WILEY | Hindawi

Research Article

Deep Learning-Assisted Short-Term Power Load Forecasting Using Deep Convolutional LSTM and Stacked GRU

Fath U Min Ullah [iD],[1] **Amin Ullah** [iD],[2] **Noman Khan** [iD],[1] **Mi Young Lee** [iD],[1] **Seungmin Rho** [iD],[3] **and Sung Wook Baik** [iD][1]

[1]*Sejong University, Seoul 143-747, Republic of Korea*
[2]*Collaborative Robotics and Intelligent Systems (CoRIS) Institute, Oregon State University, Corvallis, OR, USA*
[3]*Department of Industrial Security, Chung-Ang University, Seoul 06974, Republic of Korea*

Correspondence should be addressed to Sung Wook Baik; sbaik@sejong.ac.kr

Over the decades, a rapid upsurge in electricity demand has been observed due to overpopulation and technological growth. The optimum production of energy is mandatory to preserve it and improve the energy infrastructure using the power load forecasting (PLF) method. However, the complex energy systems' transition towards more robust and intelligent system will ensure its momentous role in the industrial and economical world. The extraction of deep knowledge from complex energy data patterns requires an efficient and computationally intelligent deep learning-based method to examine the future electricity demand. Stand by this, we propose an intelligent deep learning-based PLF method where at first the data collected from the house through meters are fed into the pre-assessment step. Next, the sequence of refined data is passed into a modified convolutional long short-term memory (ConvLSTM) network that captures the spatiotemporal correlations from the sequence and generates the feature maps. The generated feature map is forward propagated into a deep gated recurrent unit (GRU) network for learning, which provides the final PLF. We experimentally proved that the proposed method revealed promising results using mean square error (MSE) and root mean square error (RMSE) and outperformed state of the art using the competitive power load dataset.(Github Code). (Github code: https://github.com/FathUMinUllah3797/ConvLSTM-Deep_GRU).

1. Introduction

Over the decade, the global energy consumption by the large-scale machinery in factories, buildings, and transport has remarkably increased due to population growth and economic development [1, 2]. This phenomenon rapidly shifted the energy resource demands towards clean power generation and its system improvement through intelligent methods for its efficiency [3]. These days, different renewable energy resources such as solar, wind, etc., are becoming the most optimal and significant resources aiming towards green technology; therefore, an extra layer of PLF will further assist the smart grid operation and its smooth maintenance [4]. Still, there exist some challenges for energy scientists to precisely establish an accurate and smart cooperative platform between the smart grids and the consumer side. A large amount of power energy is consumed and wasted due to improper infrastructure. Therefore, forecasting this power energy is an essential and imperative step towards optimal usage to overcome its dissipation. This will also enhance its future demand through smart grid and renewable energy production [5]. Researchers and data scientists are developing efficient ways to handle energy wastage and improve its optimal usage through different machine learning and time-series modeling techniques. However, a large amount of work has been done so far, with accurate results or with some uncertainties yielding erroneous forecasting that raised the need of establishing a highly precise, generalized ability, and robust energy forecasting model. According to [6], the scenario of generating
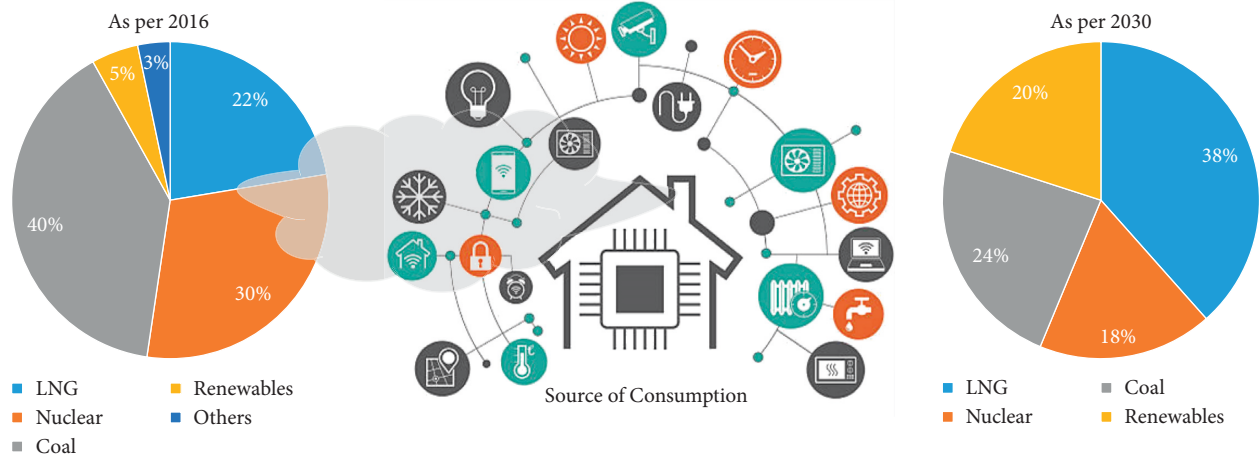
FIGURE 1: Statistics of power energy generation [6] as of 2016 and 2030. Also, different sources of power energy consumption are stepped.

the power energy from 2016 is 40%, 30%, 22%, 5%, and 3% for coal, nuclear, liquefied natural gas (LNG), renewables, and other resources, respectively. This statistic is illustrated in Figure 1, which shows the percentage of power energy generation and the consumption of resources.

Energy consumption via machinery in factories, buildings, and transport has remarkably increased over the last decades in the world due to population growth and the development in the economy. The prediction of this energy consumption is essential to reserve it for optimal use to overcome its dissipation and enhance its future demand, which is especially the current need of several countries. Researchers and scientists are developing efficient ways to handle energy wastage and improve its optimal usage in industries and residential areas. However, a large amount of work has been done so far, with accurate results or with some uncertainties. A challenge was raised to establish a highly precise, generalized ability, and robust energy prediction model. Mainly three kinds of building energy consumption models are used such as data-driven models, physical models, and hybrid models [7, 8]. Among these models, the most provocative is the data-driven model that became a popular method owning lower time in consumption with good performance. Several data-driven approaches are appropriate to cluster the buildings (residential [9] and nonresidential buildings [10]) with different time-scales such as short term [11] or long term [12]. A majority of prediction methods have been proposed over the past years in building ECP. Therefore, it is important to predict the future energy consumption and manage the energy usage accordingly. This method is a step towards efficient energy consumption. Also, it is an emerging field where the future world is widely based on energy and its utilization in industries, companies, government organizations, etc.

The utmost goal of the proposed method is the reduction of energy consumption and to ensure its efficient usage that is a prominent factor influencing the economy growth in the country. However, there exist numerous challenges in power load forecasting in buildings such as accuracy, efficient data processing, model evaluation, errors calculation, etc.

Therefore, it is important to develop a method that can achieve a fast performance to forecast and assess power energy infrastructure with a reasonable accuracy and least error. The tactics used for PLF heavily depend on the available data gained from the meters that are the foci of various problems. The mainstream methods that are mostly based on lack of preprocessing of power load data and fail in noisy condition when the data are noisy and have outliers or effects of user's behavior. This problem is handled via the refinement layer. Similarly, they provide a coarse way to collect the features for PLF, practicing low-level methods that widely skip the most discriminative features of the power data sequence. They are based on traditional feature extraction such as clustering, ensemble learning, or hand-feature engineering techniques which fail in terms of amassing the most advanced knowledge and deep characteristic from the data. In addition, the practiced approaches use a single RNN layer and passed via single hidden state that ignore the capturing of (hierarchical/think) temporal structure of the sequence. Furthermore, existing PLF techniques hands on complex architecture resulting into a large number of parameters and become computational. We handle these problems by proposing a deep learning-assisted short-term PLF method, which investigate the ConvLSTM layers giving its latter to GRU. Applying such procedure, the proposed method achieved the accurate and fast calculation results.

The key contributions of the proposed method are highlighted as follows:

(1) Existing PLF strategies practice traditional filters to overcome noisiness in data which remove the noisy disordering only. To tackle this problem, an acquisition and refinement layer is employed that refines the data through past value substitution, normalization, and organizing the data into a rolling window sequence.

(2) Employed works apply conventional learning and hand-feature engineering strategies making the load forecasting stiffer and tedious. However, for the first

time, we propose a novel ConvLSTM network for PLF that extracts the most discriminative spatiotemporal features from the power load sequence and generate a block of feature map.

(3) We employ a deep GRU for sequence learning that obtains the spatiotemporal features (feature map) from the ConvLSTM network. The deep GRU network is more suitable for learning the sequence patterns and provides effective PLF demonstrated through visual and tabular results.

(4) We experimentally prove that the proposed method shows outstanding results and outperforms state of the art through recording the least value of MSE and RMSE on the most challenging dataset. The demonstration of the proposed method's results depicts the method suitability for efficient management of energy infrastructure and ensuring saving of vast amount of energy wastage.

The rest of the article is organized as follows: Section 2 covers the literature review while the proposed PLF method and experimental analysis are discussed in Section 3 and Section 4, respectively. Finally, Section 5 describes the comparative analysis while Section 6 concludes the article.

