Short-Term Three-Phase Load Prediction With Advanced Metering Infrastructure Data in Smart Solar Microgrid Based Convolution Neural Network Bidirectional Gated Recurrent Unit

PHUONG NGUYEN THANH1,2, MING-YUAN CHO1, CHUN-LUNG CHANG1, AND MENG-JIE CHEN1

1Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 824 45, Taiwan
2Department of Electronic and Electrical Engineering, Nha Trang University, Nha Trang, Khanh Hoa 650000, Vietnam
Corresponding author: Phuong Nguyen Thanh (thanhphuong@ntu.edu.vn)

ABSTRACT
The collaboration development between the power supply, solar power, and storage batteries is a formidable commission in the smart solar microgrid. The efficient power management integrated into a smart solar microgrid is required to address these problems. The deep learning predictive tools are utilized to forecast the maximum building’s three-phase load power in the short term. In this project, historical data from the solar module, battery module, grid data, and climate data are collected by integrated advanced metering infrastructure in a smart solar microgrid in Taiwan. These historical data are utilized in predicting the building’s three-phase load power in the smart solar microgrid. This project proposed the hybrid model in the short-term three-phase loads forecasting based on convolution neural network bidirectional gated recurrent unit (CNN-Bi-GRU). The CNN-Bi-GRU utilizes the continuous-time sliding window, which is extracted features and reshaped into vectors by CNN layers. The hyperparameter optimization is utilized to construct the highest performance structure of the CNN-Bi-GRU model. The CNN-Bi-GRU performance is compared and evaluated with other state-of-arts models, which are the recurrent neural network (RNN), LSTM, gated recurrent unit (GRU), and bidirectional LSTM (Bi-LSTM), and Bi-GRU models. The experiment results prove that the forecasting accuracy could be effectively improved by a hybrid CNN-Bi-GRU model with an appropriate window size of sequential historical data. According to research knowledge, this is the first work to predict the building’s three-phase load power using multiple source data from advanced metering infrastructures (AMI) and deep learning CNN-Bi-GRU in the smart solar microgrid. The CNN-Bi-GRU model is successfully integrated into the efficient power management, which maintains the adequate battery storage level, solar power, and grid power balance under different weather conditions and working requirements of the building’s three-phase load power.

INDEX TERMS
Building’s three-phase load power prediction, convolution neural network bidirectional gated recurrent unit, deep learning machine, gated recurrent unit, long short-term memory, short-term load power, recurrent neural network.

I. INTRODUCTION
Because of the complexity of extensive electrical energy sources and continuous changes in power demand, the requirement in system power generation of smart solar microgrid should be dynamically stabilized with load changes.

The sensitivity between varying multiple renewable sources and power-consuming conditions plagues the grid connection in the smart solar microgrid. The accurate load forecasting improvement is conducive to increasing the utilization of multiple power source generation and the effectiveness of economic dispatch in smart solar microgrid’s efficient power management [1], [2]. The smart solar microgrid is a miniature power system that includes battery storage, grid power, and...
solar power. Therefore, effectively maintaining the power flow between multiple power sources is a challenging problem. A three-phase fault could easily trigger the cascaded protection equipment. Thus, predicting a building’s three-phase load is essential for the generation capacity of multiple sources between the power supply, solar power, and storage batteries. In addition, the short-term three-phase load power prediction is essential for arranging power plans, optimized generation capacity, and reasonable economic power transactions in buildings [3], [4]. The modern microgrid requires responsive power dispatching techniques to deal with dynamic grid load problems and irregular energy supplying sources [5].

Machine learning is the most preferred methodology for predicting the short-term load. Recently, deep learning machines have become the most effective techniques which rely on stacking multiple layers and refer to stochastic optimization during the learning process. Many studies are dedicated to the deep machine learning algorithm to validate and compare the forecasted performance of short-term load [6], [7]. Kim et al. proposed the hybrid 1-dimensional CNN and recurrent inception neural network for building an accurate load prediction in the energy management systems [8]. Liang et al. proposed the hybrid general regression neural network and empirical mode decomposition in short-term load prediction [9]. Sadaei et al. developed a combined algorithm based on fuzzy time and convolutional neural networks to predict hourly load power in Malaysia in 2009 and 2010 [10]. However, the LSTM has been proved the effective performance in sequential learning [11]–[16], which are the speech recognition [17], [18], the image processing implementations [19], [20], and natural language applications [21], [22]. Kong et al. utilized the LSTM to forecast individual residential household loads outperforming other rival algorithms [23]. The LSTM proved the long-term learning ability which could deal with extremely challenging problems of residential load prediction [23]. However, the hyperparameter optimization did not develop to further enhance the predicting performance on different structures of the LSTM model and diverse customers [23]. The load forecasting performances of the LSTM model had better benchmarks than traditional ARIMA, which only utilized the load data without meteorological factors [24]. These weather parameters could contain practical knowledge and provide relevant information which is not considered in the predicting model [24]. Jiao et al. proposed the LSTM to predict the non-residential load with multiple correlated factors [25]. The experiment results proved the effective utilization of the LSTM model with sequential information of the incremental sequence of time, week-day, and corresponding binary holiday mark [25]. However, the above sequential information could not provide effective information in predicting the non-residential load in massive data. Daniel et al. presented standard LSTM and sequence-to-sequence LSTM (S2S LSTM) in load prediction for more efficient energy management and minimized power wastage [26]. The experiment proved the S2S LSTM architecture performs better in both one-hour and one-minute time-step resolution [26]. In general, the predicting error of LSTM is better than other traditional predicting models in processing sequential input data. However, the LSTM model could not leverage the prior knowledge in massive historical data to extract relevant features which could effectively enhance the predicting performance and accuracy.

