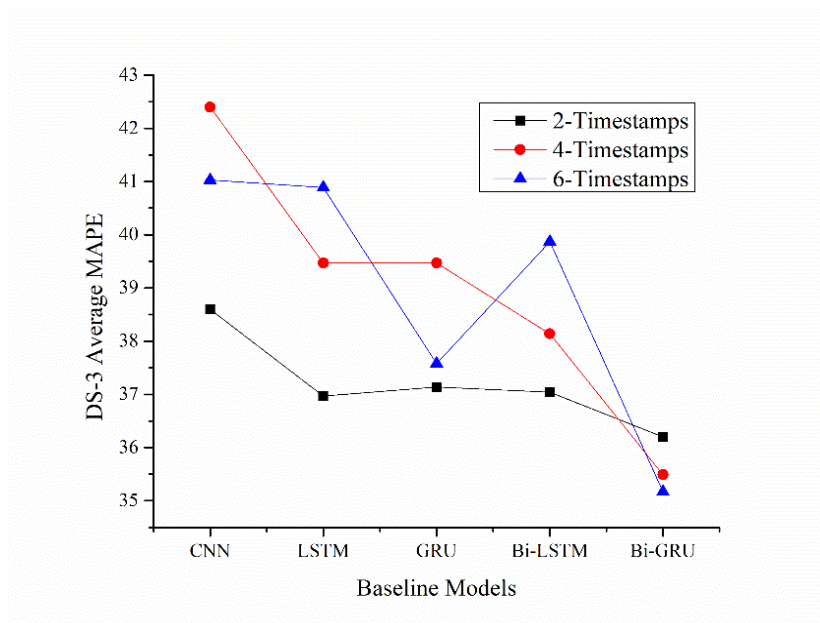


Figure 11. Average MAPE on baseline model using DS-3 with different timestamps.

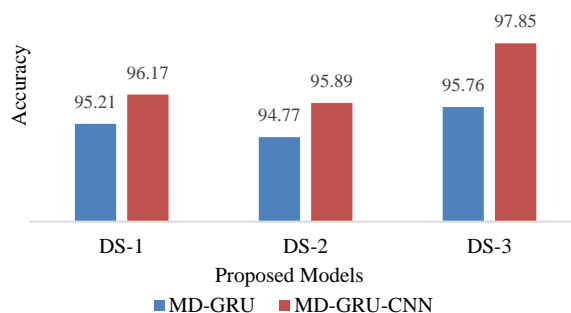


Figure 12. Accuracy was achieved on MD-GRU and MD-GRU-CNN using DS-1, DS-2 and DS-3.

Table 10. Accuracy achieved on proposed models using DS 1, DS-2 and DS-3.

Accuracy			
Models	DS-1	DS-2	DS-3
MD-GRU	95.21	94.77	95.76
MD-GRU-CNN	96.17	95.89	97.85

Table 11. Precision on proposed models using DS-1, DS-2 and DS-3.

Precision			
Models	DS-1	DS-2	DS-3
MD-GRU	96.11	95.83	95.61
MD-GRU-CNN	97.27	96.99	98.55

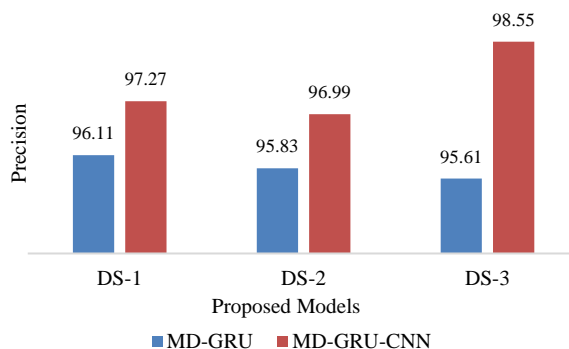
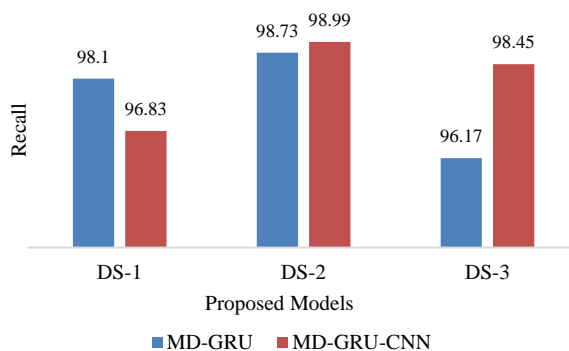
Table 12. Recall on proposed models using DS-1, DS-2 and DS-3.