2. Literature Review

Several techniques have been developed with the aim to efficiently forecast the energy consumption for buildings, industries, institutes, or residential areas. These methods are broadly based on conventional and deep learning-based techniques. The details are covered in the following sections.

2.1. Conventional Learning-Based PLF Methods. PLF methods remained popular for their promising results to forecast the power load consumption in residential buildings [13, 14], subways [15], industries [16], and households [17, 18]. Majority of the methods are based on traditional approaches. For instance, Guo et al. [19] composed a machine learning-based model to forecast the building thermal energy using extreme learning machine (ELM), multiple linear regression, and support vector machine. They analyzed the performance of each model for the heating system. Next, Peng et al. [20] used a framework that worked with multiprocessing learning based on certain defined rules to control cooling. They applied a method to adapt user scenarios with no prior knowledge. Similarly, Ngo [21] proposed an ensemble machine learning-based method to estimate the inside building cooling loads and analytically proven the ensemble learning as best performance. Hygh et al. [22] employed the Monte Carlo framework to develop multivariate linear regression model for 27 buildings that are relevant to early design, as the energy performance is sensitive to the size and building geometry. Another researcher, Wang and Ding [23], proposed an occupant-based PLF model for equipment by applying the polynomial and Markov Chain Monte method to investigate the time-varying occupancy rate. Considering time accumulation,

they calculated the consumption of equipment in the office. Furthermore, Zhong et al. [24] defined a vector field based on support vector regression for PLF in building. They transformed the model nonlinearity into linearity using these vectors. Another research in [25] performed daily and hourly analysis with the use of quadratic regression such as simple and multiple linear regression. They proved that the time interval is the relevant factor that defines the model quality. Furthermore, the researchers also proposed clustering-based energy consumption to categorize electricity usage into different levels. Hence, a majority of these methods failed and remained limited to obtain accurate forecasting and least error.

2.2. Deep Learning-Based PLF Methods. Deep learning is gaining overwhelming growth in solving different computer vision tasks such as video analytics [26] or time-series problem [27, 28]. It deeply inspires the field of energy consumption and getting involved due to model robustness and performance. For instance, Muralitharan et al. [29] proposed an optimization approach based on a neural network to analyze the energy demand through PLF. They used neural network-based genetic algorithm and particle swarm optimization methodologies. A research carried out in [30] proposed a hybrid forecasting model based on evolutionary deep learning which combined the genetic algorithm with LSTM and optimized it with the objective function. Inspired by LSTM performance, a method presented in [31] applied deep recurrent neural network (DRNN) with LSTM for PLF and photovoltaic power in a microgrid. They proved that the DRNN with LSTM performs better than multilayer perceptron and optimized the load dispatch using the particle swarm algorithm. Next, Rahman et al. [32] developed two DRNNs to forecast the electricity and applied them over medium to long horizon. They further used these models to compute the missing data scheme. Several researchers proposed hybrid approaches of combining convolutional neural network (CNN) with LSTM autoencoder to forecast future energy in residential building. Similarly, Shi et al. [33] used pooling-based DRNN that batches the group comprise customer's profiles where they addressed the problem of overfitting by increasing the data volume and its diversity. An approach presented in [34] combined the stacked autoencoders with ELM as a hybrid connection. They used ELM as a predictor and used autocorrelation analysis to determine the ELM variables.

3. Proposed Power Load Forecasting Method

The energy consumption from small buildings infrastructure to global level has greater consequences. Worldwide development and rise in technology increased energy consumption. Its management by the users is greatly impacted, bringing drastic variations in economies and different sectors. In this view, the industries and smart grids have energy deficits due to wastage of large amount of energy, improper infrastructure, inefficient supply system, and the consumption building are not synchronized to efficiently

manage it. So far, the researchers apply several techniques to manage and synchronize the energy usage through its future forecasting. However, their improper spatial and temporal structure has made it more difficult to build the most robust forecasting model. The existing state of the arts have presented several sets of procedure but failed due to misleading features tools, metering procedure, etc. Based on these assumptions, we propose a proficient deep learning-assisted intelligent PLF method that provides a useful way to overcome the energy dissipation. The proposed method reduces the error rates with a high margin and obtains the most promising results. The visuals of each steps performed in the proposed method are presented in Figure 2, while the details are covered in the following sections.

3.1. Power Data Acquisition.
This section delivers a detailed explanation of data gathering from its sources such as meter and installed sensors, and the data preprocessing is explained. To collect the data, wires across the building floors are articulated into a single edge with the main board and the meter with few sensors is installed to read and measure the energy consumed over the building setup where the data are normally collected with minute resolutions. Usually, the data collected through sensors and meter are greatly affected by the climate condition, occupant's behavior, redundancy, wire break or short circuit that brings abnormalities, outlier, and noise in the variable values. Tackling this issue is necessary for accurate forecasting; therefore, we refine the data prior to actual processing. For cleansing the data, we apply several smoothing filters such as LOESS or LOWESS that are used by numerous researchers [35, 36] for reasonable results. We remove the noise and considered the previous values on that position and remove data redundancy. In addition, we found the data attributes with different scale that are handled by applying the normalization.

3.2. Sequence Modeling via Long Short-Term Memory.
Long-term dependencies with distant characteristics are not sufficiently captured through RNN because of the vanishing gradient effect. Therefore, the gating mechanisms are introduced where the classical activation is replaced. To model the sequential data, a type of recurrent neural networks (RNNs) such as LSTM has been proven to be the most stable and powerful network which understands and deals the long-range information [37]. LSTM has the capability of learning the long-term sequence information. The most interesting fact about the LSTM is their memory cells C_t that significantly act as accumulator for the state information. These cells are accessed, controlled, and written through numerous self-parameterized gates. The cells are accumulated by keeping the input gates active I_t with the arrival of a new information as an input. The controlled information flow inside the cell allows the network to memorize long-term dependencies. Similarly, if the forget gate F_t is active, the status of past cell C_{t-1} will be forgotten. Next, the latest information that will be controlled by the final gate F_t is managed by an output gate O_t. The most significant and vital role of the memory cells and gates for the information flow is

such that the gradient will be trapped into the cell which is known to be constant error [38], is prevented from vanishing, and acts as a critical problem for vanilla RNN model [38]. The fully connected LSTM is considered to be LSTM multivariate version where the input, output, and forget gates are briefly given in Figure 3(a). The sigmoid activation determines what kind of information needs to be updated, as certain information might be ignored. The mechanism followed in an LSTM is defined by equations (1) to (6).

$$i_t = \sigma_g \left(W_i * x_t + W_i * h_{t-1} + b_i \right), \tag{1}$$

$$f_t = \sigma_g \left(W_f * x_t + W_f * h_{t-1} + b_f \right), \tag{2}$$

$$o_t = \sigma_g \left(W_o * x_t + W_o * h_{t-1} + b_o \right), \tag{3}$$

$$C_t = \sigma_c \left(W_c * x_t + W_c * h_{t-1} + b_c \right), \tag{4}$$

$$c_t = f_t * c_{t-1} + i_t * C_t, \tag{5}$$

$$h_t = o_t * \sigma_c \left(c_t \right), \tag{6}$$

where i_t, f_t, and o_t are the input, forget, and output gate, respectively, while c_t and h_t are the cell state and hidden state, respectively. Similarly, C_t is the vector value constructed for tanh at t represented by σ_c while σ_g is sigmoid function. $*$ is the elementwise multiplication. W_f, W_i, and W_{xo} are weight matrices representing forget, input, and output cell, respectively.