Moreover, the gated recurrent unit (GRU), which retained the structure and systematically minimized the update and reset gates, is a variant of the LSTM model [27]. The GRU has fewer parameters and better convergence velocity than LSTM and could achieve similar accuracy in short-term load forecasting [28]–[31]. Zheng et al. predicted the residential community load with the GRU model, which achieved faster convergence with similar accuracy compared with LSTM and RNN models [28]. The authors analyzed the influence of weather parameters on the residential load, and the average temperature was utilized in the forecasting model. Jia et al. proposed the Adam optimization with the GRU model in short-term load prediction, which achieved better accuracy and reduced computation time than LSTM and RNN models [32]. The GRU was utilized in predicting short-term load with the impact of electricity price and could improve the accuracy and achieve better performance than the LSTM model [33]. Zheng and Chen proposed the adaptive GRU and mixed gradient optimization for power prediction, which significantly improved compared with state-of-the-art methodologies [34]. Therefore, the GRU could reduce the computation time and achieve higher accuracy than the LSTM model in short-term load prediction.

In addition, many researchers utilized the CNN-LSTM [35]–[40] for feature extraction in the sequential input data, which improved the accurate performance. In addition, the CNN-GRU was also proved to provide an accurate prediction in sequential time-series data [41]–[45]. Alhussein et al. proposed the CNN-LSTM to predict the electrical load of the individual household, which achieved 40.38% mean absolute percentage error (MAPE) compared with 44.06% MAPE of the LSTM model [35]. The hour, holiday, and week-day indicators combined with historical electricity consumption data were employed in the CNN-LSTM model. The short-term load prediction is based on the clusters of various customers to identify the outlying number, which did not evaluate the meteorological data. Sajjad et al. developed a hybrid sequential CNN and GRU into the framework of energy consumption prediction, which revealed a better performance than existing energy forecasting models [41]. The proposed hybrid CNN-GRU attained the lowest values of MSE, RMSE, and MAE compared to the linear regression, SVM, CNN-LSTM, and Bi-LSTM models. Kim et al. proposed the CNN-LSTM for electric energy consumption prediction, which achieved the smallest value of RMSE compared with other methodologies [36]. The authors utilized the time indicating factors in the predicting model, which obtained a stable performance of 0.37 MSE compared with
linear regression, random forest regression, decision tree, and multilayer perceptron models. Moreover, the proposed hybrid CNN-GRU model acquired the highest performance of MAPE and RMSE compared with BPNN, GRU, and CNN methodologies [42]. Farsi et al. discussed a new hybrid parallel CNN and LSTM to predict the electrical load in German and Malaysian, which improved the accuracy to 91.18% and 98.23% in both datasets [37]. The authors only utilized historical sequential load data, and other dependent factors were not considered in the proposed methodology. Shen et al. proposed the CNN-GRU neural network for short-term busload prediction, which achieved higher precision and better accuracy in the smart grid [43]. The authors applied the date type, weather parameters, electricity price, and historical bus load data as relevant input factors in the predicting model, which reduced the average prediction error and improved the prediction accuracy. Rafi et al. propose CNN-LSTM for short-term load forecasting, which obtained effectiveness compared with LSTM, radial basis function network, and extreme gradient boosting algorithm [38]. Although the proposed CNN-LSTM/CNN-GRU acquired the advantages of both modules, various other factors that influenced load data were not examined in the predicting model. In general, these previous studies only utilized the historical electrical consumption data and time indicators in predicting CNN-LSTM/CNN-GRU models, which did not consider meteorological data and other factors.

The day-ahead electric load is predicted with the impact of meteorological factors and the LSTM model, which achieved better-forecasted performance in MAE, MAPE, and RMSE for Toronto city data [46]. However, the LSTM could not perform the feature extraction in massive input data compared with the CNN-GRU model, which achieved better accuracy and effectiveness. In previous research, both GRU and CNN can satisfactorily predict the short-term load power. The integration of CNN and GRU further decreases the predicting error and is suitable for the abrupt change in power demand. The CNN could improve the extraction of substantial and complicated patterns from the sequential load data, which produce the precise outcome and probabilistic prediction. The application of CNN and GRU could enable robust simplification in the energy prediction in the smart solar microgrid. However, the hyperparameter optimization did not utilize to maximize the accuracy and enhance the performance of the CNN-GRU model in previous studies. Currently, many factories in the industrial area, which utilize the three-phase load power for their manufacturing, have investigated the solar power plant on the roof-top to increase the energy-efficient. However, the prediction of three-phase load power has not been seriously investigated in recent years. Therefore, further investigations are necessitated to improve the accuracy and stability of the building’s short-term three-phase load prediction and mitigate the current limitation of existing methodologies. Observe that the authors in [43] only utilize the CNN-GRU in busload forecasting with the sequential historical weather data, electricity price, and energy generation. In contrast, this project proposes a novel hybrid methodology, which nominated the convolutional neural network bidirectional gated recurrent unit (CNN-Bi-GRU) with historical data from advanced meter infrastructure (AMI) in the smart solar microgrid for predicting the three-phase load power in the smart solar microgrid. The main contributions of this project are summarized as follows:

1. The employment of additional historical data, which are collected from the solar module, battery storage, grid data, and weather parameters, are utilized by the advanced metering infrastructure of the building’s smart solar microgrid.
2. The correlation matrixes of all collected features and predicted targets are computed for advanced analyses of selected inputs.
3. The hyperparameter optimization is performed to define the optimum structure of the CNN-Bi-GRU model in different time resolutions: one-hour, three-hour, and five-hour forward prediction. In addition, different sequential window sizes are evaluated for the effect on predicted accuracy.
4. The advantage performance of CNN-Bi-GRU is demonstrated through extensive comparisons with other deep learning models, which are the RNN, LSTM, GRU, Bi-LSTM, and Bi-GRU models.

In the next section, the prediction methodologies are described. Section 3 examines the hyperparameter of CNN-Bi-GRU and evaluate the effectiveness of various sequential window-size on different predicting resolution. In section 4, the experiment results of optimum structure CNN-Bi-GRU are compared with other deep learning models. Finally, section 5 is the conclusion of this paper.

