

Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data

Jaewon Chung, Beakcheol Jang

Published: November 23, 2022 • <https://doi.org/10.1371/journal.pone.0278071>

Abstract

The stress placed on global power supply systems by the growing demand for electricity has been steadily increasing in recent years. Thus, accurate forecasting of energy demand and consumption is essential to maintain the lifestyle and economic standards of nations sustainably. However, multiple factors, including climate change, affect the energy demands of local, national, and global power grids. Therefore, effective analysis of multivariable data is required for the accurate estimation of energy demand and consumption. In this context, some studies have suggested that LSTM and CNN models can be used to model electricity demand accurately. However, existing works have utilized training based on either electricity loads and weather observations or national metrics e.g., gross domestic product, imports, and exports. This binary segregation has degraded forecasting performance. To resolve this shortcoming, we propose a CNN-LSTM model based on a multivariable augmentation approach. Based on previous studies, we adopt 1D convolution and pooling to extract undiscovered features from temporal sequences. LSTM outperforms RNN on vanishing gradient problems while retaining its benefits regarding time-series variables. The proposed model exhibits near-perfect forecasting of electricity consumption, outperforming existing models. Further, state-level analysis and training are performed, demonstrating the utility of the proposed methodology in forecasting regional energy consumption. The proposed model outperforms other models in most areas.

Citation: Chung J, Jang B (2022) Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data. PLoS ONE 17(11): e0278071. <https://doi.org/10.1371/journal.pone.0278071>

Editor: Yogendra Arya, J.C. Bose University of Science and Technology, YMCA, INDIA, INDIA

Received: June 13, 2022; **Accepted:** November 8, 2022; **Published:** November 23, 2022

Copyright: © 2022 Chung, Jang. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The data underlying this study are available from GitHub at the following link: <https://github.com/Jwmc999/HybridEnergyForecastingModel>.

Funding: Funder: National Research Foundation of Korea Award number: NRF-2022R1F1A1063961 Grant Recipient: Beakcheol Jang The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Literature review

Electricity is both a primary index representing the economic welfare of a country and the primary gear of modern convenience [1, 2]. Power outages threaten the quality of life and the economy of a state [3] and power shortages damage the economic growth of emerging nations [3, 4]. Even nations with well-developed power systems may experience sudden power failures [3]. Climate change and adverse weather conditions such as heatwaves, tornadoes, and heavy snowfall increase the chances of electrical failure. Therefore, the accurate forecasting of electricity consumption has become more important to ensure a stable electricity supply.

Several studies have been conducted to predict energy consumption trends. Long short-term memory (LSTM) models combined with convolutional neural networks (CNNs) have been proposed for the hourly and daily forecast of energy demands, which outperform multilayer perceptron (MLP) and recurrent neural network (RNN) models [5–7]. Electricity load forecasting (LF) studies [5, 7, 8] have utilized historical electricity load and climate data uploaded in real time. Most energy consumption forecasting studies [9–12] have been conducted considering factors such as population, import/export values, and gross domestic product (GDP). The nomenclature utilized in this study is presented in [Table 1](#).

Nomenclature	
CNN	Convolutional Neural Network
DL	Deep Learning
EIC	Electrical Energy Consumption
GDP	Gross Domestic Product
LF	Load Forecasting
LSTM	Long-Short Term Memory
MAPE	Mean Absolute Percentage Error
MLP	Multilayer Perceptron
MTLF	Mid-term Load Forecasting
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ReLU	Rectified Linear Units
STLF	Short-term Load Forecasting
TEC	Total Energy Consumption

<https://doi.org/10.1371/journal.pone.0278071.t001>

Table 1. The nomenclature.
<https://doi.org/10.1371/journal.pone.0278071.t001>

Most existing studies on electrical energy consumption (EEC) and total energy consumption (TEC) forecasting methods focus on macroeconomic features, including population, GDP, and import/export values. On the other hand, existing studies on short-term load forecasting (STLF) and mid-term load forecasting (MTLF) consider only historical load and weather/time information [13]. However, this binary separation of components constrains the forecasting accuracy of energy demand and consumption—the components must be integrated to improve prediction performance.

