IEEE *Access*
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

# Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting

**HUI HWANG GOH[1], (Senior Member, IEEE), BILIANG HE[1], HUI LIU[1], (Senior Member, IEEE), DONGDONG ZHANG[1], (Member, IEEE), WEI DAI[1], (Member, IEEE), TONNI AGUSTIONO KURNIAWAN [2], AND KAI CHEN GOH [3]**

[1]College of Electrical Engineering, Guangxi University, Nanning 53000 China
[2]College of Environment and Ecology,Xiamen University, Fujian 361102, PR China
[3]Department of Technology Management, Faculty of Construction Management and Business, University Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia

Corresponding author: Hui Hwang Goh (hhgoh@gxu.edu.cn)

**ABSTRACT** Load forecasting is critical for power system operation and market planning. With the increased penetration of renewable energy and the massive consumption of electric energy, improving load forecasting accuracy has become a difficult task. Recently, it was demonstrated that deep learning models perform well for short-term load forecasting (STLF). However, prior research has demonstrated that the hybrid deep learning model outperforms the single model. We propose a hybrid neural network in this article that combines elements of a convolutional neural network (1D-CNN) and a long short memory network (LSTM) in novel ways. Multiple independent 1D-CNNs are used to extract load, calendar, and weather features from the proposed hybrid model, while LSTM is used to learn time patterns. This architecture is referred to as a CNN-LSTM network with multiple heads (MCNN-LSTM). To demonstrate the proposed hybrid deep learning model's superior performance, the proposed method is applied to Ireland's load data for single-step and multi-step load forecasting. In comparison to the widely used CNN-LSTM hybrid model, the proposed model improved single-step prediction by 16.73% and 24-step load prediction by 20.33%. Additionally, we use the Maine dataset to verify the proposed model's generalizability.

**INDEX TERMS** short-term load forecast, deep learning, multi-head CNN-LSTM, multi-step load prediction.

## I. INTRODUCTION

Electricity is a significant secondary energy source that has a sizable impact on both the national economy and daily life. With large-scale renewable energy connected to the power grid, we can expand the use of renewable resources, which has a number of benefits. However, there are additional complications associated with power distribution and dispatch. Short-term load forecasting (STLF) enables the generation of detailed forecasts for a time period that is critical for supply unit optimization, economy dispatch, and market transactions [1]. A detailed and reliable STLF enables the power system to operate safely and efficiently.

Due to the complexity of the factors affecting STLF, such as historical load curves, weather conditions, and calendar effects, staying on schedule is more difficult than ever. Numerous weather research experiments are currently being con-

ducted to examine the four critical STLF variables: temperature, humidity, wind speed, and air pressures. Mukhopadhyay et al [2] incorporated temperature and relative humidity into the load forecasting model. Friedrich and Afshari [3] included wind speed in their load forecasting model. Because the weather is influenced by both climate and geography, some of its forecast parameters are non-relational. Correlations between weather parameters and load data must be determined in order to determine which parameters (variables) to measure [4].The calendar effect was used to characterize the cyclical and seasonal variation in electricity consumption [5]. People's production and daily lifestyle changes can affect their energy consumption patterns, resulting in daily or weekly changes in the power load. Numerous calendar-dependent input features were incorporated into the load forecasting model to emphasize the importance of calendar

**IEEE** *Access*

H. H. Goh *et al.*: Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting

variables via coding techniques.

Historically, a variety of load forecasting algorithms were employed to determine historical loads and load-related characteristics and produced accurate prediction results. Recently, deep learning models based on long short memory networks (LSTMs) and convolutional neural networks (CNNs) have been applied to anticipate wind energy generation, photovoltaic energy generation, and load forecasting. Their predicting abilities have been demonstrated to be superior to those of conventional shallow neural networks and machine learning techniques. The LSTM model accurately captures the time series' pattern information, whereas the CNN model extracts valuable features from the time series without requiring domain expertise. LSTMs are excellent at extracting temporal features, while CNNs are excellent at extracting spatial features [6]. Thus, by integrating the technological advantages of the two models, forecasting accuracy can be increased.

The periodicity of electric load and the effect of weather parameters on electric load are discussed in this paper. As input features, periodic calendar coding and weather parameters (temperature and humidity) are used. Different from other hybrid models of CNN and LSTM [7], [8], we propose a novel combined architecture of CNN and LSTM for STLF. We use three independent 1-dimensional convolution heads to extract historical load features, calendar features and weather features, and fuse the features into LSTM layer. This hybrid model is called multi-head CNN-LSTM(MCNN-LSTM). The independent convolutional head can concentrate on the input data, which can aid in feature extraction, and each independent convolutional head can finely tune the parameters to adapt to the feature extraction. At the same time, it can avoid the slow fitting speed caused by the increase of input feature dimension, thus reducing the training time. Currently, most of the methods used in deep learning focus operate on single-step load forecasting, while power market participants rely on multi-step forecasting. For multi-step load forecasting, we also utilize the proposed multi-input multi-output model. The main contributions of this paper are as follows:

1. We propose a novel MCNN-LSTM hybrid deep learning model for short-term load forecasting.

2. We compare the proposed model's performance to that of other machine learning models and deep learning models, demonstrating the model's validity.

3. We use polar coordinates coding to encode calendar variables, and analyze the impact of calendar coding on load forecasting at different time steps.

The rest of this paper is organized as follows. Section II describes related forecasting techniques. Section III details the basic framework of MCNN-LSTM hybrid deep learning model. Section IV gives the evaluation indicators of STLF model. Section V contains data processing and simulation analysis of experiments. Section VI presents the conclusions.

## II. RELATED WORK

STLF methods can be divided into two categories: statistical methods and artificial intelligence (AI) methods. The main statistical methods are linear regression (LR) [9], multiple linear regression (MLR) [10], auto regressive moving average (ARIMA) [11]. The goal of statistical methods is to measure changes in both current and past load series to establish a mathematical relationship. While time series correlation is considered, it is difficult to attain the requisite forecasting accuracy when working with non-stationary load series due to their inherent restrictions [1]. Because of the introduction of random and intermittent energy sources in the grid, the traditional statistical approach does not work for the complex load curve. The use of machine learning models that overcame the shortcoming of statistical methods in STLF models were widely implemented for this class of problems, including support vector machine (SVM) [12], [13], light gradient boosting machine (LightGBM) [14] and artificial neural network(ANN) [15]–[17]. SVM's primary methodology is based on the principle of structural risk minimization. While SVM's time series performance is a strength, the accuracy with which the parameters are chosen has an effect on performance. LightGBM is a promotion model for gradient boosted decision trees (GBDTs) that allows for equal gains in speed and accuracy. ANN makes it much easier to track the relationship between input and output variables due to the nonlinear processing. The ANN combined with a multi-layer perceptron (MLP) has become one of the most widely used load forecasting algorithms. However, the freedom of an ANN is constrained by its initial conditions and increases as the model becomes more complex, resulting in overfitting or underfitting [17].

This is a new type of machine learning application known as deep learning. In comparison to the shallow network, big data and unsupervised learning have garnered considerable interest. Deep learning, particularly convolutional neural networks (CNN), has aided in the study of STLF [18], [19] and long-term memory network(LSTM) [20], [21]. CNN's local connection and global sharing capabilities significantly reduce the model's training parameters and training time, with some organizations utilizing CNN for load prediction. Dong et al [22] used CNN and K-means clustering to process large-scale power load data. Cai et al [23] used gated CNN to forecast the short-term power consumption of commercial buildings, which has the best performance compared with gated recurrent neural network and seasonal ARIMAX. LSTM is a recurrent neural network, which is suitable for processing time series. Rahman and Zubair [24] presented a method of applying LSTM with data construction method for hourly electric load forecasting. Bedi and Toshniwal [25] proposed a deep learning framework based on LSTM to forecast electricity demand by learning history data dependencies. Sometimes a single model is not competent for all tasks, and deep learning is no exception. In order to take advantage of different models, some researchers have mixed different deep learning models into a hybrid network model

[7], [14]. Alhussein et al [26] proposed a hybrid CNN-LSTM model using CNN layer to extract features from the input data and LSTM layer for sequence learning to forecast household load. Kim et al [27] proposed a recurrent inception convolution neural network (RICNN) that combines RNN and 1-dimensional CNN (1-D CNN) to forecast electricity load of three distribution industrial complexes in South Korea. The hybrid CNN and LSTM model has good performance in small-scale load forecasting of commercial buildings and individual household, while the STLF of large-scale regional level needs to be studied.