Recall			
Models	DS-1	DS-2	DS-3
MD-GRU	98.1	98.73	96.17
MD-GRU-CNN	96.83	98.99	98.45

Table 13. F1-Score on proposed models using DS-1, DS-2 and DS-3.

F1-Score			
Models	DS-1	DS-2	DS-3
MD-GRU	97.01	97.83	95.84
MD-GRU-CNN	97.97	97.49	98.85

Figure 12 illustrates how the proposed models, MD-GRU and MD-GRU-CNN, fared better in terms of precision than earlier neural network baseline models employing DS-2 and DS-3, obtaining 95.83% and 98.55% higher values, respectively. On the DS-2, the greatest recall rates were 98.73% and 98.99%, as shown in Figures 13 and 14. Figure 15 indicates how the proposed model surpassed baseline models by gaining an F1-score on DS-3 that was also higher, at 98.85%.

**Figure 13.** Precision on MD-GRU and MD-GRU-CNN using DS-1, DS-2 and DS-3.**Figure 14.** Recall on MD-GRU and MD-GRU-CNN using DS-1, DS-2 and DS-3.

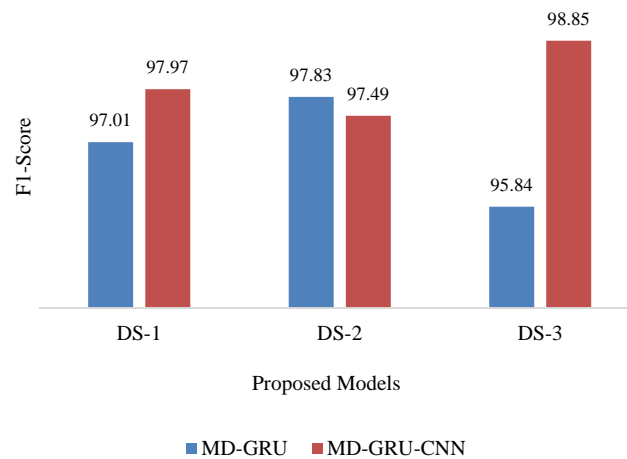


Figure 15. F1-Score on proposed models using different timestamps.

The proposed models (MD-GRU and MD-GRU-CNN) perform superior to standard architectures of CNN, LSTM, Bi-LSTM, GRU, and Bi-GRU (baseline), as demonstrated in the above tables. This is because the datasets used in this research are spatiotemporal. With the proposed hybridization (fusion of MD-GRU and CNN), there is no possibility of a loss of related information that usually occurs when flattening spatiotemporal data to dimensional and time sequence data. Resulting in the inability of recurrent and its variations models to learn the features effectively.