II. METHODOLOGIES

The general framework of the proposed methodology is presented in the following steps:

Step 1: In Fig. 1, the collected data from AMI, including the solar data, the grid data, the battery data, and climate parameters, are utilized to predict the building’s three-phase load power in the smart solar microgrid. The 15-minute data is collected for average hourly data to retain consistency on all the different weather conditions in Taiwan. The collected data are accumulated every hour over more than one year, covering the converted ranks of X and Y variables. The Spearman correlation coefficient statistically measures the monotonic
The framework of the predicting methodology in smart solar microgrid.

relationship’s strength between two variables [47], [48]. The Spearman correlation is an original principle from the Pearson coefficient. The Spearman coefficient does not reflect the gradient of positive or negative signs. The strength of correlation could be represented for a different absolute range of $r_s$, where: 0.00-0.19 describes “very weak,” 0.20-0.39 describes “weak,” 0.40-0.59 describes “moderate,” 0.60-0.79 describes “strong,” and 0.80-1.00 describes “very strong” correlation.

$$\hat{x}_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}}$$

(1)

$$r_s = \frac{cov(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}}$$

(2)

**Step 3:** Based on the selected features, the hyperparameter optimization approach is implemented to define the optimum structure of the CNN-Bi-GRU model, which is trained with the first 80% of collected data, and the remaining 20% dataset is utilized in the validating process. In addition, the hyperparameter optimization also evaluates the effectiveness of various sequential input data on the predicting accuracy when combining the CNN with the Bi-GRU model.

**Step 4:** The following deep learning algorithms: RNN, LSTM, GRU, Bi-LSTM, Bi-GRU, and CNN-Bi-GRU, are utilized for training 10-fold CV approach to examine their performances. The 10-fold CV approach could evaluate the predicting accuracy without losing any information. The entire data is randomly split into ten equal separate combinations, in which nine parts are utilized for training, and the last one is applied for validation.

**Step 5:** The mean square error (MSE) and mean absolute error (MAE) are utilized to evaluate the performance of different deep learning algorithms, as presented in (3) and (4), where $y_i$ and $\hat{y}_i$ are the actual and predicted values. The models’ benchmarks are averaged across ten combinations of 10-fold cross-validation approaches.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

(3)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

(4)

**A. THE FEATURES SELECTION FOR PREDICTING MODEL**

The data in this project were accumulated from a smart solar microgrid that is integrated into the building in Taiwan. The AMI modules collected the data from solar, grid data, battery module, and weather factors in 14 months, which are described in Fig. 1. The AMI collected the solar data, which are solar voltage and solar current. Whereas the grid data, which includes the three-phase grid current and grid active power, were accumulated every fifteen minutes by the AMI. The battery module, the AMI gathered every fifteen minutes the battery voltage, battery charged power, battery discharged power, remaining battery power (kWh), and the battery remaining capacity (%). Moreover, the accumulated three-phase load powers are considered the target variable in the predicting method. To better understanding the correlation between collected variables, the correlation matrix that
The Spearman correlation map between collected features and target variables is presented in Table 1. The green color with a positive value represents the variable increases with the increase of other variables. In contrast, the blue color with a negative value symbolizes the decrement of the variable with the increment of others. The absolute values of the Spearman coefficient illustrate the strength of correlation between two variables.

In the collected data from the grid module’s AMI in Table 1, the three-phase load power correlates with the three-phase grid current, which is 0.79, 0.17, and 0.40, respectively. Moreover, the three-phase active power also highly increases with the increase of three-phase load power, which has the 0.71, 0.23, and 0.38 Spearman coefficients, respectively. These strong correlations are explained by the dispatch ability and response from the smart solar microgrid’s energy management system to meet the requirement of the load. In addition, the three-phase grid power controls the superior response with the building’s load power to maintain the power supply concerning other sources. Therefore, in the grid module, the three-phase grid current and active power are considered input features in predicting the three-phase load power.

In the battery module, the three-phase load power has weak correlations with the battery discharged and the battery remaining (kWh) which maintains the security of the battery storage power in case of power shortage. This means that
by the growth of three-phase load power, the smart solar microgrid will increase the discharge of the stored energy from the battery and decrease the remaining storage power in the battery module. In addition, the increment of the three-phase load power also has a weak negative correlation with the battery voltage because of discharging process to maintain the battery storage capability with supply and demand balance. The battery charged power negatively correlates with discharge power, which achieves a −0.4 Spearman coefficient, and could be considered input parameters in predicting model. The battery’s remaining capacity (%) has a very weak correlation with three-phase load power and is not included as an input feature of the proposed model. Therefore, in the battery module, the voltage, the discharged power, the charged power, and the remaining power (kWh) of the battery are involved as input features in predicting three-phase load power.

In the collected weather data from AMI, the temperature has a moderate correlation with three-phase load power, which obtains 0.1, 0.48, and −0.42 Spearman coefficients, respectively. Moreover, the temperature manages a weak correlation with relative humidity and wind speed, which acquire −0.24 and 0.25 Spearman coefficients. Although the precipitation has a very weak correlation with three-phase load power, this collected feature has a moderate correlation with relative humidity with a 0.44 Spearman coefficient; and maintains weak correlations with battery voltage, battery remaining (kWh), and battery remaining capacity (%), which obtains −0.29, −0.29, and 0.24 Spearman coefficients. Therefore, all the collected weather parameters from the AMI module are utilized in predicting the three-phase load power.

In the solar module, the AMI only collects the solar voltage and solar current, which obtain a very weak correlation with three-phase load power. However, the solar voltage and solar current acquire a strong correlation with collected data in the battery module. When the solar voltage and solar current grow, the smart solar microgrid enhances the battery voltage and battery-charged power. Consequently, the remaining battery power (kWh) also expands. In contrast, the battery discharged power also decreases when the solar energy increases. These correlations represent the adjusting of the imported solar power with the grid power and loads. Therefore, the smart solar microgrid concentrated on charging the battery module for future usage in case of a power shortage. In addition, the solar data also have a negative correlation with relative humidity and a positive relationship with temperature and wind speed, which achieve −0.42, −0.47, and 0.66, 0.32, 0.47, 0.56 Spearman coefficients, respectively. Therefore, the solar voltage and solar current could be contemplated in the proposed model to forecast the three-phase load power.

**B. THE SEQUENTIAL PREDICTING MODELS**

The recurrent neural network transforms sequential input data into sequential output by utilizing the step-by-step process for each component of the input sequence [49], [50]. The previous timestep is encoded into a memory hidden state in the RNN cell, which is upgraded at each training step. The encoded information in a hidden state could retain the memory of the previous state and learn long-term temporal dependencies in sequential data. At a specified timestep, the RNN cell produces the output vector, \( y_t \), based on the current input vector, \( x_t \), and the hidden state from the previous timestep, \( h_{t-1} \), and updates the current hidden state, \( h_t \).