3.3. ConvLSTM Network.
Input data that have to be collected in a longer time horizon can be reduced and filtered based on the convolution operations incorporated in LSTM networks or LSTM cell directly. Such approaches intend in the improvement of prediction accuracy of long-term sequence based on additional input data processing through projecting the data into lower dimensional space. Approaches to incorporate the convolution operations in LSTM are present in [39]. In the previous cases, the network is capable to model locally distributed relations and extract the corresponding features. LSTM, on the other hand, is useful to learn the temporal dependencies, so that the composition of the networks in the stacked form shows the best prediction results. Using convolutional LSTM, features can capture long-term horizon which makes them able of incorporating a larger amount of past information in the prediction. Fully connected LSTM is powerful in handling the temporal correlation and face redundancy in the spatial data. Tackling such an issue needs an extension of an entirely connected LSTM that has a convolutional' structure with both input and state-to-state transition. To form an encoding forecasting mechanism, multiple ConvLSTM layers are stacked together which not only build the model for precipitation forecasting but also build the model for spatiotemporal sequencing forecasting procedure. In fully connected LSTM, the inputs are unfolded into 1D vectors before actual processing to handle the spatiotemporal data; as result of this, the important information is lost. This
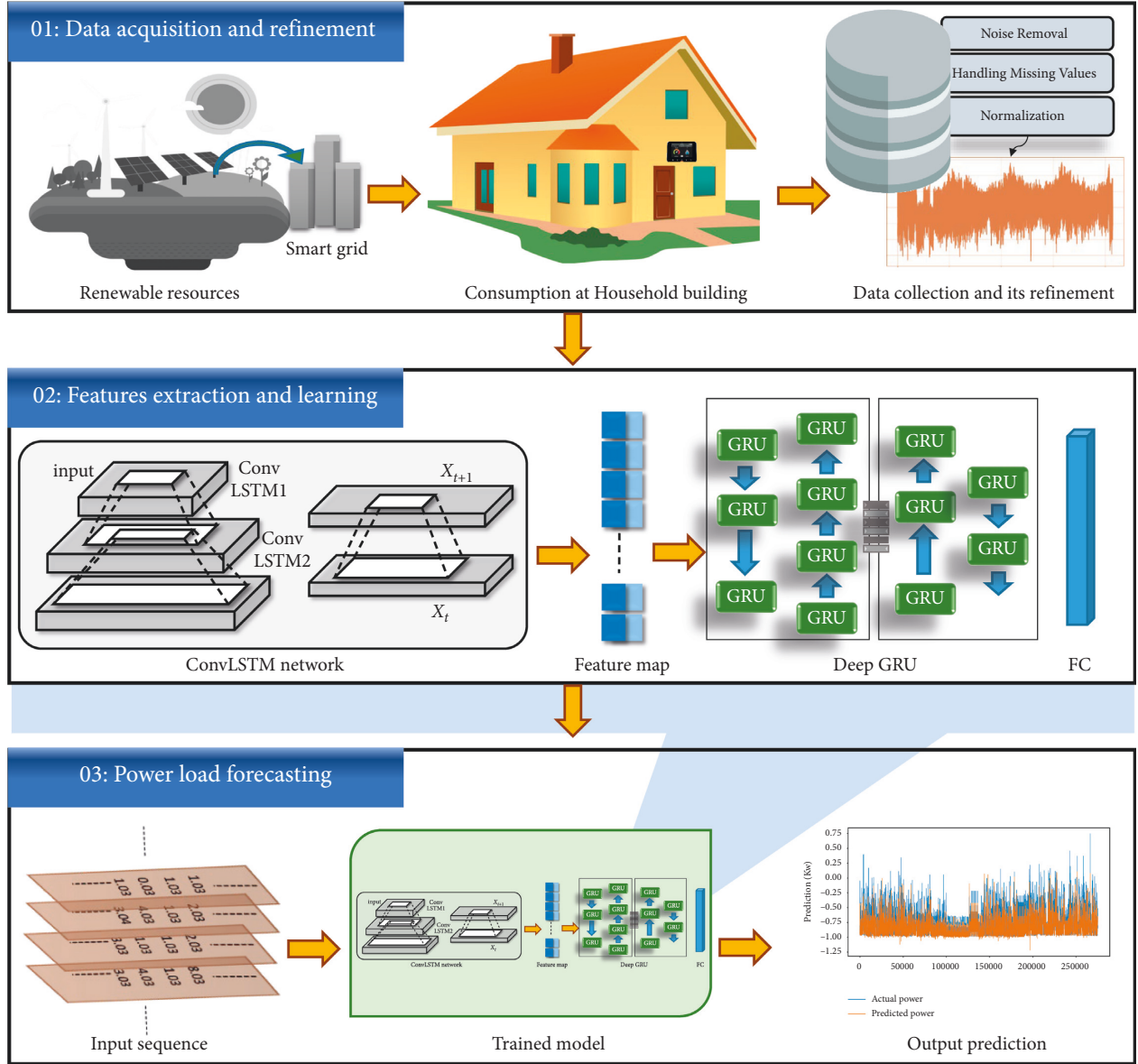
FIGURE 2: Overview of the proposed framework. Step 1: Data acquisition from sensors and the smart meters that contain the user behavior for data management and response to energy usage. Also, its refinement is performed to stabilize and remove the user behavior. Step 2: Detailed feature extraction from the energy data through ConvLSTM. Step 3: The prediction of future energy consumption based on the input data resolution.

problem can be overcome such that all the inputs into ConvLSTM are likely to be 3D tensors where the last dimensions are spatial row and column. ConvLSTM defines next state of some cells into a grid through inputs and previous states around its neighbor sides. This strategy can be achieved through convolution operation from the input-to-state and also state-to-state transitions. Furthermore, the deep mechanism along with the key equations of the process is given in equations (7) to (11).

$$i_t = \sigma\left(W_{xi} * x_t + W_{hi} * \mathcal{H}_{h-1} + W_{ci} o C_{t-1} + b_i\right), \quad (7)$$

$$f_t = \sigma\left(W_{xf} * x_t + W_{hf} * \mathcal{H}_{T-1} + W_{cf} o C_{t-1} + b_f\right), \quad (8)$$

$$C_t = F_t o C_{t-1} + i_t o \tanh\left(W_{xc} * X_t + W_{hc} * \mathcal{H}_{t-1} + b_C\right), \quad (9)$$

$$o_t = \sigma\left(W_{xo} * x_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} o C_t + b_o\right), \quad (10)$$

$$H_t = o_t o \tanh\left(C_t\right). \quad (11)$$

Similar to simple LSTM, the ConvLSTM can also be adopted which is the building block for the complex type of structure. The structure presented in Figure 3(c) solves the forecasting problem of our spatiotemporal sequence. The structure of this building block consists of an encoding and a forecasting network. To form such a
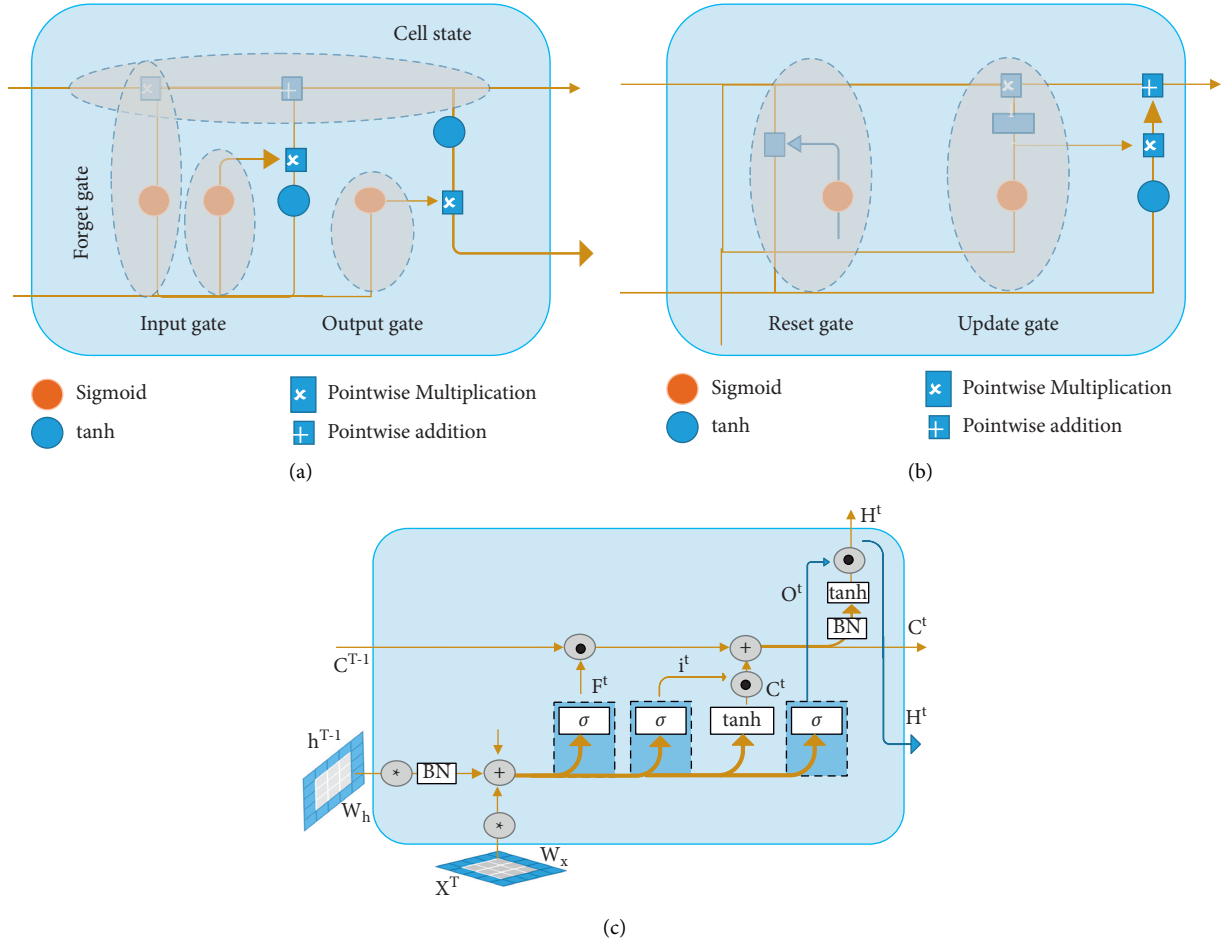
FIGURE 3: Internal structure of each sequential network. (a) LSTM has three main gates such as input, output, and forget. (b) GRU has two gates including reset and update. (c) Internal mechanism of ConvLSTM.

network, multiple ConvLSTM layers are stacked together, where the states in forecasting network are concatenated and fed into a 1×1 convolutional layer to generate final forecasting. This strategy is performed, as the input has the same dimensionality to target the prediction. This structure is considered with a similar viewpoint as in [37]. The encoding LSTM compresses the given sequence into state of hidden tensor where the forecasting LSTM unfolds the hidden state that gives the final prediction. The structure of this network is similar to LSTM as a future prediction model [40], but the input and output elements in our model are 3D tensors where the spatial information is preserved. As multiple ConvLSTM layers are stacked together, they give a strong representation and empower to give a fine prediction for a complex sequence such as PLF.