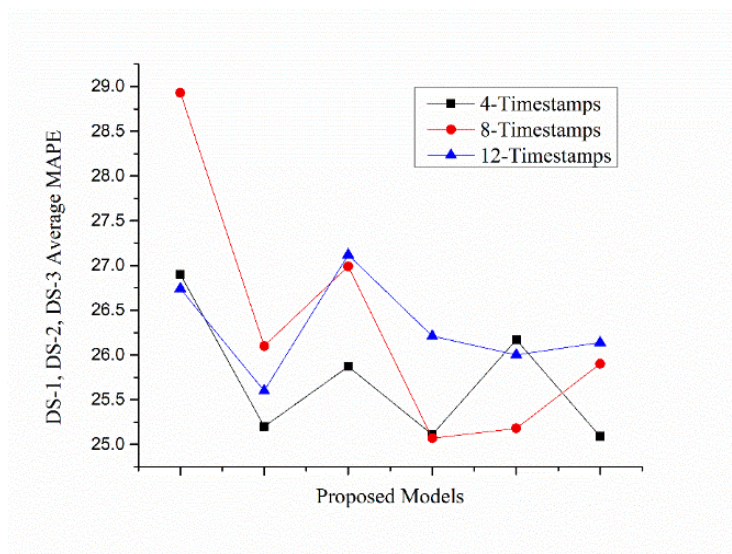
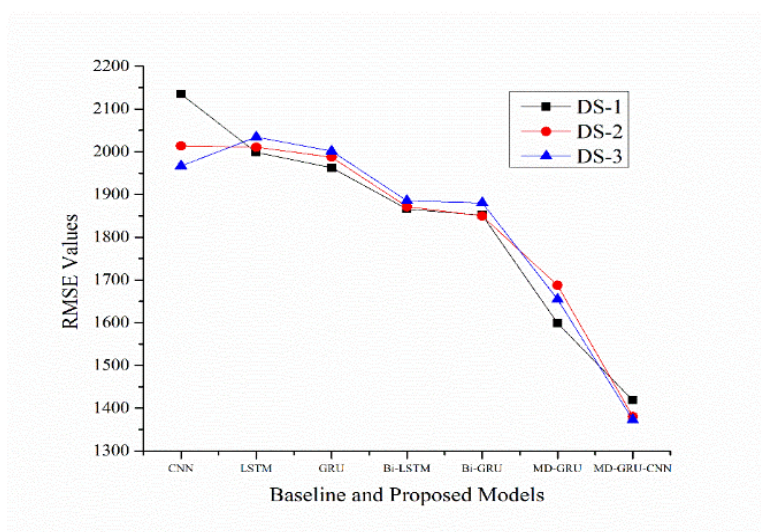
Also, the proposed models have the lowest average MAPE values out of all the baseline models shown in Table 14 and Figure 16. From the three datasets, our proposed hybrid model can effectively learn both the time sequence and spatiotemporal data and extract additional features. For a better assessment, the baseline and proposed models' RMSE representations are listed in Table 15 and illustrated in Figure 17. As shown in the mentioned figure, though the RMSE values of the Bi-GRU and Bi-LSTM models are nearly identical, the Bi-GRU model surpasses the GRU and Bi-LSTM models. The MD-GRU model values fluctuate between CNN, GRU, Bi-LSTM and MD-GRU-CNN revealing that the proposed model consistently surpasses the baseline models, while MD-GRU is lower in performance than the MD-GRU-CNN model.

Table 14. Average MAPE on proposed methods using DS-1, DS-2 and DS-3.

Models	DS-1, DS-2, DS-3 (Average MAPE)		
	4-Timestamps	8-Timestamps	12-Timestamps
MD-GRU	26.9	28.93	26.74
	25.2	26.1	25.6
	25.87	26.99	27.12
MD-GRU-CNN	25.11	25.07	26.21
	26.17	25.18	26
	25.09	25.9	26.14

Table 15. RMSE on baseline and proposed models using DS-1, DS-2 and DS-3.

Models	RMSE		
	DS-1	DS-2	DS-3
CNN	2134.45	2013.7	1966.75
LSTM	1998.5	2010.6	2034.45
GRU	1962.8	1987.44	2001.69
Bi-LSTM	1865.7	1871.4	1885.6
Bi-GRU	1851.4	1849.77	1880.5
MD-GRU	1598.6	1687.5	1654.36
MD-GRU-CNN	1418.5	1380.5	1372.96

**Figure 16.** Average MAPE on proposed methods using DS-1, DS-2 and DS-3 with 4, 8 and 12 timestamps.**Figure 17.** RMSE on baseline and proposed models using DS-1, DS-2 and DS-3 with different timestamps.