The general structure of the RNN cell is presented in Fig. 2, and the common implementation is presented by the following equations, which \( W_h \), \( W_x \), \( W_y \), are the dense matrices and \( b_h \), \( b_y \) the bias vectors. The dense matrices and bias vectors are updated and optimized during the learning process of RNN.

\[
\begin{align*}
    h_t &= \text{tanh} \left( W_{hh} h_{t-1} + W_{hx} x_t + b_{hh} \right) \quad (5a) \\
    y_t &= \text{tanh} \left( W_{hy} h_t + b_{hy} \right) \quad (5b)
\end{align*}
\]

However, in the standard RNN architecture, the network output could exponentially blow up because of the influence of the given input, which is often referred to as the vanishing gradient problem [51]. Therefore, the Long Short-Term Memory (LSTM) was introduced by Hochreiter and Schmidhuber to overcome this problem [52]. The LSTM could manage long-term dependencies and determine the optimal time series problems. The LSTM could especially define the desirable features for short-term traffic prediction, which manage the relationship between a priori knowledge and the length of historical input. The LSTM architecture is multiplicative self-connected memory cells that include input, output, and forget gates. These gates provide a continuum of writing, reading, and resetting operations during the training process. These controllable gates could store and access the information for the long term and minimize the vanishing gradient problems. When the input gate remains close, the new input information could not overwrite the activation cell and could be available for a sequential process. The forget gate decides to maintain or overpass the information from the previous timestamps. Fig. 3 presents an illustration of general LSTM architecture, and the predicted output is iteratively updated by the
The general LSTM architecture.

The structure diagram of the GRU cell.

The general LSTM architecture.

The deep bidirectional LSTM/GRU neural network.

The structure diagram of the GRU cell.

The gated recurrent unit (GRU) includes a new type of memory cell, which was motivated by the LSTM in 2014 [53]. The GRU cell combines the forget and input gate into an update gate which merges the cell state and hidden state. Therefore, the GRU reduces the number of gates and decreases the parameters, which converge faster than traditional LSTM. The GRU could achieve better final solutions in some completed problems compared with the LSTM model [54]. The GRU cell, which contains only the reset and update gates, could deal with the exploding gradient during the training process. The GRU model, which acquires the most advantage over the LSTM model, proved the excellent power to memorize fluctuating temporal variables in multiple exogenous historical data [34]. The memory structure of the GRU cell is illustrated in Fig. 4, and the output of different gates is computed as followings, where \( r_t \) and \( z_t \) are the reset gate and update gate; \( W \) and \( b \) are the bias vectors and matrix weights; and \( \tilde{h}_t \) is the candidate status of

The following equations, where \( x^t \) is the sequential historical input, \( h_t \) is the hidden vector sequence, \( o_t \) is the output flow sequence, the \( W \) term denotes the weight matrices, \( b \) is the hidden bias vector, and \( \odot \) denotes the element-wise product.

\[
\begin{align}
    f_t &= \text{sigmoid} \left( W_h h^{t-1} + W_i x^t + b_f \right) \quad (6a) \\
    i_t &= \text{sigmoid} \left( W_i h^{t-1} + W_o x^t + b_i \right) \quad (6b) \\
    z_t &= \tanh \left( W_{cz} x^t + W_{ch} h^{t-1} + b_z \right) \quad (6c) \\
    o_t &= \text{sigmoid} \left( W_{ox} x^t + W_{oh} h^{t-1} + b_o \right) \quad (6d) \\
    c_t &= f_t \odot c_{t-1} + i_t \odot z_t \quad (6e) \\
    h_t &= o_t \odot \tanh \left( c_t \right) \quad (6f)
\end{align}
\]

The LSTM or GRU (LSTM/GRU) are only able to make use of previous information. The sequential historical data contained feature information that could be utilized in the predicting model. Therefore, the Bidirectional LSTM/GRU could process the historical data in both directions with two separate forward and backward layers. As presented in Fig. 5, a Bi-LSTM/GRU computes the forward sequence, \( \tilde{h}_t \) and backward sequence, \( \tilde{h}_t \) from the GRU or LSTM layers as in (8c). The combination of bidirectional feature and LSTM/GRU could process the long-range sequential data in both directions. The deep learning Bi-LSTM/GRU could be constructed by multiple stack LSTM/GRU layers, which transfer the output sequence of the layer to the input sequence of the next layers.

\[
\begin{align}
    \tilde{h}_t &= \text{sigmoid} \left( \tilde{W}_{oh} x^t + \tilde{W}_{oh} h^{t-1} + b_o \right) \quad (8a) \\
    \tilde{h}_t &= \text{sigmoid} \left( \tilde{W}_{ox} x^t + \tilde{W}_{oh} h^{t-1} + \tilde{b}_o \right) \quad (8b) \\
    y_t &= W_{\bar{h}} \tilde{h}_t + W_{\bar{h}} \tilde{h}_t + b_y \quad (8c)
\end{align}
\]

C. THE CONVOLUTIONAL NEURAL NETWORK

The convolutional neural network (CNN) is the most popular neural network, which conceptually resembles a multiple perceptron layer [55], [56]. The CNN is widely utilized in the feature extraction ability, which performs excellently in the image processing application. The CNN contains the feature extractor, composed of convolutional and pooling layers. The convolutional layers utilize the convolution kernels to obtain deeper patterns from sequential data and utilize the maximum pooling layers to achieve the key features. The filter number, the convolution stride, and the filter size decide the CNN layers’ performance. The convolutional layer
performs convolution operations on the sequential process of preceding filters. The spliced results from pooling layers are the output of the CNN which is reduced in size of the feature map. The pooling layers output the feature map by its maximum or average values, which advantageously provide smaller sampling features from the convolutional band. The leaky rectifier linear unit (LeakyRelu) is the activation function for CNN layers [57]. The pooling layers reduce the dimensional sequential data after the convolutional process, enhancing the calculating velocity and decreasing the overfitting probability for the next operations. The max-pooling layers are utilized in deep CNN structures in this project.