3.4. Deep GRU Network. GRU is an improved form of RNN which uses the time-series data sample for forecasting purposes. Traditional neural networks are characterized via the interconnection established between the input towards the hidden layer and then towards the input layer where a direct node in every layer is connected [41]. Consequently,

RNN memorizes the previously passed information that is applied to compute and find the current output. Several improvements are made in LSTM that solve the common errors and the shortcomings in the long-term sequential application of RNN. For instance, LSTM holds three gates such as input, output, and forget. Here, the forget gate is used to control the information and its rate of forgotten, while the output gate is applied to control the status of current unit condition that is strained out. The GRU overcomes the deficiencies present in RNN which is unable to handle the long-term dependencies in an effective manner; however, GRU on the other hand, makes the structure simpler and brings efficiency by preserving the effectiveness of LSTM [42]. The GRU network is well visualized in Figure 3(b). The GRU contains two gate functions, namely update and reset. The update gate function is to control the state information of the previous moment that is brought to current state (rate of updating state information). The greater the value of the update function, the more information will be brought in the previous moment. Similarly, the reset gate controls to avoid the information obtained from the previous moment (rate of forgetting information). The minute value of the reset gate indicates more information will be forgotten.

Subsequently, the unit state and output are combined into a single state H where the input is X_t, the state previously passed by the hidden layer is H_{t-1}, and similarly the information from the previous node is included via X_t and H_{t-1}. The GRU gets the output from the hidden state with the control gate while the information is passed into the next hidden layer. Two gates with states are obtained that are based on information Z_t and R_t, which are given in the following equations.

$$p_t = f\left(w_p [x_t, h_{t-1}] + b_p\right), \tag{12}$$

$$q_t = f\left(w_q [x_t, h_{t-1}] + b_q\right), \tag{13}$$

where W_p and W_q are weights of neurons, f shows the sigmoid function controlling the values between 1 and 0, which are used to obtain the gating signal. After getting the signal, first, the reset gate is used and then the data are obtained from the reset function that is combined with tanh activation function that give $\sim H_t$.

4. Experimental Results

This section discusses the experiments performed over the competitive energy dataset such as the household power dataset [43]. We comprehensively inspect the energy consumption and discuss the details of the dataset in the subsequent sections. Similarly, we visualize the results of energy consumption and its forecasting. The comparison of forecasting results on the existing dataset with state of the art is also covered which prove the effectiveness of the proposed method.

4.1. Implementation Settings. We verify and analyze the results of the proposed method using different kinds of experiments to check and evaluate the method's performance. The proposed method is implemented in Python (Version 3.5) with Keras (Deep Learning Framework) supported by the TensorFlow at the backend and ADAM-prop optimizer is used. Next, as we are dealing with regression problem, we apply four basic evaluation metrics such as MSE, RMSE, mean absolute error (MAE), and mean absolute percentage error (MAPE). These metrics are abundantly used for performance evaluation in regression problems throughout the energy forecasting-related literature. MSE is the basic error metric used in the PLF, renewable energy generation forecasting, weather prediction, humidity, etc. The formulation of each metric is described below:

Suppose $y_i^{\sim}$ indicates the variable values for n number of predictions for energy consumption and y_i indicates the observed values, so equations (14) to (16) show the MSE, RMSE, and MAE formulation where the RMSE is the square root of MSE values. Similarly, to measure the performance of forecasting that compute the correctness of the proposed method, MAPE is used, which gives the absolute error in percentage and compute the mean of the error and is given in equation 17.

$$MSE = \frac{1}{n} \sum_{i}^{n} \left(y_i - y_i^{\sim}\right)^2, \tag{14}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i}^{n} \left(y_i - y_i^{\sim}\right)^2}, \tag{15}$$

$$MAE = \frac{1}{n} \sum_{i}^{n} |y_i - y_i^{\sim}|, \tag{16}$$

$$MAPE = \frac{100\%}{n} \sum_{i}^{n} \left|\frac{y_i - y_i^{\sim}}{y_i}\right|. \tag{17}$$

4.2. Dataset. Standard and publicly available dataset is used to verify and evaluate the proposed method. The dataset is publicly available at [43] and its further details are covered in the following sections.

4.2.1. House Power Dataset. We evaluate and analyze the proposed ConvLSTM-GRU network using several kinds of experiments to gauge its performance on the household power dataset [43] that is available on the UCI official (deep learning) repository. This dataset is collected between 2006 and 2010—4 years of data. It contains 2075259 instances, where 25979 instances contain missing values making 1.25% of total data. Usually, the use of missing values creates problems in incorrect forecasting of energy consumption. Researchers use various techniques to overcome this problem. To tackle this problem, we pass the data from the refinement step that is previously explained in the proposed method. Next, this dataset is covered with 1 minute's time horizons of electric power consumption over the building located in France. In this dataset, the global active power indicates the total power consumed by submetering 1, 2, and 3 over single minutes provided in watt-hours. Testing the proposed method, we use different time steps that explain the PLF for each time horizon. The variables used in this dataset are given in Table 1 with their detailed remarks. Furthermore, we provide the quantitative details of the household power dataset in Table 2, where 11.12 is the maximum value for active power given in kilowatts and 0.076 is the minimum value. If we analyze the attribute values, the maximum energy is consumed over submetering 1 such as 88.000 that is dissipated over the daily usage devices such as microwave oven, dishwasher, etc.

4.3. Result Analysis and Discussion. This section describes the detailed experimental evaluation of the proposed method on the household power dataset [43].

We perform an ablation study where each model is implemented and trained on the given dataset to inspect the performance of the proposed method. These models include GRU and its variants such as encoder-decoder GRU (ED-GRU) and CNN-GRU. Each deep learning network is

TABLE 1: Household power dataset [43] with its attribute description.

#	Given attributes	Given symbol	Used units	Description
1	Date	D	Days/months/years	1~30/1~12/2006~2010
2	Timing	T	Hours/minutes/seconds	1~24/1~60/1~60.
3	Active power	AP	Kilowatts	Total power consumed in each minute
4	Reactive power	RP	---	Total reactive power consumed in each minute
5	Voltage	V	Volts	Total voltage in each minute
6	Intensity	I	Ampere	Current consumed in each minute.
7	Submetering-1	S(1)	Watts per hour	Energy consumed for dishwasher and microwave oven.
8	Submetering-2	S(2)	---	Energy consumed in living room over washing machines, drier, and refrigerator
9	Submetering-3	S(3)	---	Energy consumed for air conditioner and water heater

TABLE 2: Detailed description of the household power dataset [43] with its attributes and quantitative analysis.

Attribute	D	AP	RP	V	I	S1	S2	S3
Average	/	1.089	0.124	240.844	4.618	1.117	1.289	6.453
Min	16.12.2006	0.076	0	223.200	0.200	0	0	0
Max	26.11.2010	11.12	1.390	254.150	48.400	88.000	80.000	31.000
Std. Dev	/	0.055	0.113	3.239	4.435	6.139	5.794	8.436

compared with the proposed method for every time horizon such as minute, hour, day, and week. The deep learning networks are trained up to 100 epochs. Originally, the data in the household power dataset are given in minutes' resolution. For experiments, we convert the data into hour, day, and weekly horizons. After conversion, the number of instances become smaller which can be easily identified from the patterns given in Figure 4 showing the representation of the data.

4.3.1. Performance Evaluation of the GRU Network. Conducting the experiments, the deep GRU is initially evaluated to check its performance over the household power dataset [43]. Basically, the GRU network tries to solve the vanishing gradient problem which is originated with the standard RNN. This network is also considered to be a variation of LSTM because both have similarities and sometimes produce equally excellent results. A simple GRU network uses update and reset gates which are basically two vectors that control the information which need to be passed as an input. Its internal details are covered in the previous section while its internal structure is given in Figure 3(b).

For experiments, two GRU layers are stacked together followed by the dense layer. The total obtained parameters for this network are 382607. Furthermore, the values obtained for MSE, RMSE, MAE, and MAPE for GRU on minutes' horizons are 0.3569, 0.5974, 0.4012, and 0.4083, respectively. The other error values obtained for the household power dataset on each time horizon such as minute, hour, day, and weekly are provided in Table 3. The forecasting obtained for energy consumption over minute and hour time horizons using the GRU network is graphically presented in Figure 5, while the daily and weekly based results are given in Figure 6.