To capitalize on the CNN and LSTM hybrid model's advantages, many implementations of CNN and LSTM hybrid deep learning models have been developed in a variety of domains. Bixuan et al [6] compared the forecasting performance of five different CNN-LSTM structures and proposed that the solar irradiance features be decomposed using complete ensemble empirical mode decomposition adaptive noise (CEEMDAN), and the prediction model of historical solar irradiance features obtained by multiple CNNs be fused into LSTM for optimal performance. Behnam et al [28] introduced a parallel deep LSTM-CNN (PLCNet) model that relied on the parallel CNN layer and LSTM as the upper layer to extract the features of load data, and then connected the LSTM layer and the Dense layer to anticipate the final load data. Hourly load forecasting accuracy was 98.23% on the Malaysian dataset and 91.18% on the German dataset.The hybrid model of CNN and LSTM enables the incorporation of additional data features. Taoying et al [29] introduced a cascaded CNN-LSTM model for PM2.5 prediction and demonstrated that the multivariate CNN-LSTM model is more accurate than the univariate CNN-LSTM model. Kejun et al [8] proposed the LSTM-Convolutional Network forecasting model for photovoltaic power generation. To begin, an LSTM layer is utilized to extract the temporal features affecting photovoltaic power, followed by a CNN layer to retrieve the spatial data. It is more accurate than a single CNN model or LSTM model. While our proposed model incorporates historical load and load-related characteristics and makes use of the CNN and LSTM hybrid model's multivariate feature processing capabilities, it also addresses the issue of training costs associated with the increased data dimension generated by multivariate features. It contains important references for tasks involving load forecasting with different input features.

## III. THE PROPOSED FRAMEWORK
### A. 1-DIMENSIONAL CONVOLUTION NEURAL NETWORK
Convolutional neural network (CNN) was originally developed for image classification, and have been used in the field of computer vision with great success. Recently, CNNs are not only used in the field of natural language, but they are also extend to solving energy related problems, such as power quality issues [30], wind power prediction [31], solar irradiation prediction [32] and load prediction [23] etc. In comparison to a traditional full connected network, CNN

has two main characteristics: local connection and weight sharing. The structure of CNN mainly includes convolution layer, pooling layer and full connection layer. Figure 1(a) shows the basic framework of one-dimensional CNN.
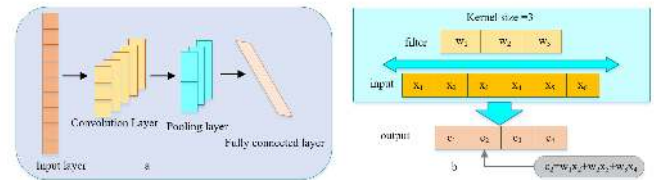


**FIGURE 1.** (a)The basic structure of 1-Dimensional convolution neural network; (b) operation of 1-Dimensional convolution.

Unlike a two-dimensional convolution image processing kernel, which moves in two dimensions from left to right and from top to bottom, the one-dimensional convolution kernel moves in just single dimension (time dimension). Time series can be defined as one-dimensional vector, and one-dimensional CNN is suitable for learning the characteristics of time series. In this study, CNN is applied for feature extraction of load forecasting. The operation example of one-dimensional convolution is shown in Figure 1(b). Convolution is the core concept of CNN, which is used to combine two sets of information into one. Convolution transforms the input data, and the convolution calculation formula is as follows:

$$c = f(w_n \otimes x + b_n) \tag{1}$$

where w, n, x and b denote the weighting factors in kernels, number of kernels, input series and bias. The symbol $\otimes$ signifies the convolution operation. f is the activation function. The Rectified Linear Unit (ReLU) can effectively deal with the vanishing gradient problem and make network more trainable. It is adopted as the activation function in this paper. The formula is as follow:

$$f(x) = max(0, x) \tag{2}$$

Pooling layer is mainly used to reduce the number of parameters. In this paper, maximum pooling is adopted.

### B. LONG SHORT-TERM MEMORY NETWORK
In the traditional feedforward neural network, information flows from the input node to the hidden layer and then to the output layer in one direction, so the neural network cannot remember the input of different time series. Recursive network (RNN) can make use of internal state storage to realize memory function. Long short-term memory (LSTM) network is a special RNN, which can learn long-term dependence. It was proposed by Hochreiter and schmidhub in 1997 [33], and has been improved and promoted by many people in the following works. It uses more complex internal structure cells instead of primitive low cell neurons to solve

**IEEE** *Access*

H. H. Goh *et al.*: Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting

the problem of gradient explosion and gradient disappearance of primitive RNN. The basic unit of LSTM is shown in Figure 2.
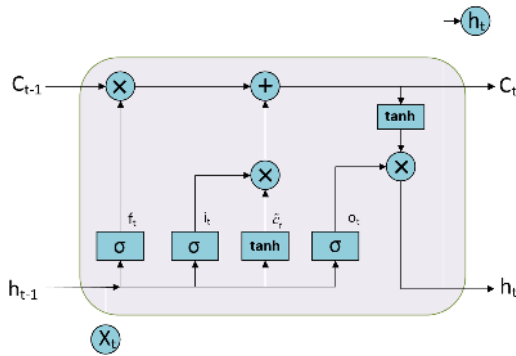


FIGURE 2. The basic LSTM cell.

There are four important elements in the cell structure of LSTM: input gate, forgetting gate, output gate and cell state. Input gate, forgetting gate and output gate are used to update, maintain and delete the information contained in the status of control unit. The LSTM model updates the cell status and calculates the output as follows:

$$i_t = \sigma(w_i h_{t-1} + w_i x_t + b_i) \tag{3}$$

$$f_t = \sigma(w_f h_{t-1} + w_f x_t + b_f) \tag{4}$$

$$o_t = \sigma(w_o h_{t-1} + w_o x_t + b_o) \tag{5}$$

$$\widetilde{c}_t = tanh(w_c h_{t-1} + w_c x_t + b_c) \tag{6}$$

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{7}$$

$$h_t = o_t * tanh c_t \tag{8}$$

where $\sigma$ is sigmoid activation function. $i_t$, $f_t$, $\widetilde{c}_t$, $o_t$ are the output value of the input gate, forget gate, cell state and output gate. $h_{t-1}$ is the hidden state of former time step $t-1$, $x_t$ is the input value at present time $t$, $c_t$ is cell output, $h_t$ is the output at time $t$. $w_i$, $w_f$, $w_c$, $w_o$ are the weight matrices, $b_i$, $b_f$, $b_c$, $b_o$ is bias vectors. The weights and bias are updated based on the difference between the Back Propagation Trough Time (BPTT) algorithm calculated output and the actual value.