D. PROPOSED CONVOLUTIONAL NEURAL NETWORK BIDIRECTIONAL GATED RECURRENT UNIT (CNN-BI-GRU)
The convolutional neural network (CNN) is the most popular neural network, which conceptually resembles a multiple perceptron layer [55], [56]. The convolutional layers are utilized to extract the main feature in massive data, and the pooling layers are applied to reduce the dimensional sequential input. Therefore, the combination of convolutional layers and max-pooling layers improves the robustness of feature extraction in preprocessing sequential data. The primary extracted features from CNN layers are transferred to Bi-GRU layers for the continuous learning process. The hybrid of CNN and Bi-GRU is proposed and illustrated in Fig. 6. The CNN-Bi-GRU includes two stages, the first step is feature extraction based on the advantage of CNN layers, and the second stage is for three-phase load prediction. The CNN could effectively extract the essential patterns in time-series information, and the Bi-GRU can productively handle large amounts of sequential data, which provides better features from historical knowledge. The corresponding Bi-GRU could tackle the gradient disappearance or explosion in a traditional recurrent neural network in the long term and resolve the requirement to process both directional information. The output of the Bi-GRU layers is flattened into the feature vector, which transfers to the fully connected layers. The CNN-Bi-GRU is utilized in predicting the energy consumption for the building, which is integrated with the smart solar microgrid.

III. EVALUATING THE PERFORMANCE OF CNN-BI-GRU
In this section, hyperparameter optimization (HO) is utilized to acquire the optimum structure of the CNN-Bi-GRU model. These preprocessed hourly selected features are utilized in the proposed prediction method, and the evaluating results of CNN-Bi-GRU are obtained. For better prediction accuracy, hyperparameter optimization is utilized to achieve the optimum structure of neural networks. A simple grid search is utilized to optimize the network structure. Although the simple grid search could suffer the curse of dimensionality of the configuration space, it could evaluate the effectiveness of every setting parameter on the performance of the CNN-Bi-GRU model. In this experiment, the variant of sequential historical input data, the number of CNN layers, the number of CNN filters in each layer, the number of Bi-GRU layers, and the neural nodes in each Bi-GRU layer are examined, as illustrated in Table 2. The HO was evaluated for the number of sequential inputs 36, 48, and 60. For CNN and Bi-GRU layers, the HO evaluates the variants of 2 and 3 layers in the structure of the CNN-Bi-GRU model.

![FIGURE 6. The overall framework of the proposed CNN-Bi-GRU architecture.](image)

TABLE 2. The summary of the setting parameter of CNN-Bi-GRU in the hyperparameter optimization process.

| Setting parameters               | Variant values |
|----------------------------------|----------------|
| Number of sequential inputs      | 36, 48, 60     |
| Number CNN Layers                | 2, 3           |
| Number of CNN filter             | 32, 64         |
| Number Bi-GRU Layers             | 2, 3           |
| Number Bi-GRU Neural             | 32, 64, 128    |
| Activation function              | ReLU           |
| Learning rate                    | 0.001          |
| Kernel initialization            | Uniform        |
| Batch size                       | 32             |
| Number of epochs                 | 50             |
number of CNN filters is either 32 or 64, while the number of Bi-GRU neural nodes varies from 32, 64, and 128. The number of training epochs is 50, and the value of the learning rate is 0.001. There are a total of 72 combinations which are analyzed the relative impact of each setting parameter against the predicting accuracy. The experiments attempt to prove how the parameter importance’s variant becomes the benchmark for the performance of the prediction model.

FIGURE 7. The effectiveness of hyperparameter optimization on the benchmark metrics in training and validating processes of the proposed CNN-Bi-GRU model: (a) Effect of Bi-GRU Layers; (b) Effect of the Bi-GRU neural nodes.
In this experiment, the variants of each selected parameter are analyzed to evaluate the significant relative impact on the statistic benchmarks of the proposed CNN-Bi-GRU model. Fig. 7 and 8 illustrate the impacts of each selected parameter against the MSE and MAE benchmarks during the training and validating processes. The average of each benchmark metric also proves how the essential parameters affect the performance of the proposed CNN-Bi-GRU model. Fig. 7(a) presents the impact of Bi-GRU layers against the value of MSE and MAE metrics. The number of Bi-GRU layers varies from 2 to 3, which impacts the predicting accuracy. The experiment results prove that the increment of Bi-GRU will decrease the predicting performance of the proposed model. Therefore, the two layers of Bi-GRU achieve the highest accuracy of both MSE and MAE values in training and validating processes.

Fig. 7(b) illustrated the impact of hyperparameter neural nodes on the CNN-Bi-GRU’s performance. The number of
Bi-GRU neural nodes varies from 32, 64, and 128 to evaluate the accuracy of three-phase load power prediction. Fig. 7(b) proves that this parameter has a significant influence on the predicting performance, which increases the stability of the learning process. The higher value of neural nodes increases the predicting accuracy of the CNN-Bi-GRU model. However, the increment of neural nodes magnifies the training process due to the bi-directional operations. The MAE benchmark trends to oscillate with a higher neural node in the validating procedure. However, the 128-neural node could be considered an appropriate parameter for the CNN-Bi-GRU model, which achieves the most significant performance during the training operations compared with 32 and 64 setting values.

Fig. 8(a) presents the box plot chart of the window size in predicting the three-phase load power. According to the mean of MSE and MAE benchmarks which are collected from simulation results, the numbers of historical data tend to improve the accuracy of the proposed methodology. The lower quartile data of MSE and MAE in the validating process are also increased at a high window size. Although the increment of window size expands the complexity of the calculating process and computing resources, the predicting accuracy could obtain better performance in return. The 60-window size acquires the most significant accuracy in training operation compared with 36 and 48 sequential input data. Therefore, this window size is selected as a suitable parameter in the construction of the CNN-Bi-GRU model.