4.3.2. Performance Evaluation of the ED-GRU Network. Subsequently, we also use the ED-GRU network to identify its results and performance in the forecasting of energy consumption over the buildings. Setting the internal structure of ED-GRU, the encoder is set to stack several layers of GRU in such a way that each unit accepts a single input element sequence and collects the most important information from it and forward propagate it. Internally, the encoder vector is produced from the encoder which is known to be final hidden state. This state encapsulates the information of all the input elements to make an accurate load forecasting of power energy. It acts to be the hidden state of the model decoder part. Next, the decoder stacks several units which predict the output y at time t. Each of the decoder unit accepts the hidden state from the previous and produce its own hidden state. The hidden state h_i can be calculated using the formula given in (18) while the output y_t at time t is calculated through the formula given in (19). The layered GRU is followed by the time distributed layer. The number of obtained parameters for ED-GRU is 154,051. The values obtained on each time resolution for the household power dataset are given in Table 3. Next, the forecasting for the energy consumed considering minute and hour resolution is visually presented in Figure 7 while the daily and weekly based forecasting is given in Figure 8. The MSE obtained for ED-GRU on the household power dataset considering minute horizon is 0.3246 while RMSE, MAE, and MAPE are 0.5697, 0.3635, and 0.3485, respectively. Similarly, its performance is somehow improved on hour resolution with the least MSE which is 0.3134. Similar to its performance on the household power dataset, its performance becomes better when daily resolution is considered, where the value becomes 0.3054 while the RMSE, MAE, and MAPE are 0.5526, 0.3519, and 0.3401, respectively. Hence, this analysis clearly exhibits the good performance of ED-GRU after CNN-GRU and the proposed method.
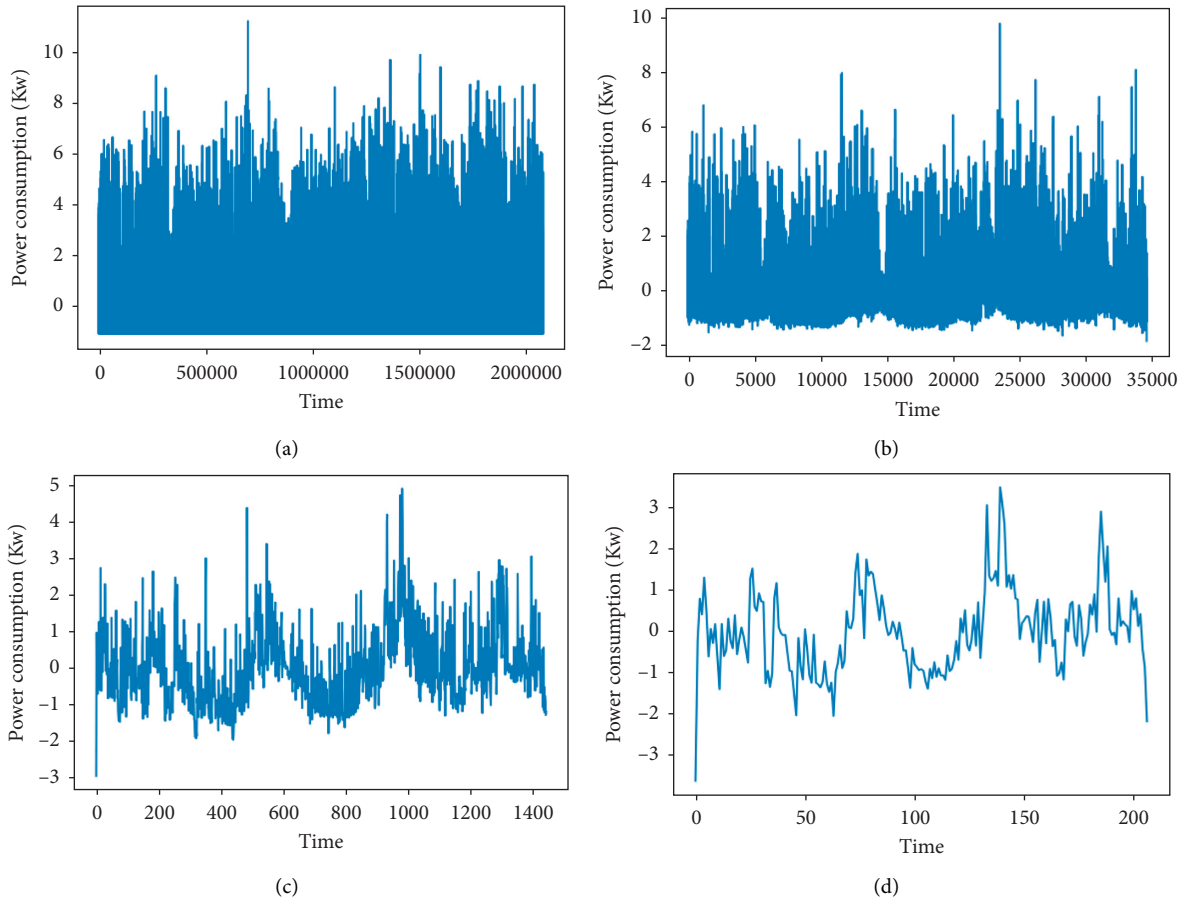
Figure 4: Data representation of the publicly available household power dataset [43] with different horizons. (a) Minutely data. (b) Hourly data. (c) Daily data. (d) Weekly data.

Table 3: Performance results of the proposed method and its comparison with other competitive deep learning models on the household power dataset for all time horizons such as minute, hour, daily, and week.

| Methods	Minutely			
	MSE	RMSE	MAE	MAPE
GRU	0.3569	0.5974	0.4012	0.4083
ED-GRU	0.3246	0.5697	0.3635	0.3485
CNN-GRU	0.3215	0.5670	0.3618	0.3064
Proposed method	**0.3101**	**0.5568**	**0.3467**	**0.2902**
	Hourly			
GRU	0.3251	0.5701	0.3814	0.3962
ED-GRU	0.3134	0.5598	0.3585	0.3415
CNN-GRU	0.2897	0.5382	0.3607	0.3042
Proposed method	**0.2384**	**0.4882**	**0.3435**	**0.2897**
	Daily			
GRU	0.3099	0.5566	0.3691	0.3586
ED-GRU	0.3054	0.5526	0.3519	0.3401
CNN-GRU	0.2821	0.5311	0.3524	0.3021
Proposed method	**0.2315**	**0.4811**	**0.3419**	**0.2879**
	Weekly			
GRU	0.3198	0.5655	0.3796	0.3697
ED-GRU	0.3157	0.5618	0.3627	0.3591
CNN-GRU	0.2981	0.5459	0.3619	0.3032
Proposed method	**0.2497**	**0.4996**	**0.3517**	**0.2882**
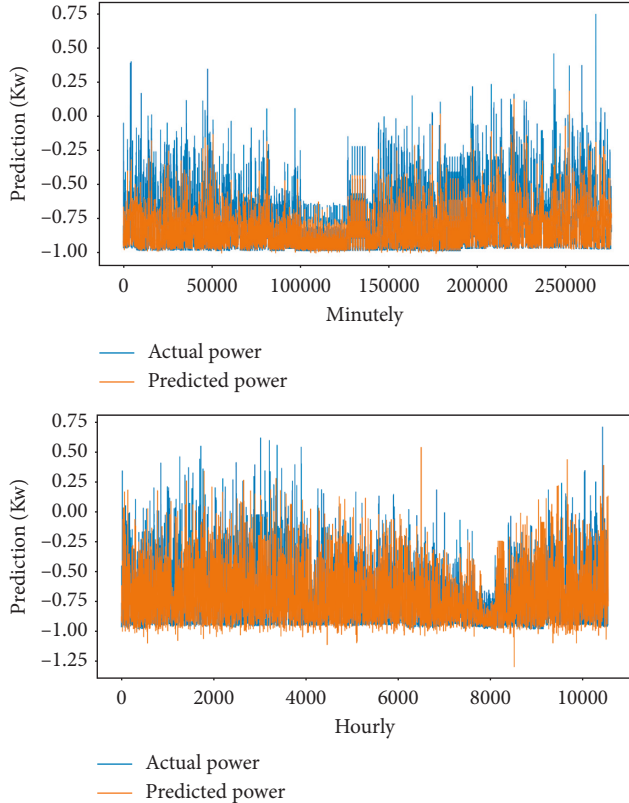
FIGURE 5: Visual representation of GRU-based forecasting results considering minute and hour horizons on the household power dataset.

FIGURE 6: Visual representation of GRU-based forecasting results considering daily and weekly horizons on the household power dataset.

$$hs_t = f\left(W^{(hh)}hs_{t-1}\right),$$
$$y_t = \text{software}\left(W^s hs_t\right). \quad (18)$$

4.3.3. Performance Evaluation of the CNN-GRU Network.

Recently, CNNs have shown the most promising results in different fields such as computer vision, time-series analysis, energy informatics, and energy monitoring system. To evaluate and analyze its performance, we incorporate its several layers for PLF problem. We combine it as hybrid connection with deep GRU and add three convolutional layers followed by the max pooling and flatten layer. The features obtained from these layers are given to deep GRU where two GRU layers are connected and followed by the time distributed layer. Furthermore, the error values obtained for each time horizon on the CNN-GRU network for the household power dataset are given in Table 3. Moreover, the forecasting of energy consumption on this network considering minute and hour resolution is graphically illustrated in Figure 9, while the daily and weekly based consumption forecasting is given in Figure 10. The CNN-GRU network has better performance than other deep learning models for consumption forecasting in terms of MAPE and MAE metrics.