## C. MULTI-HEAD CNN-LSTM HYBRID NEURAL NETWORK

Hybrid model has better performance than single model [7], [34]. CNN and LSTM are the most widely used models for deep learning. In this paper, a novel hybrid MCNN-LSTM deep learning load forecasting model is proposed. The upper layer of MCNN-LSTM consists of three convolution heads with different positions. Convolutional neural network can receive various variables, such as historical load, date, time and weather that affect load changes. Figure 3 demonstrates the architecture of MCNN-LSTM. CNN layer extracts the

characteristics of historical load, time and weather factors respectively across three different positions of one-dimensional CNN. Using an independent one-dimensional CNN, which can obtain the characteristics of independent time series, and accurately identify the key features of the specific period (such as weekends and working day information) and weather changes that affect the load. The lower layer of MCNN-LSTM is LSTM. The time feature information related to power demand is extracted by CNN layer, and then processed by LSTM layer. LSTM is responsible for finding hidden time patterns from the extracted features. LSTM has the ability to remember the past events. Data passes through the unit neuron, which has two inputs: past historical data and current data. It can form internal memory that capture the actions through the whole sequence.
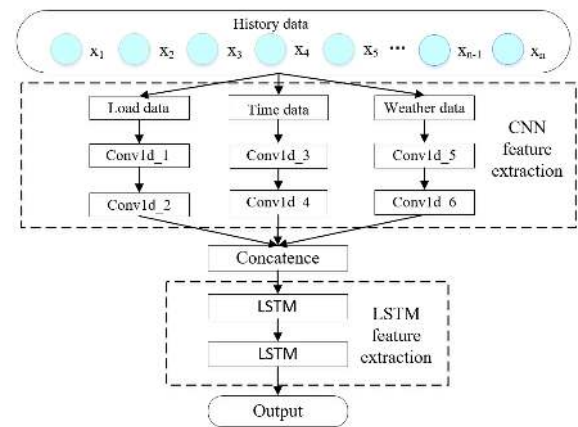


FIGURE 3. The proposed MCNN-LSTM hybrid model framework.

## D. ARCHITECTURE OF THE PROPOSED MODEL

The proposed MCNN-LSTM hybrid model can adjust the parameters of each layer according to the composition of the network. The convolution layer can adjust the parameters of filter, kernel and stride, and the LSTM layer can adjust the parameters of hidden layer. Tuning the parameters will adjust the model's learning rate and learning effect of the model. Each model will differ. Dropout works to avoid over-fitting on the results. To improve continuity, the parameters of CNN layer and LSTM layer of MCNN-LSTM as shown in Table 1.

## E. SIMULATION SOFTWARE AND HARDWARE

The proposed model is implemented in Python using back-end libraries such as Keras and TensorFlow. This work is being conducted on a personal computer equipped with an Intel Core i7-4720HQ processor, a NVIDIA GeForce GTX 950m graphics card, and 16.0 GB of random access memory (RAM).

## IV. PREDICTION EVALUATION

In order to evaluate the forecasting performance of the model proposed in this paper, the error metrics of five different

**TABLE 1.** The specific structure configuration of the proposed model

| The proposed MCNN-LSTM model | | |
|---|---|---|
| CNN layers | Conv1d_1 | filters=32; kernel size=2; stride=1 |
| | MaxPooling | kernel size=2; stride=1 |
| | conv1d_2 | filters=64; kernel size=3; stride=1 |
| | MaxPooling | kernel size=2; stride=1 |
| | conv1d_3 | filters=32; kernel size=2; stride=1 |
| | MaxPooling | kernel size=2; stride=1 |
| | conv1d_4 | filters=64; kernel size=2; stride=1 |
| | MaxPooling | kernel size=3; stride=1 |
| | conv1d_5 | filters=32; kernel size=2; stride=1 |
| | MaxPooling | kernel size=2; stride=1 |
| | conv1d_6 | filters=64; kernel size=3; stride=1 |
| | MaxPooling | kernel size=2; stride=1 |
| LSTM layers | LSTM1 | hidden nodes=128 |
| | LSTM2 | hidden nodes=128 |
| Dropout layer | dropout | 0.01 |

viewpoints are adopted. They are MAE(Mean Absolute Error), RMSE(Root Mean Square Error), MAPE(Mean Absolute Percentage Error), R-squared and error as fellow:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (9)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (10)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \qquad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{\sum_{i=1}^{N}(y_i - \overline{y}_i)} \qquad (12)$$

$$error = y_i - \hat{y}_i \qquad (13)$$

where $y_i$ is the true load data, $\hat{y}_i$ is the forecasting value, $\overline{y}_i$ is the average of true load data, and $N$ is the number of $y_i$.

## V. CASE STUDY

### A. DATASET DESCRIPTION

Ireland's 2016-2018 datasets and Maine's 2013-2015 datasets were used for the case studies. Ireland's electrical power data is provided by the EirGrid Group[1], and the country's weather data is provided by the Irish Meteorological Service[2]. Maine's electrical power and weather data are provided by ISO New England[3]. The data above is granular to the hour. The datasets from these two regions are divided into three components in this paper: training set, verification set, and testing set. The proportions of these three components are 6:2:2.

## B. DATA PREPROCESSING AND FEATURE ENGINEERING

Outlier analysis and missing value filling are examples of data preprocessing. Due to the continuity of load data, the abnormal load value included in the data calculation and analysis process will have a negative impact on the load forecast result. A box plot is a graphical representation of digital data by its quartiles, it is a straightforward and effective method for visualizing outliers. Using the upper and lower tentacles as the distribution's boundaries, any data points that are higher or lower than the upper tentacles can be considered outliers. In terms of missing value filling, our data has sparse missing values, which we fill using interpolation.

The purpose of this study is to forecast future load at the large region level using historical load data, meteorological data, and calendar variables. The primary exogenous factors affecting STLF problems are weather conditions. Seasonal weather has an effect on electricity consumption. For example, the increased use of air conditioning and heating systems during the summer and winter seasons will affect the load curve's trend. We determine the correlation between load and four weather variables (temperature, humidity, air pressure, and wind speed) using the Pearson correlation coefficient method, and identify temperature and humidity as the most relevant input features.

The load on the large level region accumulates the power system's large-scale energy consumption. In comparison to the stochastic microgrid, the periodicity of power load changes is more apparent [4]. STLF benefits from an appropriate mechanism for encoding calendar information in a way that accurately indicates the periodic change pattern. The most common methods of encoding a calendar in the traditional sense are natural encoding and one-hot encoding [11]. However, because these two coding methods do not take into account the periodicity of electricity behavior, they are ineffective for load forecasting. For instance,if natural coding is used, for example, 23:00 on the first day and 0:00 on the second day are adjacent, but their distance differs by 23. When one-hot encoding is used, the Euclidean distance between two distinct time points is always $\sqrt{2}$ and the distance between two identical time points is 0, making it difficult to distinguish the distance between the two time points.We use polar coordinates to encode calendar effect features in order to reflect the periodicity of temporal data. The polar coordinate coding method makes use of the sin and cos functions to identify calendar variables within a given period. Electricity consumption is influenced by people's production and lifestyle choices and exhibits varying trends over time. For instance, the load on weekdays and weekends is different, as is the load at various times of the day. As a result, we perform periodic coding for the hour of the day and the day of the week or year. The encoding process is by(14-19) and Figure 4. gives an example of coding 24 hours of the day.

$$p^{hsin} = sin(2\pi n^h/T^h) \qquad (14)$$

$$p^{hcos} = cos(2\pi n^h / T^h) \tag{15}$$

$$p^{wsin} = sin(2\pi n^w / T^w) \tag{16}$$

$$p^{wcos} = cos(2\pi n^w / T^w) \tag{17}$$

$$p^{dsin} = sin(2\pi n^d / T^d) \tag{18}$$

$$p^{dcos} = cos(2\pi n^d / T^d) \tag{19}$$

h

**TABLE 2.** Predictive evaluation of Ireland dataset with different input set lengths.

| Input length | MAE(MW) | RMSE(MW) | MAPE(%) | train time(s) |
|---|---|---|---|---|
| 24-step | 48.02 | 71.36 | 1.49 | 275.81 |
| 48-step | 47.29 | 68.07 | 1.48 | 540.52 |
| 72-step | 45.32 | 66.71 | 1.42 | 692.05 |
| 144-step | 56.34 | 81.28 | 1.74 | 931.37 |
| 168-step | 57.98 | 82.81 | 1.77 | 1132.77 |

72-, 144-, and 168-step series lengths.The proposed model input set is historical load series and corresponding length of historical temperature (T), humidity (H) and calendar variables($p^{hsin}, p^{hcos}, p^{wsin}, p^{wcos}, p^{dsin}, p^{dcos}$).
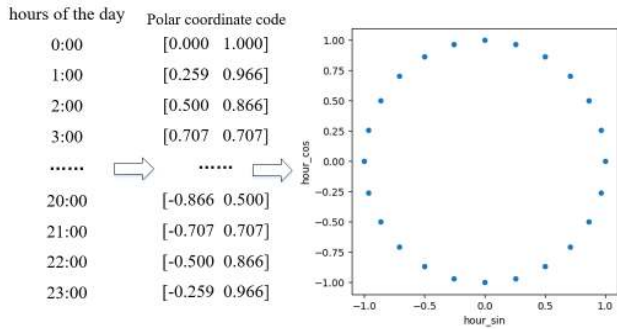


**FIGURE 4.** Polar coordinate coding for 24 hours a day.

where $n$ is time point, $T$ is the length of a period. $[p^{hsin}, p^{hcos}]$, $[p^{wsin}, p^{wcos}]$ and $[p^{dsin}, p^{dcos}]$ denote the polar coordinate code of the hour in one day and day in one week or year.