The HO was performed by combining the number of CNN layers with the window size, the number of Bi-GRU neural nodes, and the number of Bi-GRU layers. The box plot performance of variant CNN layers is illustrated in Fig. 8(b), the box plot graphs prove that the higher CNN layers provide a better MSE and MAE in training operations. The increment of CNN layers could reduce the number of sequential data for the Bi-GRU layers, which shows a faster convergent velocity in the proposed methodology. However, the higher CNN layers tend to oscillate the performance of the CNN-Bi-GRU model in the validating MSE process. Therefore, the 3-CNN layers are selected as appropriate parameters in the architecture of the CNN-Bi-GRU model.

The comparison of different CNN filters on the predicting performance is presented in Fig 8(c). From the mean and median of quartile data, the higher CNN filter tends to enhance the performance of the CNN-Bi-GRU model in the training process. The lower quartile data in the box plot proves the effectiveness of increment CNN filters in predicting the three-phase load power. The number of CNN filters is the most influential parameter in the increase of predicting performance. However, the addition of a CNN filter could fluctuate the MSE value in the validating process, which increases the higher quartile data. The continuous increment of the CNN filter will decrease the performance and stability of the proposed model. Moreover, the increment of the CNN filter will enlarge the computed operations in the Bi-GRU layers, which substantially impact the prediction sensitivity. Therefore, the 64 CNN filter is the most appropriate parameter in the HO process of the CNN-Bi-GRU model.

The HO process is performed to specify the most optimum parameters for the architecture of the CNN-Bi-GRU model, which are the window size, the CNN layers, the CNN filter, the Bi-GRU layers, and the Bi-GRU neural nodes. These selected parameters are utilized in predicting the three-phase load power in the smart solar microgrid. The comparative performances of the proposed architecture CNN-Bi-GRU with other persistent models are presented in the next section.

IV. COMPARATIVE PERFORMANCE OF CNN-BI-GRU MODEL

In this part, the performance of the CNN-Bi-GRU model is evaluated and compared with other benchmark models. The benchmark models are constructed with the same optimum structure as the CNN-Bi-GRU model. The comparative table is presented in Table 3. The average benchmarks between methodologies are summaries to evaluate the accuracy and performance.

In Table 3, the average MSE values during the training process with 10-fold cross-validation are 0.00798, 0.00471, 0.00484, 0.00441, 0.00421, 0.00224 for the RNN, LSTM, Bi-LSTM, GRU, Bi-GRU, and CNN-Bi-GRU, respectively. For the average MAE benchmark in the training operation, the RNN, LSTM, Bi-LSTM, GRU, Bi-GRU, and CNN-Bi-GRU achieve 0.03790, 0.02709, 0.02760, 0.02712, 0.02672, and 0.01912, respectively. In the training process, the CNN-Bi-GRU obtains the greatest improvements of 71.88% and 49.543% for MSE and MAE, respectively. The improvement is explained by not only the additional CNN layers in extracting essential data features but also the forward and backward operations of Bi-GRU layers. The combination between CNN and Bi-GRU achieves higher performances than all other methodologies in predicting the three phases of load power in the smart solar microgrid.

In addition, the performances of the validating process are also evaluated and compared. For average MSE benchmarks, the RNN, LSTM, Bi-LSTM, GRU, Bi-GRU, and CNN-Bi-GRU acquire 0.00614, 0.00635, 0.00606, 0.00636, 0.00633, and 0.00614, respectively. Using the 10-fold cross-validation, the RNN, LSTM, Bi-LSTM, GRU, Bi-GRU, and CNN-Bi-GRU achieve an average MAE of 0.03409, 0.02709, 0.02760, 0.02712, and 0.01912, respectively. In the training process, the CNN-Bi-GRU achieves the highest enhancement of 3.405% MSE and 11.574% MAE during the validating operation, respectively. The higher performances indicate the effectiveness of CNN layers in extracting important historical data patterns. Moreover, the bidirectional operations also generate a significant impact on the performance of the CNN-Bi-GRU model compared with other traditional methodologies.

With the same network structure and 10-fold cross-validation data, the CNN-Bi-GRU model acquires better achievements compared with the original Bi-GRU model.
The additional CNN layers obtain enhancements of 46.720% MSE, 28.442% MAE in training operation, and 2.966% MSE, 6.245% MAE in the validating process. The CNN extracts the important data features with two layers and transfers the results to the Bi-GRU layers. Moreover, the proposed CNN-Bi-GRU method obtains better performances than the original GRU model, with the greatest improvement of 49.102% MSE, 29.505% MAE in training operation, and 3.405% MSE, 0.022% MAE in the validating process. The predicting accuracy of the CNN-Bi-GRU model outperforms not only the Bi-GRU method but also the GRU, LSTM, Bi-LSTM, and RNN models. The enhancing performance of 10-fold cross-validation in the proposed CNN-Bi-GRU method is due to the fit selection of structure and appropriate window size during the HO process. By combining the feature extraction of CNN layers, a suitable window size, as well as an optimized Bi-GRU structure, the proposed CNN-Bi-GRU can improve the predicting performance, accuracy, and stability in the 10-fold cross-validation.

Fig. 9 illustrates the box plot chart of training and validating operations with 10-fold cross-validation in different methodologies. According to the median experiment performance, the proposed model can improve the performance by adding the number of CNN layers to two layers. In addition, the lower quartiles MSE and MAE of CNN-Bi-GRU prove the effectiveness of extracting data features before applying them in the backward and forward processes of Bi-GRU layers. In addition, the performance of the proposed CNN-Bi-GRU is more stable than other methodologies with a small range of MSE and MAE during the training and validating operations. Although other methodologies could achieve better values of MSE and MAE in the validating process, the proposed CNN-Bi-GRU could acquire a smaller range of benchmarks which proves the stability in predicting three-phase load power with 10-fold cross-validation. Therefore, a comparison of some methods with feature extraction demonstrates the CNN-Bi-GRU model resulting in better predicting stability compared with traditional Bi-GRU, GRU, and other methods. Fig. 10 illustrates the comparative MSE values in testing data with different predicting horizons. For one-hour ahead horizon, the CNN-Bi-GRU obtains the 0.00214 MSE, which is the highest performance compared with other methods. In 3-hour ahead, the CNN-Bi-GRU outperforms other methodologies with a 0.00206 MSE benchmark. For 5-hour ahead prediction, the CNN-Bi-GRU surpasses the traditional methods with 0.00194 MSE. Therefore, the proposed CNN-Bi-GRU method acquires the highest accuracy, which proves the performance enhancement for predicting the three-phase load power in the smart solar microgrid.
V. CONCLUSION

This research provides some essential contributions in predicting the three-phase load power with a deep learning machine. (1) the proposed CNN-Bi-GRU model, utilizing the CNN for extracting important features in sequential historical data and bidirectional operations in Bi-GRU, improves the accuracy and stability in predicting experiments. (2) The hyperparameter optimization illustrates the significant impact of important parameters on predicting accuracy. (3) The results of optimized CNN-Bi-GRU architecture and the appropriate window size of sequential historical data could be utilized in developing models with other applications. The experiment results prove that the proposed CNN-Bi-GRU could be successfully utilized to predict the three-phase load power in the efficient management system of the smart solar microgrid.