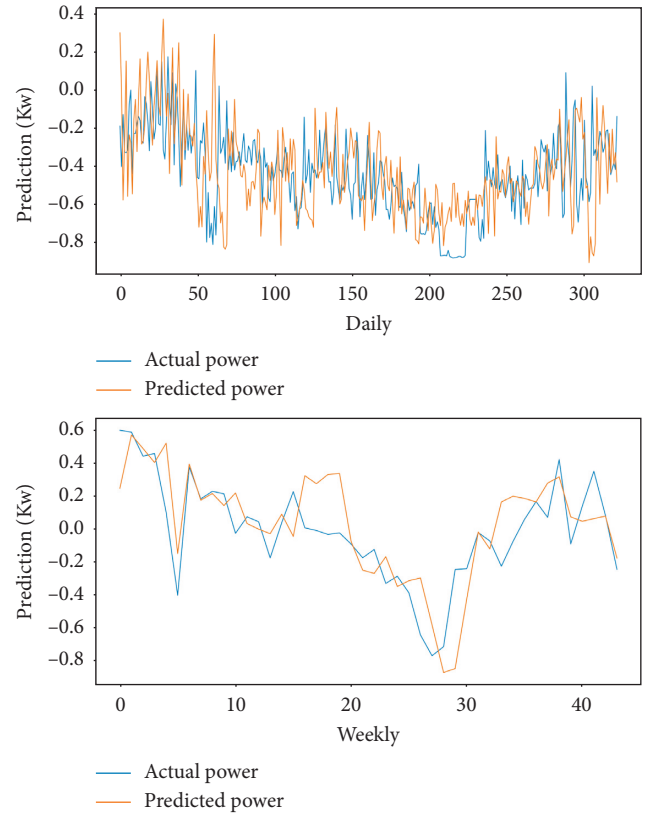Furthermore, CNN-GRU has better performance over all the methods and become a runner up. CNN-GRU shows the same response as ED-GRU in terms of its performance for resolution. For instance, CNN-GRU achieves an MSE of 0.3215 value considering the minute horizon; however, its results become better in hour resolution where MSE obtained is 0.2897. The value of RMSE, MAE, and MAPE obtained for CNN-GRU considering minutes resolution is 0.5670, 0.3618, and 0.3964, respectively. After deep analysis, we realize that CNN-GRU reflects its better results on the household power dataset rather than in terms of considering the minute resolution. In the final phase, we pose the results obtained for the proposed ConvLSTM with the GRU network. The values obtained by the proposed method on the household power dataset for metrics of MSE, RMSE, MAE, and MAPE for the minute horizon are 0.3101, 0.5568, 0.3467, and 0.2902, respectively, and are given in Table 3. Next, the visual representation of forecasting results considering minute and hour resolution is given in Figure 11, while the daily and weekly based resolution results are depicted in Figure 12.

5. Comparison with State of the Art

In this section, we analyze and compare the results of the proposed method with existing competitive state of the art in terms of performance considering the basic metric using the household power dataset. For a fair comparison, we consider and evaluate the minute horizon of the proposed method as
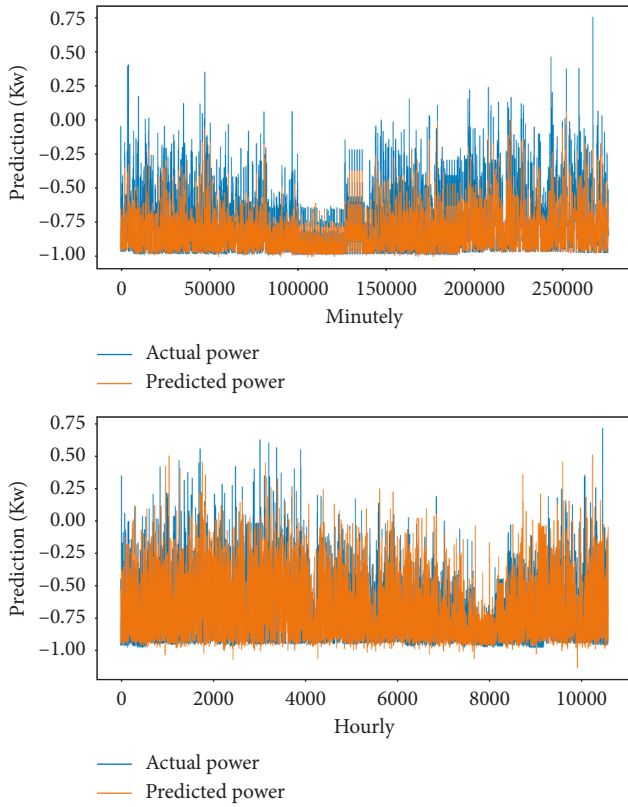
FIGURE 7: Visual representation of ED-GRU-based forecasting results considering minute and hour horizons on the household power dataset.
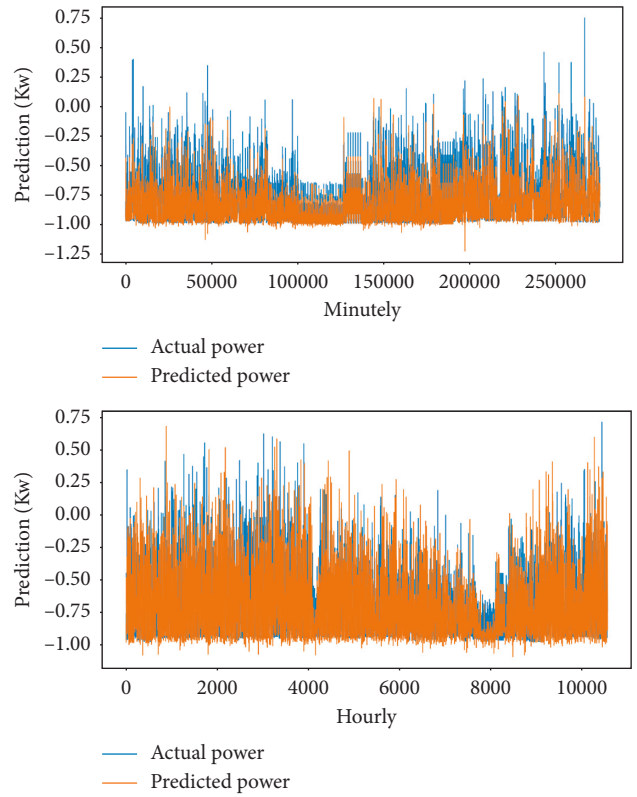

FIGURE 9: Visual representation of CNN-GRU-based forecasting results considering minute and hour horizons on the household power dataset.
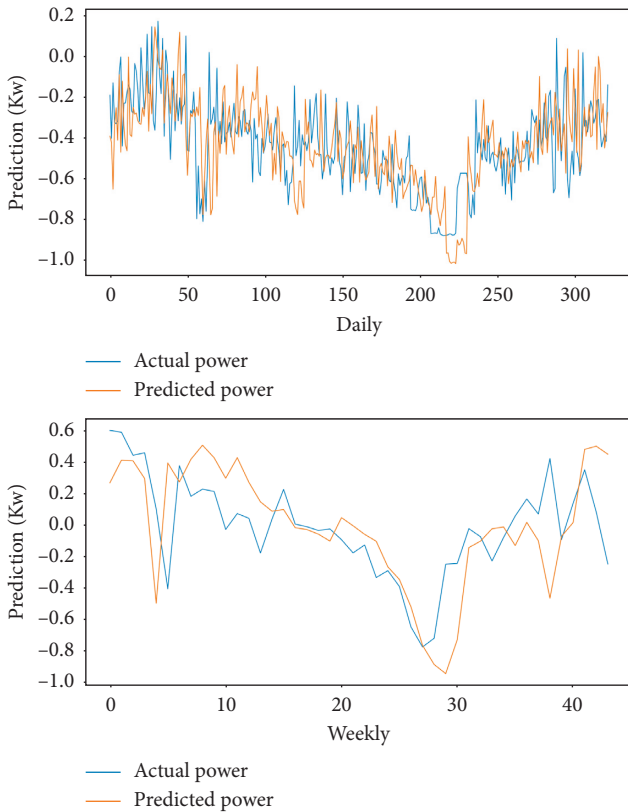