Table 2 presents an overview of existing load and energy consumption forecasting, including their input and target variables, forecasting time intervals, and forecasting areas. In terms of forecasting time intervals, energy forecasting can be categorized into two groups: 1) monthly, seasonal, and annual forecasting [1, 10–12, 14–16] and 2) per-minute, hourly, and daily forecasting [5–8, 17]. The former type considers socio-economic input variables from GDP, population, sales index, production index, import/export, demographic, personality, and Google Trends data. The latter type involves observations of the actual electric load, voltage, submetering, and weather. Further, the target areas of the two types are different. Methods of the first group predict energy consumption at state [9, 14, 15], national [11, 12, 14, 16], and global [10] levels whereas those in the second group [6–8, 17] predict the energy consumption of relatively smaller areas. In [9], a neural network ensemble approach based on a novel sparse adaboost framework and an echo state network was used to improve generalization ability and construct nonlinear relations between electricity demand and other factors. The proposed model was validated by using it to predict the industrial electricity consumption of Hubei Province of China. However, further real-life applications in different regions of China are required to establish its generality. In [10], Google Trends data, including the search history of certain keywords, were used as the input dataset and an online big data-driven oil consumption forecasting approach was constructed for application alongside statistical and machine learning (ML) techniques. It was the first attempt to use the trends of google search data to predict uncertain yet essential oil consumption. It also proved the predictive power of Google search trends through relationship investigation. However, the proposed prediction improvement techniques did not explore deep learning (DL) modules. In [14], the forecasting speed of residential electricity consumption was prioritized, contrary to most residential electricity consumption studies that focus on forecasting accuracy. The authors proposed a hybrid Improved Whale Optimization Algorithm—Optimized Grey Seasonal Variation Index which exhibited both high accuracy and fast convergence. Consideration of forecasting speed is practical because real-life implementation requires fast forecasting. The aforementioned model also exhibited “excellent” forecasting accuracy. In [8], the periodic part of the household residual load forecasting was modeled based on the behavioral patterns of overestimated/underestimated residual components. The model exhibited a significant improvement in the prediction accuracy of periodic residual demand. When combined with climatic data, it also improved total power consumption prediction. Nevertheless, as the real dataset of the experiment pertained to a single household, its application in broader residential households remains to be established. To improve the performance of RNN on electric load forecasting at a specific time, a recurrent inception convolution neural network combining RNN and 1D CNN was proposed in [6]. To this end, a 1D CNN inception model was used to balance prediction time and hidden state vector values. Its performance was verified based on power usage data obtained from three distribution complexes in South Korea. The model outperformed the benchmarks of MLP, RNN, and 1D CNN. However, because there are multitudinous power distribution complexes in South Korea, its application to the entire country requires further verification.

Ref.	Input Variables	Output variables	Production Interval	Prediction Area
[9]	Industrial value, production price index, export delivery value, export value, import value etc.	Industrial electricity consumption	Monthly	Hubei, China
[10]	Google trend data	Oil consumption	Monthly	World
[11]	GDP, population, export value, import value	EEC/TEC	Annual	China
[12]	Seasonal residential electricity consumption history	EEC, Residential electricity consumption	Seasonal	USA, 4 regions of China
[14]	Electricity consumption history	EEC	Monthly	Brazil, Algeria
[15]	Time electricity generation, GDP, population, installed capacity, export value, export value, total submetering	EEC	Annual	Turkey
[16]	Energy consumption history, moving average, flow, maximum, rate of change etc.	TEC	Monthly	USA
[8]	Electricity load history, climatic data	Residential electricity load	Hourly	An individual household in Montreal, Canada
[17]	Active power, reactive power, voltage, global intensity etc.	Residential electricity consumption	15 minutes	An individual residential building in Paris, France
[1]	Global active power, global reactive power, voltage, global intensity, sub-metering	Residential power consumption	Hourly	An individual household in France
[21]	Electricity load history (2003–2010), 51 years of electricity load of New England, New York	Electricity load	Hourly	New England and New York, USA
[6]	Electricity load history, climatic data, time information, electricity rates, number of houses	Electricity load	Daily	Three distribution complexes of South Korea

Table 2. Some typical types of energy forecasting techniques.
<https://doi.org/10.1371/journal.pone.0278071.t002>

Research gap and motivation

In this paper, we present a novel and accurate hybrid CNN-LSTM model that uses multivariable data to forecast EEC. Further, we explain the potential problems and roles of various state-level forecasting components, such as regional import and export values. The proposed technique utilizes broad and necessary components of state-level time-series forecasting techniques. The proposed technique uses LSTM layers to process and predict based on time-series data, and the output depends not only on the current input but also on the inputs from the past. The hybrid CNN-LSTM also utilizes convolutional and pooling layers to extract various aspects and objects in a time series, and achieves differentiation at a relatively low computational cost. Moreover, it uses a multivariable dataset containing the data of electric load, export values, and import values to estimate the electricity usage of the following month. Industrial electricity usage and imports/exports are directly proportional to each other; hence, the value of international trade represents the present and future costs of electricity closely. Our technique also allows users to analyze and predict electricity consumption of any region, in addition to that on a national level.

South Korea was deemed to be an adequate data source to identify the correlation between import/export values, industrial electricity usage, and total electricity consumption. As an advocate for climatic environmentalism, the South Korean government has recently reduced the number of nuclear and thermal power plants in the country. This endeavor has led to nationwide concerns regarding electricity shortage, as 72.9% of the GDP [18] of the country relies on importing/exporting sectors of the manufacturing industry, which require more than one-third of the nation’s total EEC [19]. Moreover, manufacturing and production firms in South Korea are primarily distributed in specific regions, necessitating accurate state-level energy forecasting for smooth operation. Therefore, 16 regions in South Korea are analyzed to train a novel nationwide DL-based EEC forecasting model in this paper.