### C. SINGLE-STEP PREDICTION ANALYSIS

To evaluate the proposed MCNN-LSTM hybrid model, it is compared to traditional machine learning and deep learning models: (1) multi-layer perceptron (MLP), (2) support vector machine (SVM), (3) light gradient boosting machine (Light-GBM), (4) convolutional neural network (CNN), (5) long short-term memory neural network (LSTM), and (6) CNN-LSTM hybrid network. For each model, we perform a single-step load forecasting to forecast the load for the next hour.

#### 1) Ireland single-step load forecasting

The length of the proposed model series is crucial since it determines the amount of historical data that will be used in the training process. If the series is too brief, it is impossible to extract sufficient data to forecast the load. If the length is excessive, the presence of a significant amount of weakly correlated data increases the training cost and decreases the accuracy. We determine the length with the strongest connection by computing the load series' autocorrelation and partial autocorrelation coefficients. The autocorrelation coefficients of load series denote the correlation between a given load point and the load point at time t lagging, whereas the partial autocorrelation coefficients denote the correlation between the load point and the load series trailing over a specified time period. As illustrated in Figure 5, we use Python's statsmodels tool to plot the autocorrelation and partial autocorrelation of the Ireland load series. According to Figure 5, we design five models with 24-, 48-,
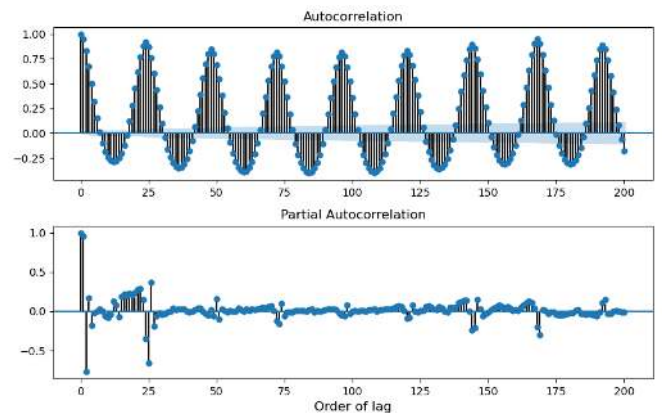


**FIGURE 5.** The autocorrelation coefficients and partial autocorrelation coefficients of load on Ireland dataset.

Table 2 compares the performance of MCNN-LSTM models with varying input set lengths.As the length of the input set rises, the training time increases as well, but the prediction accuracy does not. When the first 72 steps were used on the input set of the Ireland dataset, the prediction accuracy was the highest.

For Ireland's testing set, the results of all models' single-step load forecasting are calculated using the performance indicators MAE, RMSE, and MAPE. The evaluation indicators for the seven algorithms in Table 3 indicate that: (i) Deep learning models outperform traditional machine learning models in terms of prediction accuracy, while hybrid models outperform single models. (ii) Our proposed hybrid model's MAE, REMSE, and MAPE are the smallest and have the best predictive performance. (iii) When compared to the worst-precision SVM model, the proposed model's MAE is decreased by 42.83, RMSE is decreased by 42.58 MAPE is decreased by 1.35% and its R-squared is increased by 2.12%. (iv) Compared with the CNN-LSTM model, the proposed model's MAE decreases by 8.32, RMSE decreases 5.07, MAPE is decreased by 0.28%, and its R-squared is increased by 0.7%. Figure 6 illustrates error distributions using box plots. It can be seen that the proposed model's error

is concentrated near the zero-valued baseline, indicating that the predicted load value is closer to the actual value.

**TABLE 3.** Evaluation index of single step forecasting result of Ireland dataset
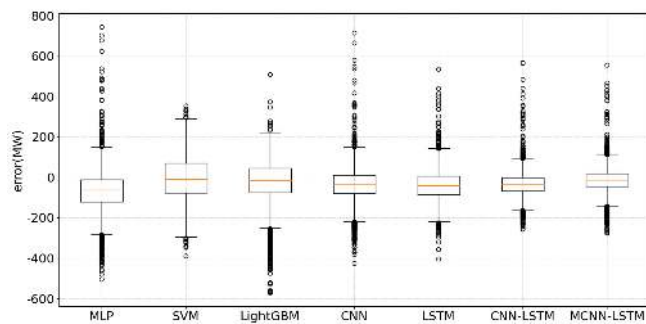
| Methods | MAE (MW) | RMSE (MW) | MAPE (%) | $R^2$ score |
|---|---|---|---|---|
| SVM | 88.15 | 109.29 | 2.77 | 96.61% |
| MLP | 82.78 | 110.11 | 2.56 | 96.56% |
| LightGBM | 75.77 | 102.13 | 2.34 | 97.04% |
| CNN | 67.612 | 94.46 | 2.08 | 97.46% |
| LSTM | 65.22 | 84.07 | 2.21 | 97.99% |
| CNN-LSTM | 53.34 | 71.78 | 1.70 | 98.03% |
| MCNN-LSTM | 45.32 | 66.71 | 1.42 | 98.73% |
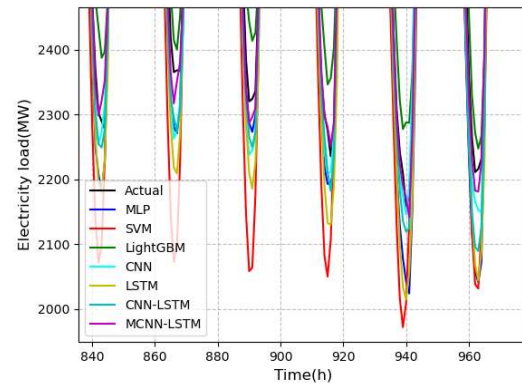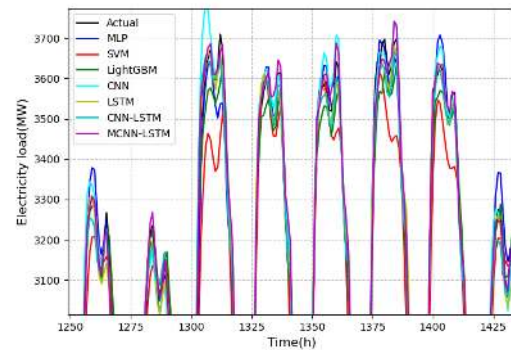


**FIGURE 6.** The error distribution of each actual value and the predicted value for Ireland dataset.
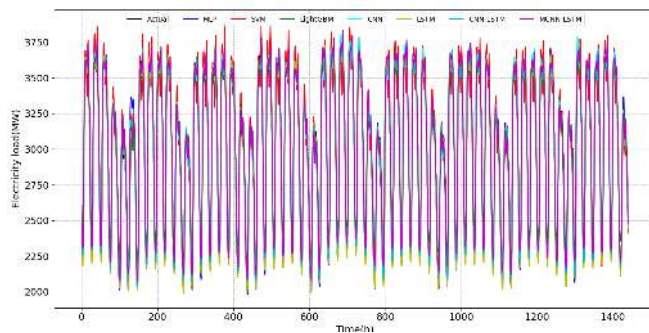


**FIGURE 7.** Single step load forecasting results of model for Ireland dataset.

Figure 7 shows the simulation results of load forecasting curves of different models. All models predict values in a consistent manner relative to the actual values. To illustrate the predictive performance of various models, the details of the lower and peak load in Figure 7 have been partially enlarged and are shown in Figure 8 (a) and (b). In comparison to other models, the proposed model's lower and upper peak loads are more similar to the actual value in terms of trend shape and degree of fit.