6. Discussion

The results presented in Section IV demonstrate that the proposed methodology consistently outperforms many previous works. Results reveal that the combination of multi-directional gated recurrent unit (MD-GRU) and convolutional neural network enhances the effectiveness of deep learning models. As opposed to this work, which examines a single deep learning architecture across many diverse datasets from various sources referred to as DS-1, DS-2 and DS-3, many earlier efforts typically provide findings on a single dataset. Additionally, it has been established that all sequential models that have bidirectional behavior and multidirectional abilities dominate traditional sequential models including GRU and LSTM. When combining several datasets to forecast load and energy, the integration of MAPE and RMSE is crucial.

Additionally, it has been found that many machine learning approaches are insufficiently equipped to handle categorical variables. In this scenario, the categorical values should always be transformed into statistical values by using a label-learning algorithm. The initial version of this study presented a multi-directional GRU (MD-GRU) that employs temporal GRUs to autonomously analyze every row and column in different positions. Through each level, hidden layers are figured out and added collectively. In the vertical position, two of the four temporal GRUs are employed. The use of Multidirectional GRUs, as depicted in Figure 2, results from the other Temporal GRUs' use in vertical directions. Because of the computational complexity and parameter optimizations required to learn long-term dependencies, other variants such as LSTM, Bi-LSTM, GRU, and Bi-GRU are insufficient. We conducted multiple experiments to compare the effectiveness of proposed models to baseline models on similar datasets as shown in the above tables using multiple evaluation metrics such as MAPE (Loss Function) and RMSE values.

Secondly, the integration of two neural networks that utilize the multi-directional GRU-CNN model with global average pooling (GAP) proposed in this research improves the accuracy. The multidirectional GRU (MD-GRU) module is targeted at analyzing the time sequence data to learn long-term dependencies, whereas CNN focuses on processing the spatiotemporal data. The convolution and global average pooling layer are employed effectively in this study to directly extract local features from spatiotemporal data and provide effective representation.

A global average pooling layer replaces the conventional max-pooling layer that is applied to the output produced by a convolutional layer and also replaces the fully connected layer and dropout since it acts as a regularizer to test whether the computational complexity is reduced or not. Additionally, it precisely correlates between feature mapping and classifications. Rather than including all of the layers of the particular feature map, we emphasized the average value related to each feature map. By separating characteristics from changeable data that affect the power system, the proposed models quickly and correctly forecast energy and load. The MAPE and RMSE values for the MD-GRU-CNN model are the lowest.

This work has the potential to be expanded in a variety of ways. For each dataset, hyperparameter tuning and optimization can be examined in conjunction with other parameters such as weather conditions, climate changes, and holidays. Simultaneously, to train the deep neural network more effectively as well as to shorten the model's training time new techniques other than MAPE and RSME are required. Finally, data augmentation in the time-series domain and advanced learning techniques can assist to address the issue of data limitation.

7. Conclusions and future work

Accurate forecasting of energy and load is essential for improving energy systems and their functionality. These are now producing a large amount of data that is complex, diverse, and of a spatiotemporal nature as a result of digitalization. To learn long-term dependencies linked to sequential data, most classical and deep learning techniques fall short. We introduced two novel deep learning architectures based on consecutive layers that integrate multi-directional (vertical and horizontal) GRU with convolutional neural network employing GAP as hybridization. This fusion uses MD-GRU to extract feature vectorization from time sequence data, whereas CNN extracts high-dimensional data. Multiple datasets were used to train and test the model, and several measures including accuracy, precision, recall and F1-Score, were used to assess its performance. Additionally, the MD-GRU-CNN model produced the lowest MAPE and RMSE values across baseline models for CNN, LSTM, GRU, Bi-LSTM, and Bi-GRU. In the future, similar work can be used for other diverse applications such as image dephasing towards scenically improving the degraded visibility.