FIGURE 8: Visual representation of ED-GRU-based forecasting results considering daily and weekly horizons on the household power dataset.
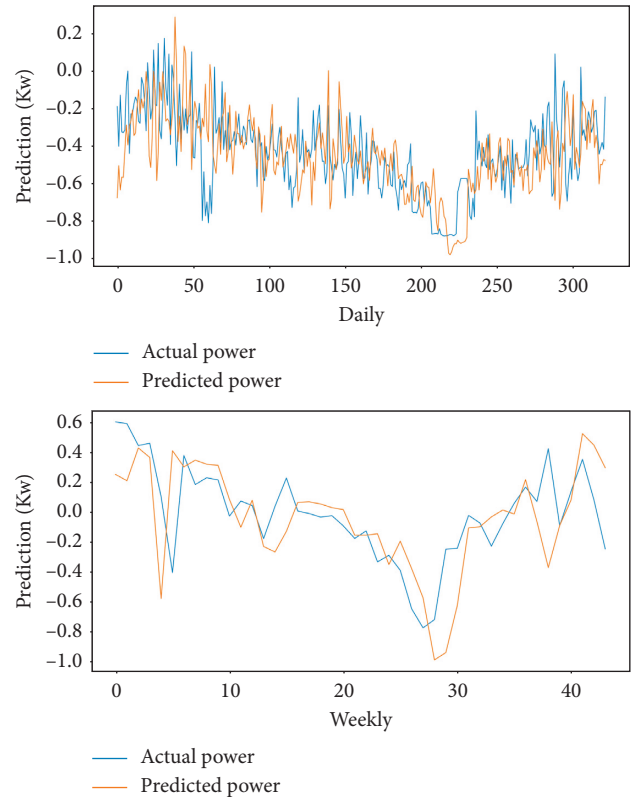

FIGURE 10: Visual representation of CNN-GRU-based forecasting results considering daily and weekly horizons on the household power dataset.
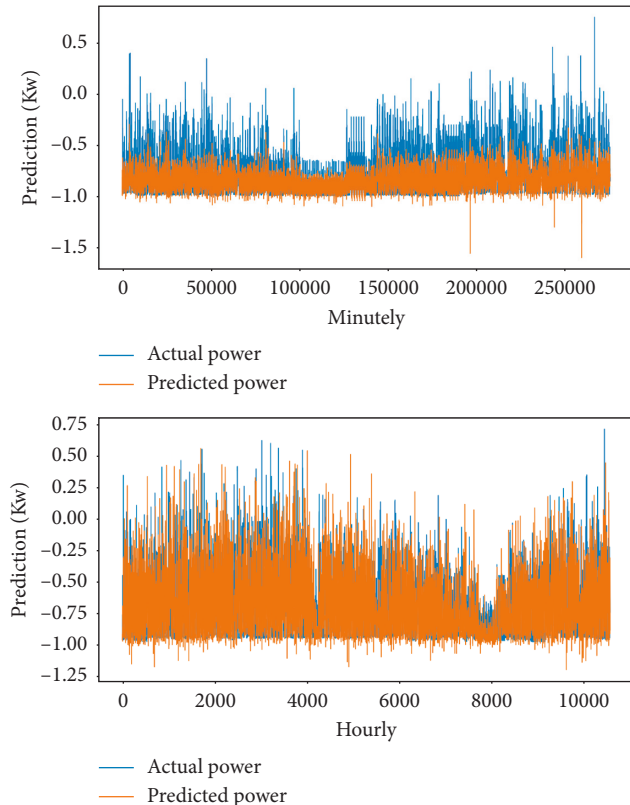
Figure 11: Forecasting results obtained using the proposed method for the household power dataset considering minute and hour time horizons.

Figure 12: Forecasting results obtained through the proposed method for the household power dataset over day and week time horizons.

considered by other works. Similarly, we consider the same metrics for comparison. A method presented in [47] proposed a three-stage hybrid network of CNN with a multi-layer BLSTM network to forecast the power load. They first practiced LSTM and then BLSTM to assess their performance by obtaining 0.3446 and 0.3295 MSE values, respectively. However, their proposed method's outcomes obtained for MSE, RMSE, MAE, and MAPE on this method are 0.3193, 0.5650, 0.3469, and 0.2910, respectively. Next, Mocanu et al. [44] investigated two main models to estimate the energy consumption in buildings such as the conditional restricted Boltzmann machine and the factored conditional restricted Boltzmann machine. They further considered the support vector machine and RNN to investigate their method. Their method used a single layer of factored conditional restricted Boltzmann machine to fit the needs for representing different useful parameters. They used RMSE as an evaluation metric and obtained 0.6663 value for it. They also computed correlation coefficient (R) and p-value achieving 0.4552 and 0.0070, respectively. Furthermore, Kim and Cho [45] proposed a deep learning-based method to forecast the energy demand. For this purpose, they used a state explainable autoencoder-based model and obtained 0.3840 value for the RMSE metric. A research presented in [46] used the hybrid approach of CNN with LSTM and reported 0.3738, 0.6114, 0.3493, and 0.3484 values for MSE, RMSE, MAE, and MAPE, respectively. The comparative results are summarized in Table 4.
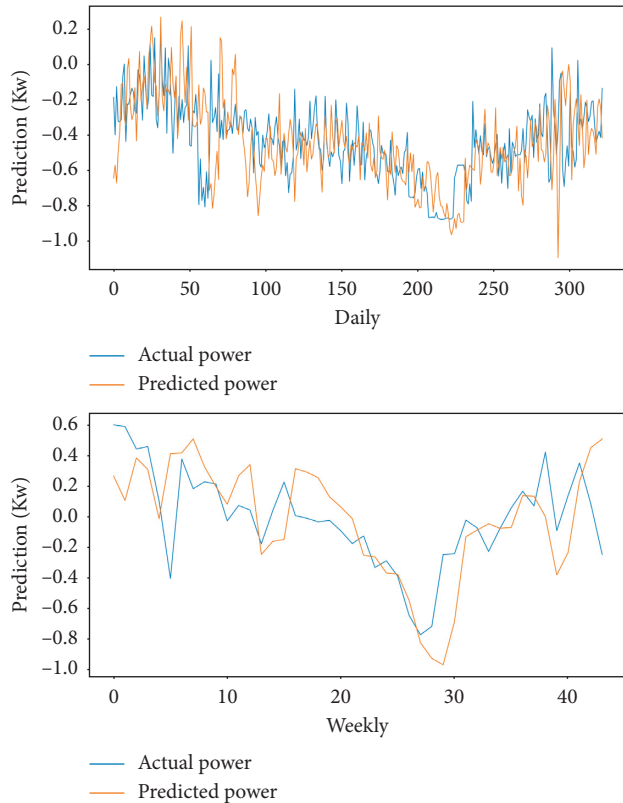
Table 4: Detailed comparative analysis with competitive state of the arts with the proposed method on the household power dataset.

Techniques	Evaluation metrics			
	MSE	RMSE	MAE	MAPE
FCBRM [44]	—	0.6663	—	—
BPTT [45]	0.3840	0.6196	0.3953	—
CNN-LSTM [46]	0.3738	0.6114	0.3493	0.3484
CNN-M-BLSTM [47]	0.3193	0.5650	0.3469	0.2910
Proposed method	0.3101	0.5568	0.3467	0.2902

Primarily, the aforementioned methods apply the traditional way of collecting the feature information from the sequence, which are old machine learning practices yielding lower correctness in prediction problem. Similarly, in the case of deep learning usage by these methods, their networks apply complex architectures where the training consumes more time. Next, if overview, the PLF methods are heading towards convolutional networks and sequential learning mechanisms such as RNN, LSTM, or BLSTM, which are the recent state-of-the-art learning methods [48]. These methods highly rely on the input model parameters and mostly hunt towards error reduction for the precise prediction. After thorough exploration, these methods have high error rates considering the minute horizon and have complex architecture. We improved our method by reducing the
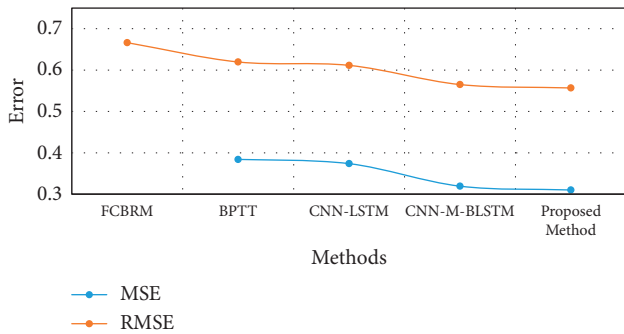
FIGURE 13: Comparative analysis of the proposed method with state of the arts. The blank area for FCBRM shows that there is no MSE value provided by them.

error to 0.3101 while the runner up method is our previously published work with an error rate of 0.3193. The comparison of our method with other PLF methods is visually presented in Figure 13.

6. Conclusions

Over the decades, the energy demand is growing throughout the world due to an increase in technology, industrial machinery, and population. This results in the wastage of a large amount of energy due to a lack of efficient usage and its storage from the grid or renewable energy resources. Therefore, energy generation companies and smart grid authorities are investigating new ways to tackle this issue. To this end, we proposed an intelligent deep learning-based architecture for energy to boost the PLF for the proper establishment of energy infrastructure. To carry out, initially, the data collected through various installed sensors and meters are fed into the acquisition layer for refinement purposes. Next, the refined data are passed into the ConvLSTM network to extract the deep features and generate the final feature map. Further, the feature map is passed into deep GRU to learn the series which gives us final forecasting of the energy. In addition, we proved that the proposed method outperforms the existing state of the art using different error metrics that are applied for regression model evaluation. The proposed method is verified and tested on a publicly available household power dataset.