### 2) Maine single-step load forecasting
Figure 9 depicts the autocorrelation coefficients and partial autocorrelation coefficients of Maine datasets. The prediction



(a) Lower peak points



(b) Upper peak points

**FIGURE 8.** The local enlargement (lower peaks and upper peaks) comparison of forecasting results for Ireland dataset.

evaluation of different length load series in the Maine dataset is shown in Table 4, and the findings indicate that the proposed model performs best when a 48-step input set length is used. As a result of the variances in load curves, load data sets from different locations must select appropriate sequences to get the highest prediction performance.
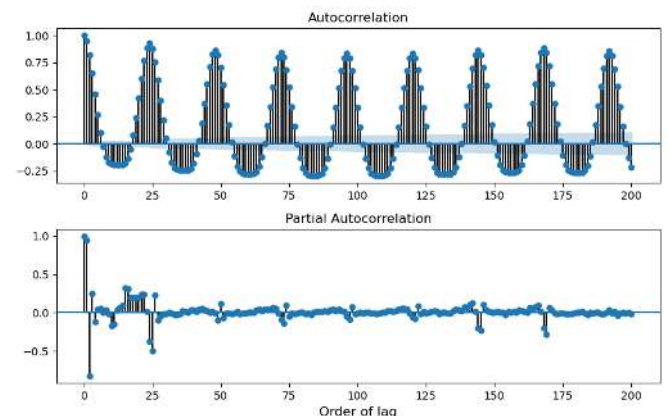


**FIGURE 9.** The autocorrelation coefficients and partial autocorrelation coefficients of load on Maine dataset.

Additionally, we forecast single-step load on Maine's testing set. Table 5 illustrates the evaluation indicators for

**IEEE** *Access*

H. H. Goh *et al.*: Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting

**TABLE 4.** Predictive evaluation of Maine dataset with different input set lengths.

| Input length | MAE(MW) | RMSE(MW) | MAPE(%) | Train time(s) |
|---|---|---|---|---|
| 24-step | 21.78 | 28.79 | 1.63 | 211.48 |
| 48-step | 18.92 | 25.04 | 1.43 | 462.81 |
| 72-step | 19.95 | 27.01 | 1.50 | 535.62 |
| 144-step | 22.93 | 30.15 | 1.73 | 766.31 |
| 168-step | 25.77 | 34.38 | 1.96 | 959.72 |

**TABLE 5.** Evaluation index of single step forecasting result of Maine dataset

| Methods | MAE(MW) | RMSE(MW) | MAPE(%) | $R^2$ score |
|---|---|---|---|---|
| SVM | 30.45 | 37.57 | 2.33 | 96.72% |
| MLP | 36.66 | 46.77 | 2.84 | 94.92% |
| LightGBM | 25.36 | 31.77 | 1.97 | 97.65% |
| CNN | 26.93 | 35.80 | 2.05 | 97.02% |
| LSTM | 24.71 | 32.53 | 1.86 | 97.54% |
| CNN-LSTM | 22.47 | 30.48 | 1.70 | 97.83% |
| MCNN-LSTM | 19.01 | 24.98 | 1.43 | 98.55% |

various models: (i)The proposed model performs the best in terms of prediction accuracy, while the MLP model performs the worst. (ii) When compared to the MLP model, the proposed model's the proposed model's R-squared increases 3.63%, MAE decreases 17.65, RMSE decreases 21.79 and MAPE has a reduction of 1.41%. (iii)When compared to the second-performing CNN-LSTM, R-squared increases 0.72%, MAE decreases 3.46, RMSE decreases 5.50 and MAPE is reduced by 0.27%. Additionally, Figure 10 box plot also demonstrates that the error distribution of the proposed model is the smallest overall.

Figure 10 illustrates the load forecasting simulation curves for various models. Figure 12 (a) and (b) are enlarged versions of the lower and peak load curves from Figure 11. The findings indicate that the proposed model is capable of accurately forecasting the electricity load and has superior fitting performance for lower and upper peak loads with a higher degree of randomness.
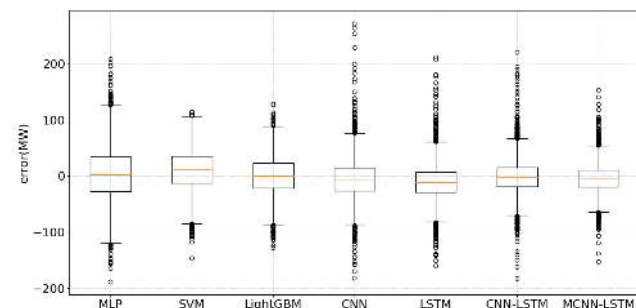


**FIGURE 10.** The error distribution of each actual value and the predicted value for Maine dataset.
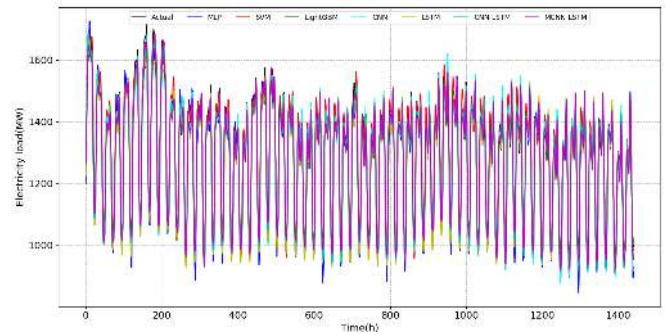


**FIGURE 11.** Single step load forecasting results of model for Maine dataset.
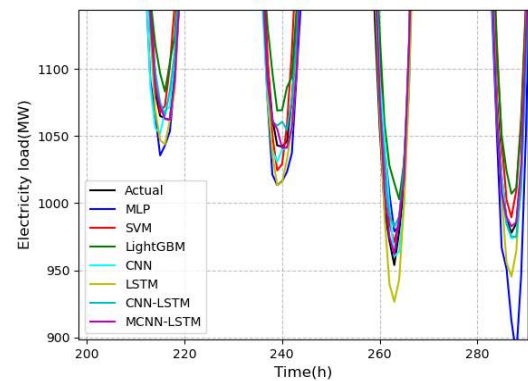
(a) Lower peak points
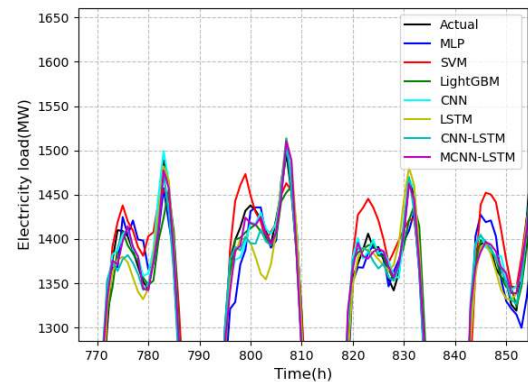
(b) Upper peak points

**FIGURE 12.** The local enlargement (lower peaks and upper peaks) comparison of forecasting results for Maine dataset.

### D. MULTI-STEP PREDICTION ANALYSIS

We compare the proposed MCNN-LSTM model to other models for 24-step load forecasting. For example, a sliding window multi-input multi-output forecasting method for MLP, CNN, LSTM, CNN-LSTM, and the proposed MCNN-LSTM models. The previous time-steps of data are used as input, and the predicted load for 24 time-steps is the output result. Because the SVM and LightGBM models do not support multiple outputs, multi-step load forecasting is accomplished by creating 24 models corresponding to 24 time steps.