Acknowledgment

This research was funded by Princess Nourah bint Abdulrahman University and Researchers Supporting Project number (PNURSP2023R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

All authors declare no conflicts of interest in this paper.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

References

1. Y. Lu, G. Wang, A load forecasting model based on support vector regression with whale optimization algorithm, *Multimed Tools Appl.*, **82** (2023), 9939–9959. <https://doi.org/10.1007/s11042-022-13462-2>
2. H. Habbak, M. Mahmoud, K. Metwally, M. M. Fouda, M. I. Ibrahim, Load forecasting techniques and their applications in smart grids, *Energies*, **16** (2023), 1480. <https://doi.org/10.3390/en16031480>
3. L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, et al., A review of machine learning in building load prediction, *Appl. Energy*, **285** (2021), 116452. <https://doi.org/10.1016/j.apenergy.2021.116452>
4. M. Zulfiqar, M. Kamran, M. B. Rasheed, T. Alquthami, A. H. Milyani, A short-term load forecasting model based on self-adaptive momentum factor and wavelet neural network in smart grid, *IEEE Access*, **10** (2022), 77587–77602. <https://doi.org/10.1109/ACCESS.2022.3192433>

5. R. Liu, T. Chen, G. Sun, S. M. Muyeen, S. Lin, Y. Mi, Short-term probabilistic building load forecasting based on feature integrated artificial intelligent approach, *Electr. Pow. Syst. Res.*, **206** (2022), 107802. <https://doi.org/10.1016/j.epsr.2022.107802>
6. I. Yazici, O. F. Beyca, D. Delen, Deep-learning-based short-term electricity load forecasting: A real case application, *Eng. Appl. Artif. Intell.*, **109** (2022), 104645. <https://doi.org/10.1016/j.engappai.2021.104645>
7. A. Goia, C. May, G. Fusai, Functional clustering and linear regression for peak load forecasting, *Int. J. Forecast.*, **26** (2010), 700–711. <https://doi.org/10.1016/j.ijforecast.2009.05.015>
8. A. H. Nury, K. Hasan, A. M. J. Bin, Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh, *J. King, Saud. Univ. Sci.*, **29** (2017), 47–61. <https://doi.org/10.1016/j.jksus.2015.12.002>
9. G. Y. Chen, M. Gan, G. L. Chen, Generalized exponential autoregressive models for nonlinear time series: Stationarity, estimation and applications, *Inf. Sci.*, **438** (2018), 46–57. <https://doi.org/10.1016/j.ins.2018.01.029>
10. S. Deng, F. Chen, X. Dong, G. Gao, X. Wu, Short-term load forecasting by using improved GEP and abnormal load recognition, *ACM Trans. Inter. Technol.*, **21** (2021), 1–28. <https://doi.org/10.1145/3447513>
11. J. Lee, Y. Cho, National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? *Energy*, **239** (2022), 122366. <https://doi.org/10.1016/j.energy.2021.122366>
12. T. Alquthami, M. Zulfiqar, M. Kamran, A. H. Milyani, M. B. Rasheed, A performance comparison of machine learning algorithms for load forecasting in smart grid, *IEEE Access*, **10** (2022), 48419–48433. <https://doi.org/10.1109/ACCESS.2022.3171270>
13. Z. Li, J. Wang, J. Huang, M. Ding, Development and research of triangle-filter convolution neural network for fuel reloading optimization of block-type HTGRs, *Appl. Soft Comput.*, **136** (2023), 110126. <https://doi.org/10.1016/j.asoc.2023.110126>
14. S. Deng, F. Chen, D. Wu, Y. He, H. Ge, Y. Ge, Quantitative combination load forecasting model based on forecasting error optimization, *Comput. Elec Engin.*, **101** (2022), 108125. <https://doi.org/10.1016/j.compeleceng.2022.108125>
15. S. Sun, Y. Liu, Q. Li, T. Wang, F. Chu, Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks, *Energy Convers. Manage.*, **283** (2023), 116916. <https://doi.org/10.1016/j.enconman.2023.116916>
16. Z. Xiao, S. J. Ye, B. Zhong, C. X. Sun, Short term load forecasting using neural network with rough set, Conference: Advances in Neural Networks-ISNN 2006, Third International Symposium on Neural Networks, Chengdu, China, May 28-June 1, 2006, Proceedings, Part II. https://doi.org/10.1007/11760023_183
17. C. X. Li, D. X. Niu, L. M. Meng, Rough set combine BP neural network in next day load curve forecasting, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5264 LNCS: 2008, 1–10. https://doi.org/10.1007/978-3-540-87734-9_1
18. Z. Xiao, S. J. Ye, B. Zhong, C. X. Sun, BP neural network with rough set for short term load forecasting, *Expert Syst. Appl.*, **36** (2009), 273–279. <https://doi.org/10.1016/j.eswa.2007.09.031>
19. D. Yi, S. Bu, I. Kim, An Enhanced Algorithm of RNN Using Trend in Time-Series, *Symmetry*, **11** (2019), 912. <https://doi.org/10.3390/sym11070912>