REFERENCES

[1] H. S. Hippert, C. E. Pedreira, and R. C. Souza, “Neural networks for short-term load forecasting: A review and evaluation,” IEEE Trans. Power Syst., vol. 16, no. 1, pp. 44–55, Feb. 2001.

[2] G. Gross and F. D. Galiana, “Short-term load forecasting,” Proc. IEEE, vol. 75, no. 12, pp. 1558–1573, Dec. 1987.

[3] M. A. Mahmud, “Isolated area load forecasting using linear regression analysis: Practical approach,” Energy Power Eng., vol. 3, no. 4, pp. 547–550, 2011.

[4] J. W. Taylor and P. E. McSharry, “Short-term load forecasting methods: An evaluation based on European data,” IEEE Trans. Power Syst., vol. 22, no. 4, pp. 2213–2219, Nov. 2007.

[5] Z. Wang and J. Wang, “Self-healing resilient distribution systems based on sectionalization into microgrids,” IEEE Trans. Power Syst., vol. 30, no. 6, pp. 3139–3149, Nov. 2015.

[6] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, “Short-term load forecasting with deep residual networks,” IEEE Trans. Smart Grid, vol. 10, no. 4, pp. 3943–3952, Jul. 2018.

[7] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, “Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting,” IEEE Access, vol. 7, pp. 88058–88071, 2019.

[8] J. Kim, J. Moon, E. Hwang, and P. Kang, “Recurrent inception convolution neural network for multi short-term load forecasting,” Energy Buildings, vol. 194, pp. 328–341, Jul. 2019.

[9] Y. Liang, D. Niu, and W.-C. Hong, “Short term load forecasting based on feature extraction and improved general regression neural network model,” Energy, vol. 166, pp. 653–663, Jan. 2019.

[10] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, “Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series,” Energy, vol. 175, pp. 365–377, May 2019.

[11] M. S. Hossain and H. Mahmood, “Short-term load forecasting using an LSTM neural network,” in Proc. IEEE Power Energy Conf. Illinois (PECI), Feb. 2020, pp. 1–6.

[12] K. Passini, M. Khoutari, A. Same, F. Gananias, and L. Oukhellou, “LSTM encoder-predictor for short-term train load forecasting,” in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, 2019, pp. 535–551.

[13] Y. Jin, H. Guo, J. Wang, and A. Song, “A hybrid system based on LSTM for short-term power load forecasting,” Energies, vol. 13, no. 23, p. 6241, Nov. 2020.

[14] J. Li, D. Deng, J. Zhao, D. Cai, W. Hu, M. Zhang, and Q. Huang, “A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network,” IEEE Trans. Ind. Informat., vol. 17, no. 4, pp. 2443–2452, Apr. 2021.

[15] J. Li, A. Mohamed, G. Zweig, and Y. Gong, “LSTM time and frequency recurrence for automatic speech recognition,” in Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU), Dec. 2015, pp. 187–191.

[16] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schluter, and H. Ney, “A comprehensive study of deep bidirectional LSTM RNNs for acoustic modeling in speech recognition,” in Proc. IEEE Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2017, pp. 2462–2466.

[17] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, “Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting,” Energies, vol. 13, no. 2, p. 391, Jan. 2020.

[18] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, “Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster,” in Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. (EAIT), Jan. 2018, pp. 1–4.

[19] W. Buyon, T. M. Brefel, F. Raus, and M. Liwicki, “Scene labeling with LSTM recurrent neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3547–3555.

[20] S. Ma and Y. Han, “Describing images by feeding LSTM with structural words,” in Proc. IEEE Int. Conf. Multimedia Expo. (ICME), Jul. 2016, pp. 1–6.

[21] D. Guo, W. Zhou, H. Li, and M. Wang, “Hierarchical LSTM for sign language translation,” in Proc. AAAI Conf. Artif. Intell., vol. 32, no. 1, 2018, pp. 6845–6852.

[22] Y. Cui, S. Wang, and J. Li, “LSTM neural reordering feature for statistical machine translation,” 2015, arXiv:1512.00177.

[23] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “Short-term residential load forecasting based on LSTM recurrent neural network,” IEEE Trans. Smart Grid, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[24] C. Zhao and S. Long-Xiang, “Short-term electrical load forecasting based on deep learning LSTM networks,” Electron. Technol., vol. 10, no. 39, pp. 122–125, 2018.

[25] R. Jiao, T. Zhang, Y. Jiang, and H. He, “Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network,” IEEE Access, vol. 6, pp. 59438–59448, 2018.

[26] D. L. Marino, K. Amarasinghe, and M. Manic, “Building energy load forecasting using deep neural networks,” in Proc. IEECON 42nd Annu. Conf., IEEE Ind. Electron. Soc., Oct. 2016, pp. 7046–7051.

[27] R. Dey and F. M. Salem, “Gate-variants of gated recurrent unit (GRU) neural networks,” in Proc. IEEE 60th Int. Midwest Symp. Circuits. Systems (MWSCAS), Aug. 2017, pp. 1597–1600.

[28] J. Zheng, X. Chen, K. Yu, L. Gan, Y. Wang, and K. Wang, “Short-term power load forecasting of residential community based on GRU neural network,” in Proc. Int. Conf. Power Syst. Technol. (POWERCON), Nov 2018, pp. 4862–4868.