In the future, we intend to enhance the method by the involvement of the Internet of things (IoT) [49], that is the deployment over resource-constrained devices, which will reduce the complexity in terms of computational power and resources as performed in [50]. This will help to reduce the bandwidth and easethe transmission of information. Furthermore, we aim to include the forecasting for years and decade-wise consumption and generation of power energy by considering various characteristics. The characteristic involves weather condition, energy consumption over industries, public transport, and occupants' behavior in response to these disruptions. We will incorporate these kinds of datasets to further confirm and verify their impact on load forecasting.

Data Availability

The data that are used to support the findings of this study are included at https://archive.ics.uci.edu/ml/datasets/ individual+household+electric+power+consumption.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] P. Zhang, X. Ma, and K. She, "A novel power-driven grey model with whale optimization algorithm and its application in forecasting the residential energy consumption in China," *Complexity*, vol. 2019, pp. 1–22, Article ID 1510257, 2019.

[2] H. Luo, H. Zhang, and J. Wang, "Ensemble power load forecasting based on competitive-inhibition selection strategy and deep learning," *Sustainable Energy Technologies and Assessments*, vol. 51, Article ID 101940, 2022.

[3] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, "Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction," *Complexity*, vol. 2019, pp. 1–13, Article ID 7414318, 2019.

[4] X. Guo, Q. Zhao, S. Wang, D. Shan, and W. Gong, "A short-term load forecasting model of LSTM neural network considering demand response," *Complexity*, vol. 2021, Article ID 5571539, 2027 pages, 2021.

[5] A. Yang, W. Li, and X. Yang, "Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines," *Knowledge-Based Systems*, vol. 163, pp. 159–173, 2019.

[6] S. Woo-hyun, "Moon's nuclear-free policy lauded but face technical hurdles," 2017, http://www.koreaherald.com/view. php?ud=20170623000670.

[7] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 796–808, 2017.

[8] Y. Chen, Z. Dong, Y. Wang et al., "Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history," *Energy Conversion and Management*, vol. 227, Article ID 113559, 2021.

[9] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes," *Energy and Buildings*, vol. 130, pp. 519–531, 2016.

[10] A. Tascikaraoglu and B. M. Sanandaji, "Short-term residential electric load forecasting: a compressive spatio-temporal approach," *Energy and Buildings*, vol. 111, pp. 380–392, 2016.

[11] M. H. Alobaidi, F. Chebana, and M. A. Meguid, "Robust ensemble learning framework for day-ahead forecasting of household based energy consumption," *Applied Energy*, vol. 212, pp. 997–1012, 2018.

[12] P. Lusis, K. R. Khalilpour, L. Andrew, and A. Liebman, "Short-term residential load forecasting: impact of calendar effects

and forecast granularity," *Applied Energy*, vol. 205, pp. 654–669, 2017.

[13] F. U. M. Ullah, N. Khan, T. Hussain, M. Y. Lee, and S. W. Baik, "Diving deep into short-term electricity load forecasting: comparative analysis and a novel framework," *Mathematics*, vol. 9, no. 6, p. 611, 2021.

[14] N. Khan, I. U. Haq, F. U. M. Ullah, S. U. Khan, and M. Y. Lee, "CL-net: ConvLSTM-based hybrid architecture for batteries' state of health and power consumption forecasting," *Mathematics*, vol. 9, no. 24, p. 3326, 2021.

[15] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Systems*, vol. 82, pp. 29–40, 2015.

[16] A. Abdoos, M. Hemmati, and A. A. Abdoos, "Short term load forecasting using a hybrid intelligent method," *Knowledge-Based Systems*, vol. 76, pp. 139–147, 2015.

[17] R. Talavera-Llames, R. Pérez-Chacón, A. Troncoso, and F. Martínez-Álvarez, "Big data time series forecasting based on nearest neighbours distributed computing with Spark," *Knowledge-Based Systems*, vol. 161, pp. 12–25, 2018.

[18] L. Harrington, L. Aye, and B. Fuller, "Impact of room temperature on energy consumption of household refrigerators: lessons from analysis of field and laboratory data," *Applied Energy*, vol. 211, pp. 346–357, 2018.

[19] Y. Guo, J. Wang, H. Chen et al., "Machine learning-based thermal response time ahead energy demand prediction for building heating systems," *Applied Energy*, vol. 221, pp. 16–27, 2018.

[20] Y. Peng, A. Rysanek, Z. Nagy, and A. Schlüter, "Using machine learning techniques for occupancy-prediction-based cooling control in office buildings," *Applied Energy*, vol. 211, pp. 1343–1358, 2018.

[21] N.-T. Ngo, "Early predicting cooling loads for energy-efficient design in office buildings by machine learning," *Energy and Buildings*, vol. 182, pp. 264–273, 2019.

[22] J. S. Hygh, J. F. DeCarolis, D. B. Hill, and S. Ranji Ranjithan, "Multivariate regression as an energy assessment tool in early building design," *Building and Environment*, vol. 57, pp. 165–175, 2012.

[23] Z. Wang and Y. Ding, "An occupant-based energy consumption prediction model for office equipment," *Energy and Buildings*, vol. 109, pp. 12–22, 2015.

[24] H. Zhong, J. Wang, H. Jia, Y. Mu, and S. Lv, "Vector field-based support vector regression for building energy consumption prediction," *Applied Energy*, vol. 242, pp. 403–414, 2019.

[25] N. Fumo and M. Rafe Biswas, "Regression analysis for prediction of residential energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 47, pp. 332–343, 2015.

[26] F. U. M. Ullah, M. S. Obaidat, K. Muhammad et al., "An intelligent system for complex violence pattern analysis and detection," *International Journal of Intelligent Systems*, vol. 12, 2021.

[27] C. Fan, Y. Sun, Y. Zhao, M. Song, and J. Wang, "Deep learning-based feature engineering methods for improved building energy prediction," *Applied Energy*, vol. 240, pp. 35–45, 2019.

[28] N. Khan, F. U. M. Ullah, A. Ullah, A. Ullah, M. Y. Lee, and S. W. Baik, "Batteries state of health estimation via efficient neural networks with multiple channel charging profiles," *IEEE Access*, vol. 9, pp. 7797–7813, 2021.

[29] K. Muralitharan, R. Sakthivel, and R. Vishnuvarthan, "Neural network based optimization approach for energy demand prediction in smart grid," *Neurocomputing*, vol. 273, pp. 199–208, 2018.

[30] A. Almalaq and J. J. Zhang, "Evolutionary deep learning-based energy consumption prediction for buildings," *IEEE Access*, vol. 7, pp. 1520–1531, 2019.

[31] L. Wen, K. Zhou, S. Yang, and X. Lu, "Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting," *Energy*, vol. 171, pp. 1053–1065, 2019.

[32] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Applied Energy*, vol. 212, pp. 372–385, 2018.

[33] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.

[34] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, "Building energy consumption prediction: an extreme deep learning approach," *Energies*, vol. 10, p. 1525, 2017.

[35] I. Ullah, R. Ahmad, and D. Kim, "A prediction mechanism of energy consumption in residential buildings using hidden Markov model," *Energies*, vol. 11, no. 2, p. 358, 2018.

[36] M. Fayaz and D. Kim, "A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings," *Electronics*, vol. 7, no. 10, p. 222, 2018.

[37] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, vol. 1, pp. 3104–3112, 2014.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[39] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM2. 5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, 2018.

[40] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using LSTMs," *International conference on machine learning*, pp. 843–852, 2015.

[41] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[42] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[43] A. B. Georges Hebrail, "Individual household electric power consumption Data Set," 2012, https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption.

[44] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, 2016.

[45] J.-Y. Kim and S.-B. Cho, "Electric energy consumption prediction by deep learning with state explainable autoencoder," *Energies*, vol. 12, no. 4, p. 739, 2019.

[46] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, 2019.

[47] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, "Short-term prediction of residential power energy consumption via CNN and multilayer Bi-directional LSTM networks," *IEEE Access*, vol. 8, Article ID 123380, 2020.

[48] T. Hussain, F. U. Min Ullah, K. Muhammad et al., "Smart and intelligent energy monitoring systems: a comprehensive literature survey and future research guidelines," *International Journal of Energy Research*, vol. 45, no. 3, pp. 3590–3614, 2021.

[49] F. U. M. Ullah, K. Muhammad, I. U. Haq et al., "AI-assisted edge vision for violence detection in IoT-based industrial surveillance networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5359–5370, 2022.

[50] M. Babar, M. U. Tariq, and M. A. Jan, "Secure and resilient demand side management engine using machine learning for IoT-enabled smart grid," *Sustainable Cities and Society*, vol. 62, Article ID 102370, 2020.