20. V. Kusuma, A. Privadi, A. L. S. Budi, V. L. B. Putri, Photovoltaic Power Forecasting Using Recurrent Neural Network Based on Bayesian Regularization Algorithm. *ICPEA 2021-2021 IEEE International Conference in Power Engineering Application*, (2021), 109–114. <https://doi.org/10.1109/ICPEA51500.2021.9417833>
21. G. Li, H. Wang, S. Zhang, J. Xin, H. Liu, Recurrent neural networks based photovoltaic power forecasting approach, *Energies*, **12** (2019), 2538. <https://doi.org/10.3390/en12132538>
22. A. Buonanno, M. Caliano, A. Pontecorvo, G. Sforza, M. Valenti, G. Graditi, Global vs. local models for short-term electricity demand prediction in a Residential/Lodging scenario, *Energies*, **15** (2022), 2037. <https://doi.org/10.3390/en15062037>
23. R. Quan, Z. Li, P. Liu, Y. Li, Y. Chang, H. Yan, Minimum hydrogen consumption-based energy management strategy for hybrid fuel cell unmanned aerial vehicles using direction prediction optimal foraging algorithm, *Fuel Cells*, **23** (2023), 221–236. <https://doi.org/10.1002/fuce.202200121>
24. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
25. J. Chung, C. Gulcehre, K. H. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, *NIPS 2014 Deep Learning and Representation Learning Workshop*, 2014. <https://doi.org/10.48550/arXiv.1412.3555>
26. A. K. Tyagi, N. Sreenath, Cyber physical systems: Analyses, challenges and possible solutions, *Int. Thing. Cyber-Physical Syst.*, **1** (2021), 22–33. <https://doi.org/10.1016/j.iotcps.2021.12.002>
27. J. Moon, S. Park, S. Rho, E. Hwang, A comparative analysis of artificial neural network architectures for building energy consumption forecasting, *Int. J. Distrib. Sens. N.*, **15** (2019). <https://doi.org/10.1177/1550147719877616>
28. T. Walser, A. Sauer, Typical load profile-supported convolutional neural network for short-term load forecasting in the industrial sector, *Energy AI.*, **5** (2021), 100104. <https://doi.org/10.1016/j.egyai.2021.100104>
29. X. Ke, L. Shi, W. Guo, D. Chen, Multi-Dimensional traffic congestion detection based on fusion of visual features and convolutional neural network, *IEEE T. Intell. Transp.*, **20** (2019), 2157–2170. http://www.ieee.org/publications_standards/publications/rights/index.html
30. P. H. Kuo, C. J. Huang, A high precision artificial neural networks model for short-term energy load forecasting, *Energies*, **11** (2018) 213. <https://doi.org/10.3390/en11010213>
31. J. Walther, D. Spanier, N. Panten, E. Abele, Very short-term load forecasting on factory level—A machine learning approach, *Procedia CIRP*, **80** (2019), 705–710. <https://doi.org/10.1016/j.procir.2019.01.060>
32. T. Hong, J. Wilson, J. Xie, Long term probabilistic load forecasting and normalization with hourly information, *IEEE T. Smart Grid*, **5** (2014), 456–462. <https://doi.org/10.1109/TSG.2013.2274373>
33. B. M. Hodge, D. Lew, M. Milligan, Short-term load forecast error distributions and implications for renewable integration studies, *IEEE Green Technologies Conference*, (2013), 435–442. <https://doi.org/10.1109/GreenTech.2013.73>
34. H. M. Al-Hamadi, S. A. Soliman, Long-term/mid-term electric load forecasting based on short-term correlation and annual growth, *Electr. Pow. Syst. Res.*, **74** (2005), 353–361.
35. X. Sun, Z. Ouyang, D. Yue, Short-term load forecasting model based on multi-label and BPNN. *Comm. Comp. Infor. Sci.*, **761** (2017), 263–272. https://doi.org/10.1007/978-981-10-6370-1_26