[29] G. Xiuuyun, W. Ying, G. Yang, S. Chengzhii, X. Wen, and Y. Yimiao, “Short-term load forecasting model of GRU network based on deep learning framework,” in Proc. 2nd IEEE Conf. Energy Internet Energy Syst. Integr. (EI2), Oct. 2018, pp. 1–4.

[30] D. Li, G. Sun, S. Miao, Y. Gu, Y. Zhang, and S. He, “A short-term electric load forecast method based on improved sequence-to-sequence GRU with adaptive temporal dependence,” Int. J. Electr. Power Energy Syst., vol. 137, May 2022, Art. no. 107627.

[31] S. Jung, J. Moon, S. Park, and E. Hwang, “An attention-based multilayer GRU model for multistep-ahead short-term load forecasting,” Sensors, vol. 21, no. 5, p. 1639, Feb. 2021.

[32] P. Jia and Y. Miao, “Research on short-term forecast of power load based on ADAM-GRU,” J. Phys. Conf. Ser., vol. 2020, pp. 1693, no. 1, 2020, Art. no. 12116.

[33] W. Wu, W. Liao, J. Miao, and G. Du, “Using gated recurrent unit network to forecast short-term load considering impact of electricity price,” Energy Proc., vol. 158, pp. 3369–3374, Feb. 2019.

[34] W. Zheng and G. Chen, “An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization,” IEEE Trans. Cybern., early access, Nov. 3, 2021, doi: 10.1109/TCYB.2021.3121312.

[35] M. Alhussein, K. Ang徘徊, S. K. Haidar, “Hybrid CNN-LSTM model for short-term individual household load forecasting,” IEEE Access, vol. 8, pp. 180544–180557, 2020.

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

[37] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, “On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach,” IEEE Access, vol. 9, pp. 31191–31212, 2021.
[38] S. H. Rafi, S. R. Deeba, and E. Hossain, “A short-term load forecasting method using integrated CNN and LSTM network,” IEEE Access, vol. 9, pp. 32436–32448, 2021.

[39] X. Guo, Q. Zhao, D. Zheng, Y. Ding, and Y. Gao, “A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price,” Energy Rep., vol. 6, pp. 1046–1053, Dec. 2020.

[40] H. Choi, S. Ryu, and H. Kim, “Short-term load forecasting based on ResNet and LSTM,” in Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm), Oct. 2018, pp. 1–6.

[41] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, “A novel CNN-GRU-Based hybrid approach for short-term residential load forecasting,” IEEE Access, vol. 8, pp. 143759–143768, 2020.

[42] L. Wu, C. Kong, X. Hao, and W. Chen, “A short-term load forecasting method based on GRU-CNN hybrid neural network model,” Math. Problems Eng., vol. 2020, pp. 1–10, Mar. 2020.

[43] M. Shen, Q. Xu, K. Wang, M. Tu, and B. Wu, “Short-term bus load forecasting method based on CNN-GRU neural network,” in Proc. Purple Mountain Forum Int. Forum Smart Grid Protection Control, 2020, pp. 711–722.

[44] M. Afrasiabi, M. Mohammadi, M. Rastegar, and A. Kargarian, “Probabilistic deep neural network price forecasting based on residential load and wind speed predictions,” IET Renew. Power Gener., vol. 13, no. 11, pp. 1840–1848, Aug. 2019.

[45] M. Suresh, M. S. Anbarasi, J. Divyabharathi, D. Harshavardeni, and S. Meena, “Household electricity power consumption prediction using CNN-GRU techniques,” in Proc. Int. Conf. Comput., Autom. Netw. (ICSCAN), Jul. 2021, pp. 1–6.

[46] M. Bashari and A. Rahimi-Kian, “Forecasting electric load by aggregating meteorological and history-based deep learning modules,” in Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM), Aug. 2020, pp. 1–5.

[47] L. J. Myers and M. Siros, “Spearman correlation coefficients, differences between,” Encyclopedia Stat. Sci., vol. 12, Jul. 2004.

[48] J. H. Zar, “Spearman rank correlation,” Encyclopedia Biostatistics, vol. 7, Jul. 2005.

[49] W. Zaremba, I. Sutskever, and O. Vinyals, “Recurrent neural network regularization,” 2014, arXiv:1409.2329.

[50] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, “Recurrent neural network based language model,” in Proc. Interspeech, vol. 2, no. 3, 2010, pp. 1045–1048.

[51] A. Graves, “Hierarchical subsampling networks,” in Supervised Sequence Labelling With Recurrent Neural Networks. Cham, Switzerland: Springer, 2012, pp. 109–131.

[52] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[53] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” 2014, arXiv:1406.1078.

[54] J. Chang, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” Tech. Rep., Dec. 2014.

[55] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Netw., vol. 61, pp. 85–117, Jan. 2014.

[56] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, “Convolutional neural networks for time series classification,” J. Syst. Eng. Electron., vol. 28, no. 1, pp. 162–169, Feb. 2017.

[57] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1026–1034.

PHUONG NGUYEN THANH received the master’s degree from the Royal Melbourne Institute of Technology (RMIT), in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, National Kaohsiung University of Science and Technology. His research interests include artificial applications in smart power systems, utilizing the neural network models in some predicting applications in the smart solar microgrid, and developing neural network models for processing sequential data.

MING-YUAN CHO received the M.S. and Ph.D. degrees in electrical engineering from National Sun Yat-sen University, Kaohsiung, Taiwan, in 1989 and 1992, respectively. He is with the Department of Electrical Engineering, National Kaohsiung University of Science and Technology (NKUST), where he is currently a Distinguished Professor associated with the Director of the Energy Technology Research Institute. He was the Chairman of the Department of Electrical Engineering, NKUST, from August 2000 to July 2003, where he was also the Dean of the Electrical and Information College, from August 2015 to July 2018. His research interests include smart grid, micro grid, AI applications to power systems, and energy saving technologies.

CHUN-LUNG CHANG received the degree from the Department of Electrical and Electronic Engineering, National Kaohsiung Institute of Technology, in 1994. His research interests include intelligent electrical networks and micro grid.

MENG-JIE CHEN received the master’s degree in electrical engineering from the National Kaohsiung University of Science, in 2017. His research interests include intelligent electrical networks and micro grid.