**TABLE 6.** Evaluation index of 24-step forecasting result of Ireland dataset

| Methods | MAE(MW) | RMSE(MW) | MAPE(%) | $R^2$score |
|---------|---------|----------|---------|-----------|
| SVM | 122.50 | 157.56 | 3.87 | 92.96% |
| MLP | 106.68 | 146.08 | 3.20 | 93.65% |
| LightGBM | 98.96 | 135.49 | 3.01 | 94.79% |
| CNN | 86.58 | 119.48 | 2.66 | 95.95% |
| LSTM | 86.19 | 121.35 | 2.72 | 95.82% |
| CNN-LSTM | 84.27 | 119.59 | 2.65 | 95.94% |
| MCNN-LSTM | 67.31 | 105.12 | 2.10 | 96.87% |

1) Ireland 24-step load forecasting

We conduct 24-step load forecasting on Ireland's testing set, forecast the load for the next 24 hours. Table 6 summarizes the error evaluation metrics for the 24-hour ahead, including MAE, RMSE, and MAPE. In comparison to single-step load forecasting, the forecast errors of each model increase as the output dimensions increase. However, the proposed MCNN-LSTM model still has the best predictive performance, MAE is 67.31, RMSE is 105.12, MAPE is 2.10% and R-squared is 96.87%. The performance improvement of the CNN-LSTM hybrid model over the single CNN and single LSTM models is not immediately apparent. Compared with the CNN-LSTM model, the proposed MCNN-LSTM model's MAPE has a decline of 0.55%, and its MAE decreases by 16.96, RMSE decreases by 14.47, R-squared increases by 0.93%.

To concretely demonstrate the forecast results, we evaluate different time points of each week. It can be observed from table 7 that the proposed model is smaller than other models in MAE, RMSE and MAPE, indicating that the proposed model has strong load forecasting capacity for weekdays and weekends. The simulation of a 24-hour ahead prediction curve for the Ireland testing set is shown in Figure 13. The findings indicate that deep learning-based methods such as CNN, LSTM, CNN-LSTM, and the proposed MCNN-LSTM are capable of accurately simulating the next 24-hour power consumption trend. The power consumption trends of MLP, SVM, and LightGBM, on the other hand, are poorly fitted. Additionally, the proposed MCNN-LSTM model can more accurately predict the 24-hour power consumption prediction's lower and upper peak loads.

2) Maine 24-step load forecasting

The weekly MAE, RMSE and MAPE show in Table 8, the proposed model has obvious advantages over other models. The overall evaluation index of Maine test set is shown in Table 9, and the prediction curves are shown in Figure 14. The results demonstrate that our proposed model provides the best 24-hour load forecasting performance and validates its generalizability.

*E. COMPARISON OF TRAINING TIME*

We compare the training times of four deep learning models with the best prediction performance, including CNN, LSTM, CNN-LSTM, and the proposed MCNN-LSTM. The

CNN model is composed of two stacked one-dimensional convolutions with a total of 32 or 64 convolution kernels. The LSTM model is composed of two LSTMs stacked into a two-layer LSTM, with each hidden layer having a neural unit of 128. The CNN-LSTM hybrid model is constructed by stacking two-layer CNNs and two-layer LSTMs in series. The CNN layer contains between 32 and 64 convolution kernels, while the LSTM layer contains 128 hidden layer neurons. The MCNN-LSTM model's structural parameters are described in Section III D. To compensate for randomness, we run each model ten times over 100 epochs and employ early stopping to avoid overfitting. Table 10 summarizes the average ten-time training time for various deep learning models.

According to the results in Table 10, the CNN model requires the least amount of training time, while the LSTM model requires the most. The CNN-LSTM and proposed MCNN-LSTM models have a shorter training time than the LSTM model, owing to the use of the CNN layer to extract features and the convolution operation, which reduces the number of parameters and improves training efficiency. The training time for MCNN-LSTM is less than that for CNN-LSTM, because multiple convolution heads extract features independently, avoiding the issue of slow fitting speed caused by increased dimensionality of the input features. Simultaneously, the Ireland test set has 5256 data points. Single-step load forecasting has a test time of 12.08s and an average hourly forecasting time of 0.0023s. The forecasting time for a 24-step load is 10.23s, and the average forecasting time every 24 hours is 0.047s.

*F. THE IMPACT OF PERIODIC CODING*

To demonstrate the effect of periodic coding, we compare the proposed MCNN-LSTM model to a model that does not include a convolution head for extracting calendar variables. The forecasting of single-step, six-step, twelve-step, and twenty-four-step loads is carried out using Ireland's testing set, and the error evaluation indicators are listed in Table 11.

As shown in Table 11, as the output dimension increases, the MAE, RMSE, and MAPE values increase as well. The proposed MCNN-LSTM model performs better than the MCNN-SLTM model without calendar variables. According to MAPE values, the performance of single-step load forecasting is improved by 1.39%, 6-step load forecasting is improved by 6.47%, 12-step load forecasting is improved by 10.23%, and 24-step load forecasting is improved by 20.45%. The experimental results indicate that periodic coding of calendar variables is more advantageous for forecasting loads with large time steps.

**VI. CONCLUSION**

We propose a multi-head CNN-LSTM (MCNN-SLTM) short-term load forecasting model in this article. Two datasets (the Ireland dataset and the Maine dataset) are used to evaluate the proposed MCNN-LSTM model's performance. On these two datasets, experiments with single-step and 24-step
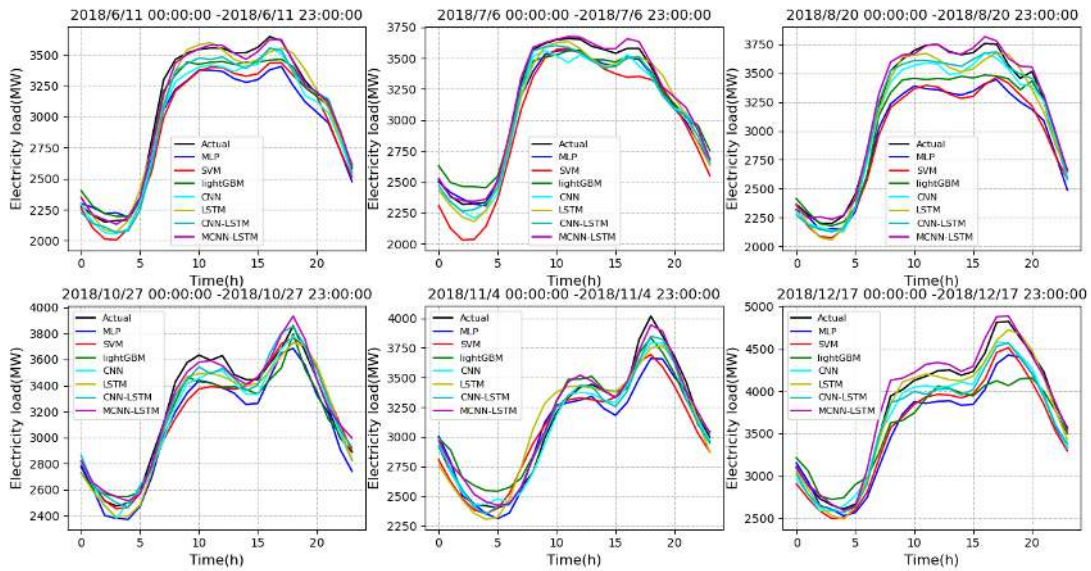
**IEEE** *Access*·