36. W. Tang, F. He, Y. Liu, YDTR: Infrared and visible image fusion via Y-shape dynamic transformer, *IEEE T. Multimedia*, (2022), 1–16. <https://doi.org/10.1109/TMM.2022.3192661>
37. A. A. Peñaloza, R. C. Leborgne, A. Balbinot, Comparative analysis of residential load forecasting with different levels of aggregation, *Eng. Proc*, **18** (2022), 29. <https://doi.org/10.3390/engproc2022018029>
38. T. Bashir, C. Haoyong, M. F. Tahir, Z. Liqiang, Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN, *Energy Rep.*, **8** (2022), 1678–1686. <https://doi.org/10.1016/j.egy.2021.12.067>
39. Y. Song, F. He, Y. Duan, Y. Liang, X. Yan, A kernel correlation-based approach to adaptively acquire local features for learning 3D point clouds, *Comput. Aided Design*, **146** (2022), 103196. <https://doi.org/10.1016/j.cad.2022.103196>
40. A. H. Nury, K. Hasan, M. J. B. Alam, Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh, *J. King Saud. Univ. Sci.*, **29** (2017), 47–61. <https://doi.org/10.1016/j.jksus.2015.12.002>
41. C. M. Lee, C. N. Ko, Short-term load forecasting using lifting scheme and ARIMA models, *Expert Syst. Appl.*, **38** (2011), 5902–5911. <https://doi.org/10.1016/j.eswa.2010.11.033>
42. A. Baliyan, K. Gaurav, S. K. Mishra, A Review of short term load forecasting using artificial neural network models, *Procedia Comput. Sci.*, **48** (2015), 121–125. <https://doi.org/10.1016/j.procs.2015.04.160>
43. J. P. Liu, C. L. Li, The short-term power load forecasting based on sperm whale algorithm and wavelet least square support vector machine with DWT-IR for feature selection, *Sustainability*, **9** (2017), 1188. <https://doi.org/10.3390/su9071188>
44. A. Jadidi, R. Menezes, N. D. Souza, A. C. D. C. Lima, Energies, E. Sciubba, Short-term electric power demand forecasting using NSGA II-ANFIS model, *Energies*, **12** (2019), 1891. <https://doi.org/10.3390/en12101891>
45. J. Zhang, F. He, Y. Duan, Y. Duan, S. Yang, AIDEDNet: Anti-interference and detail enhancement dehazing network for real-world scenes, *Front Comput. Sci.*, **17** (2023), 1–11. <https://doi.org/10.1007/s11704-022-1523-9>
46. S. Zhang, F. He, DRCND: learning deep residual convolutional dehazing networks, *Visual Comput.*, **36** (2020), 1797–1808. <https://doi.org/10.1007/s00371-019-01774-8>
47. D. Niu, Y. Wang, D. D. Wu, Power load forecasting using support vector machine and ant colony optimization, *Expert Syst. Appl.*, **37** (2010), 2531–2539. <https://doi.org/10.1016/j.eswa.2009.08.019>
48. H. H. Çevik, M. Çunkaş, Short-term load forecasting using fuzzy logic and ANFIS, *Neural Comput. Appl.*, **26** (2015), 1355–1367. <https://doi.org/10.1007/s00521-014-1809-4>
49. G. Li, H. Wang, S. Zhang, J. Xin, H. Liu, Recurrent neural networks based photovoltaic power forecasting approach, *Energies*, **12** (2019), 2538. <https://doi.org/10.3390/en12132538>
50. X. Xiong, P. Zhou, C. Ailian, Asymptotic normality of the local linear estimation of the conditional density for functional time-series data, *Commun. Stat. Theory Meth.*, **47** (2017), 3418–3440. <https://doi.org/10.1080/03610926.2017.1359292>
51. Deep Learning, Available from: <https://mitpress.mit.edu/9780262035613/deep-learning/>.
52. N. Ahmad, Y. Ghadi, M. Adnan, M. Ali, Load forecasting techniques for power system: Research challenges and survey, *IEEE Access*, **10** (2022), 71054–71090. <https://doi.org/10.1109/ACCESS.2022.3187839>