H. H. Goh *et al.*: Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting



**FIGURE 13.** 24-step load forecasting results with difference methods for ireland dataset.

**TABLE 7.** The forecast results of weekly time points for Ireland dataset.

| Time points | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| MLP | | | | | | | |
| MAE(MW) | 185.7 | 128.18 | 83.82 | 73.18 | 80.1 | 88.67 | 110.37 |
| RMSE(MW) | 222.31 | 181.9 | 120.29 | 96.53 | 107.29 | 117.47 | 138.03 |
| MAPE(%) | 5.41 | 3.68 | 2.50 | 2.20 | 2.31 | 2.81 | 3.61 |
| SVM | | | | | | | |
| MAE(MW) | 194.83 | 125.03 | 113.60 | 104.84 | 106.60 | 100.25 | 114.71 |
| RMSE(MW) | 226.29 | 177.65 | 150.37 | 129.43 | 130.99 | 123.62 | 142.39 |
| MAPE(%) | 5.84 | 3.83 | 3.60 | 3.35 | 3.31 | 3.28 | 3.95 |
| LightGBM | | | | | | | |
| MAE(MW) | 147.79 | 117.12 | 95.72 | 89.69 | 90.58 | 82.38 | 71.58 |
| RMSE(MW) | 192.37 | 169.02 | 132.82 | 115.7 | 118.62 | 103.42 | 89.78 |
| MAPE(%) | 4.42 | 3.34 | 2.85 | 2.60 | 2.64 | 2.76 | 2.51 |
| CNN | | | | | | | |
| MAE(MW) | 120.55 | 102.75 | 90.60 | 79.99 | 81.52 | 67.26 | 65.03 |
| RMSE(MW) | 169.56 | 151.57 | 117.62 | 99.32 | 104.9 | 86.18 | 82.59 |
| MAPE(%) | 3.68 | 3.09 | 2.68 | 2.39 | 2.42 | 2.17 | 2.22 |
| LSTM | | | | | | | |
| MAE(MW) | 111.88 | 97.77 | 90.18 | 79.67 | 79.78 | 70.86 | 74.38 |
| RMSE(MW) | 173.04 | 149.68 | 116.68 | 98.85 | 103.87 | 92.3 | 93.7 |
| MAPE(%) | 3.53 | 3.05 | 2.81 | 2.44 | 2.37 | 2.30 | 2.59 |
| CNN-LSTM | | | | | | | |
| MAE(MW) | 109.42 | 90.09 | 71.99 | 66.22 | 65.91 | 66.19 | 67.98 |
| RMSE(MW) | 164.43 | 142.72 | 96.78 | 86.65 | 90.54 | 85.33 | 88.95 |
| MAPE(%) | 3.39 | 2.69 | 2.16 | 2.00 | 1.95 | 2.13 | 2.32 |
| MCNN-LSTM | | | | | | | |
| MAE(MW) | 106.99 | 72.93 | 55.63 | 55.79 | 55.35 | 65.48 | 60.48 |
| RMSE(MW) | 178.90 | 126.27 | 72.95 | 81.94 | 75.91 | 81.19 | 76.44 |
| MAPE(%) | 3.37 | 2.17 | 1.67 | 1.68 | 1.63 | 2.16 | 2.07 |

**IEEE** *Access*

H. H. Goh *et al.*:Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting
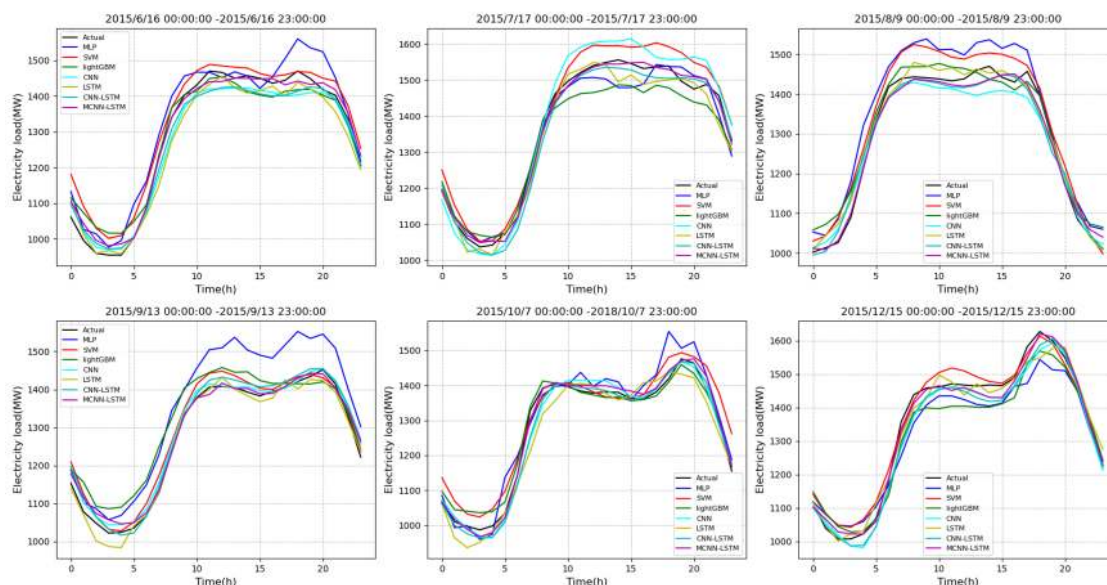


**FIGURE 14.** 24-step load forecasting results with difference methods for Maine dataset.

**TABLE 8.** The forecast results of weekly time points for Maine dataset.

| Time points | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| **MLP** | | | | | | | |
| MAE(MW) | 54.55 | 52.00 | 42.78 | 53.03 | 45.63 | 68.08 | 49.63 |
| RMSE(MW) | 70.37 | 64.49 | 54.27 | 67.53 | 57.5 | 81.15 | 63.00 |
| MAPE(%) | 3.85 | 3.88 | 3.17 | 4.05 | 3.53 | 5.35 | 3.79 |
| **SVM** | | | | | | | |
| MAE(MW) | 55.45 | 46.41 | 40.94 | 50.45 | 50.61 | 49.40 | 49.91 |
| RMSE(MW) | 70.91 | 59.52 | 52.05 | 65.01 | 66.18 | 62.47 | 61.44 |
| MAPE(%) | 4.04 | 3.38 | 3.07 | 3.86 | 3.9 | 3.95 | 3.89 |
| **LightGBM** | | | | | | | |
| MAE(MW) | 55.17 | 39.67 | 43.98 | 47.81 | 47.06 | 45.89 | 43.65 |
| RMSE(MW) | 70.68 | 55.97 | 57.23 | 61.88 | 61.64 | 56.87 | 55.26 |
| MAPE(%) | 3.96 | 2.84 | 3.16 | 3.67 | 3.68 | 3.77 | 3.40 |
| **CNN** | | | | | | | |
| MAE(MW) | 47.35 | 45.01 | 46.51 | 49.43 | 64.05 | 49.14 | 37.31 |
| RMSE(MW) | 64.59 | 62.23 | 61.14 | 67.04 | 82.4 | 65.58 | 2.88 |
| MAPE(%) | 3.42 | 3.13 | 3.31 | 3.72 | 4.82 | 3.85 | 49.64 |
| **LSTM** | | | | | | | |
| MAE(MW) | 52.01 | 43.11 | 41.97 | 46.40 | 47.39 | 39.83 | 44.04 |
| RMSE(MW) | 67.50 | 55.19 | 53.53 | 61.98 | 64.01 | 50.90 | 3.42 |
| MAPE(%) | 3.74 | 3.11 | 3.06 | 3.52 | 3.63 | 3.17 | 55.81 |
| **CNN-LSTM** | | | | | | | |
| MAE(MW) | 48.32 | 45.43 | 38.86 | 45.94 | 46.32 | 42.92 | 44.70 |
| RMSE(MW) | 65.18 | 60.48 | 49.05 | 62.58 | 65.44 | 56.05 | 57.63 |
| MAPE(%) | 3.44 | 3.21 | 2.85 | 3.50 | 3.62 | 3.41 | 3.47 |
| **MCNN-LSTM** | | | | | | | |
| MAE(MW) | 42.29 | 36.58 | 32.68 | 42.65 | 47.15 | 38.97 | 37.03 |
| RMSE(MW) | 57.07 | 53.07 | 42.85 | 56.95 | 60.17 | 50.39 | 48.40 |
| MAPE(%) | 3.03 | 2.56 | 2.34 | 3.20 | 3.36 | 3.04 | 2.82 |

**IEEE** *Access*

H. H. Goh *et al.*: Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting

**TABLE 9.** Evaluation index of 24-step forecasting result of Maine dataset

| Methods | MAE(MW) | RMSE(MW) | MAPE(%) | $R^2 score$ |
|---------|---------|----------|---------|-------------|
| SVM | 48.49 | 62.27 | 3.66 | 90.97% |
| MLP | 51.48 | 65.38 | 3.90 | 90.11% |
| LightGBM | 46.02 | 60.34 | 3.48 | 91.56% |
| CNN | 48.18 | 65.27 | 3.57 | 90.12% |
| LSTM | 45.12 | 59.67 | 3.39 | 91.74% |
| CNN-LSTM | 44.63 | 60.35 | 3.35 | 91.59% |
| MCNN-LSTM | 39.86 | 53.73 | 2.93 | 93.30% |

**TABLE 10.** The training time of different deep learning models

| Methods | Ireland dataset training time (s) | | Maine dataset training time (s) | |
|---------|-------------|---------|-------------|---------|
| | Single-step | 24-step | Single-step | 24-step |
| CNN | 149.35 | 152.25 | 121.84 | 108.72 |
| LSTM | 1293.95 | 1085.73 | 998.46 | 694.56 |
| CNN-LSTM | 975.25 | 776.60 | 585.16 | 396.97 |
| MCNN-LSTM | 692.05 | 525.97 | 462.81 | 296.40 |

load forecasting are conducted. When compared to other machine learning and deep learning models, the proposed model achieves the highest accuracy for the evaluation indicators MAE, RMSE, and MAPE. The training time of four deep learning models is also recorded in this experiment. Although the proposed model requires significantly more training time than the CNN model, it is more accurate. In comparison to the CNN-LSTM model, the proposed model can reduce parameter variables and accelerate feature extraction speed through the use of multiple convolutional heads, while maintaining high accuracy and training speed. To investigate the effect of calendar encoding, we remove the proposed MCNN-LSTM model's convolution head about calendar variables. To investigate the effect of calendar encoding, we remove the proposed MCNN-LSTM model's convolution head about calendar variables. Comparative experiments indicate that the addition of calendar variables improves load forecasting performance slightly with short output steps but significantly with long output steps.

In general, the MCNN-LSTM short-term load forecasting model proposed in this paper is capable of producing acceptable results for both single-step and multi-step load forecasting. For future research, a variety of optimization algorithms could be considered to optimize the proposed model's parameters. Holidays are not treated differently in the proposed model, and necessary holiday elements can be introduced into future model upgrades.

.

## APPENDIX A  DATA SOURCES

1. https://www.eirgridgroup.com/.
2. https://www.met.ie/.
3. https://www.iso-ne.com/.

**TABLE 11.** The evaluation index of multiple steps ahead for load forecasting in Ireland's testing set

| Steps | MCNN-LSTM | | MCNN-LSTM (lack calendar variables) | |
|-------|----------|----------|----------|----------|
| | MAE (MW) | MAPE (%) | MAE (MW) | MAPE (%) |
| Single-step | 45.32 | 1.42 | 46.89 | 1.44 |
| 6-step | 59.70 | 1.88 | 65.83 | 2.01 |
| 12-step | 66.76 | 2.04 | 72.94 | 2.27 |
| 24-step | 67.31 | 2.10 | 85.32 | 2.64 |

## REFERENCES

[1] P. Singh, P. Dwivedi, and V. Kant, "A hybrid method based on neural network and improved environmental adaptation method using controlled gaussian mutation with real parameter for short-term load forecasting," *Energy*, vol. 174, pp. 460–477, 2019.

[2] P. Mukhopadhyay, G. Mitra, S. Banerjee, and G. Mukherjee, "Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter," in *2017 7th International Conference on Power Systems (ICPS)*. IEEE, 2017, pp. 812–819.

[3] L. Friedrich and A. Afshari, "Short-term forecasting of the abu dhabi electricity load using multiple weather variables," *Energy Procedia*, vol. 75, pp. 3014–3026, 2015.

[4] X. Liu, Z. Zhang, and Z. Song, "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning," *Renewable and Sustainable Energy Reviews*, vol. 119, p. 109632, 2020.

[5] 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.

[6] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, "Hourly forecasting of solar irradiance based on ceemdan and multi-strategy cnn-lstm neural networks," *Renewable Energy*, vol. 162, pp. 1665–1683, 2020.

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

[8] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based lstm-convolutional network," *Energy*, vol. 189, p. 116225, 2019.

[9] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electric power systems research*, vol. 130, pp. 139–147, 2016.

[10] X. Sun, Z. Ouyang, and D. Yue, "Short-term load forecasting based on multivariate linear regression," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. IEEE, 2017, pp. 1–5.

[11] C.-M. Lee and C.-N. Ko, "Short-term load forecasting using lifting scheme and arima models," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5902–5911, 2011.

[12] J. Che and J. Wang, "Short-term load forecasting using a kernel-based support vector regression combination model," *Applied energy*, vol. 132, pp. 602–609, 2014.

[13] S. Karthika, V. Margaret, and K. Balaraman, "Hybrid short term load forecasting using arima-svm," in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*. IEEE, 2017, pp. 1–7.

[14] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting," *IEEE Access*, vol. 7, pp. 28 309–28 318, 2019.

[15] S. K. Sheikh and M. Unde, "Short term load forecasting using ann technique," *International Journal of Engineering Sciences & Emerging Technologies*, vol. 1, no. 2, pp. 97–107, 2012.

[16] A. K. Fard and M.-R. Akbari-Zadeh, "A hybrid method based on wavelet, ann and arima model for short-term load forecasting," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 26, no. 2, pp. 167–182, 2014.

[17] P. Singh and P. Dwivedi, "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," *Applied energy*, vol. 217, pp. 537–549, 2018.

[18] 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. 88 058–88 071, 2019.

[19] A. Kaligambe and G. Fujita, "Short-term load forecasting for commercial buildings using 1d convolutional neural networks," in *2020 IEEE PES/IAS PowerAfrica*. IEEE, 2020, pp. 1–5.

[20] S. Atef and A. B. Eltawil, "Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting," *Electric Power Systems Research*, vol. 187, p. 106489, 2020.

[21] K. Yan, W. Li, Z. Ji, M. Qi, and Y. Du, "A hybrid lstm neural network for energy consumption forecasting of individual households," *Ieee Access*, vol. 7, pp. 157 633–157 642, 2019.

[22] X. Dong, L. Qian, and L. Huang, "Short-term load forecasting in smart grid: A combined cnn and k-means clustering approach," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2017, pp. 119–125.

[23] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019.

[24] M. A. Rahman, A. Zubair *et al.*, "Electric load forecasting with hourly precision using long short-term memory networks," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, 2019, pp. 1–6.

[25] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Applied energy*, vol. 238, pp. 1312–1326, 2019.

[26] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid cnn-lstm model for short-term individual household load forecasting," *IEEE Access*, vol. 8, pp. 180 544–180 557, 2020.

[27] J. Kim, J. Moon, E. Hwang, and P. Kang, "Recurrent inception convolution neural network for multi short-term load forecasting," *Energy and buildings*, vol. 194, pp. 328–341, 2019.

[28] 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. 31 191–31 212, 2021.

[29] T. Li, M. Hua, and X. Wu, "A hybrid cnn-lstm model for forecasting particulate matter (pm2. 5)," *IEEE Access*, vol. 8, pp. 26 933–26 940, 2020.

[30] S. Wang and H. Chen, "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network," *Applied energy*, vol. 235, pp. 1126–1140, 2019.

[31] Y.-Y. Hong and C. L. P. P. Rioflorido, "A hybrid deep learning-based neural network for 24-h ahead wind power forecasting," *Applied Energy*, vol. 250, pp. 530–539, 2019.

[32] X. Zhao, H. Wei, H. Wang, T. Zhu, and K. Zhang, "3d-cnn-based feature extraction of ground-based cloud images for direct normal irradiance prediction," *Solar Energy*, vol. 181, pp. 510–518, 2019.

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

[34] X. Shao, C. Pu, Y. Zhang, and C. S. Kim, "Domain fusion cnn-lstm for short-term power consumption forecasting," *IEEE Access*, vol. 8, pp. 188 352–188 362, 2020.

• • •