53. A. S. Santra, J. L. Lin, Integrating long short-term memory and genetic algorithm for short-term load forecasting, *Energies*, **12** (2019), 2040. <https://doi.org/10.3390/en12112040>
54. W. Li, T. Logenthiran, W. L. Woo, Multi-GRU prediction system for electricity generation's planning and operation, *IET Gener. Transm. Dis.*, **13** (2019), 1630–1637. <https://doi.org/10.1049/iet-gtd.2018.6081>
55. X. Gao, X. Li, B. Zhao, W. Ji, X. Jing, Y. He, Short-term electricity load forecasting model based on EMD-GRU with feature selection, *Energies*, **12** (2019), 1140. <https://doi.org/10.3390/en12061140>
56. T. Mikolov, M. Karafiát, L. Burget, J. H. Cernocky, S. Khudanpur, Recurrent neural network based language model, Conference: INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26–30, 2010
57. H. Salehinejad, S. Sankar, J. Barfett, E. Colak, S. Valaee, Recent advances in recurrent neural networks, *Neural Evolu. Comput.*, (2018), 1–21. <https://doi.org/10.48550/arXiv.1801.01078>
58. M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, *IEEE T. Signal Proces.*, **45** (1997), 2673–2681. <https://doi.org/10.1109/78.650093>
59. T. Mikolov, S. Kombrink, L. Burget, J. Černocký, S. Khudanpur, Extensions of recurrent neural network language model, *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, (2011), 5528–5531. <https://doi.org/10.1109/ICASSP.2011.5947611>
60. A. G. Ororbia, T. Mikolov, D. Reitter, Learning simpler language models with the differential state framework, *Neural Comput.*, **29** (2017), 3327–3352. https://doi.org/10.1162/neco_a_01017
61. S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, *Int. J. Uncertain Fuzz.*, **6** (1998), 107–116. <https://doi.org/10.1142/S0218488598000094>
62. B. Y. Lin, F. F. Xu, Z. Luo, K. Zhu, Multi-channel BiLSTM-CRF model for emerging named entity recognition in social media, *Proceedings of the 3rd Workshop on Noisy User-generated Text*, Stroudsburg, PA, USA, Association for Computational Linguistics, (2018), 160–165. <https://doi.org/10.18653/v1/W17-4421>
63. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, (2015), 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
64. K. He, J. Sun, Convolutional neural networks at constrained time cost, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2015), 5353–5360. <https://doi.org/10.48550/arXiv.1412.1710>
65. F. Abid, M. Alam, M. Yasir, C. Li, Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter, *Future Generation Computer Systems*, **95** (2019), 292–308. <https://doi.org/10.1016/j.future.2018.12.018>
66. S. Wang, J. Jiang, Learning natural language inference with LSTM, *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016-Proceedings of the Conference*, (2016), 1442–1451. <https://doi.org/10.18653/v1/N16-1170>
67. N. F. F. da Silva, E. R. Hruschka, E. R. Hruschka Jr., Tweet sentiment analysis with classifier ensembles, *Decis. Support Syst.*, **66** (2014), 170–179. <https://doi.org/10.1016/j.dss.2014.07.003>

68. S. Makonin, F. Popowich, L. Bartram, B. Gill, I. V. Bajić, AMPds: A public dataset for load disaggregation and eco-feedback research, *2013 IEEE Electrical Power & Energy Conference*, Halifax, NS, Canada, (2013) 1–6. <https://doi.org/10.1109/EPEC.2013.6802949>
69. Smart-Grid Smart-City Customer Trial Data |Datasets| data.gov.au-beta Available from: <https://data.gov.au/dataset/ds-dga-4e21dea3-9b87-4610-94c7-15a8a77907ef/details>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)