In this study, the Open Data Portal [20], Open MET Data Portal [21], and Trade Statistics [22] government databases of South Korea are used, with each offering monthly data on energy consumption, import/ export values, and temperature and rainfall data during the period between January 2004 and December 2020. Our augmented multivariable CNN-LSTM is observed to exhibit a root mean square error (RMSE) of 0.165 in the nation-level experiment, outperforming conventional deep learning-based energy

forecasting algorithms—the RMSEs of MLP, RNN, and LSTM are 0.4521, 0.1713, and 0.174, respectively. Moreover, in regional forecasting, the CNN-LSTM model using the proposed data processing method outperforms the alternatives in all the areas, accurately predicting 16 states in terms of RMSE and mean absolute percentage error (MAPE).

Contribution

As mentioned in the Literature Review and the Research Gap and Motivation sections, previous approaches to the prediction of electricity consumption and demands have not considered the two types of input factors simultaneously. However, greater integrity in input is necessary to improve prediction accuracy. Moreover, monthly state-level predictions are required to account for variable energy consumption patterns of different states. Finally, we utilize an interpolation technique for convergence between monthly data points. While the time units of LF datasets are as detailed as seconds and milliseconds, the finest unit of time in the case of EEC/TEC is a month. Our interpolation bridges these gaps between different data points and ensures better prediction scores. The various contributions of this paper are as follows:

- We present a novel EEC forecasting method based on small datasets using data augmentation.
- The proposed model integrates input components of EEC/TEC forecasting and LF to improve prediction performance.
- To the best of our knowledge, this is the first study to conduct the state-level monthly EEC forecasting for South Korea.

Paper organization

The rest of this paper is structured as follows. In Materials and Methods, the proposed deep learning model is introduced along with the requisite background information. Experimental validation of the proposed method is presented in the Results section. In the Discussion section, the motivation of the study is re-stated and the results obtained using the proposed method are analyzed. Finally, the study is concluded in the Conclusions section.

Materials and methods

The proposed energy-consumption forecasting model comprises three steps. In the first step, the univariable and multivariable time series data based on continuous observations of climatological/economic information are analyzed and a collective dataset is constructed. In the second step, a resampling method and spline interpolation are applied to generate the simulation dataset comprising continuous polynomial segments. An overview of deep learning technology for sequential data is introduced in the third step with the objective of preserving an accurate one-day and one-month forecasting horizon for electricity consumption.

Extraction of important features

In [1, 2], a close correlation was reported between GDP and power consumption. However, the annuity of GDP impedes its application in deep learning, as annual GDP index data is too sparse for DL model training. To offset this challenge, monthly macroeconomic features, including import/export values, volume, and trade gains and losses from Korea Customs Service, are collected in this study. Fig 1 depicts the Pearson correlation coefficients of all 11 features. The import and export values exhibit the highest correlation with power usage. Climatic factors, such as temperature and precipitation, exhibit weak relationships with the target variable.

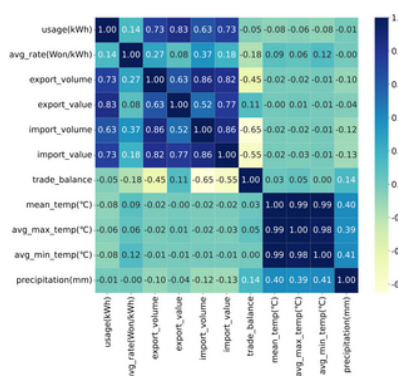


Fig 1. Correlation coefficients of input variables.

<https://doi.org/10.1371/journal.pone.0278071.g001>

Regional analysis is another important sector of energy forecasting because energy consumption is correlated to the local conditions of each community. These conditions include weather, population, and economic factors. However, although studies have been conducted on provincial energy forecasting [5, 9, 14], nationwide forecasting of each region remains to be performed.

Data augmentation

Certain input variables cannot be collected at the frequency of a target variable. Specifically, macroeconomic factors, such as import and export values, are usually collected monthly, quarterly, and annually, while empirical observations of temperature, rain, and electric load are measured every second, minute, or hour. To preserve the characteristics of both domains, a new dataset of daily frequency can be used as a middle ground.

Fig 2 illustrates the workflow of the proposed method. Once the rows of new daily values are generated, interpolation is used to generate the simulated values. In power forecasting, as mentioned in [23], piecewise cubic polynomial interpolation represents a reasonable compromise between computational cost and flexibility. In [24] as well, piecewise cubic interpolation was preferred over other spline methods to smooth the down-sampled data, suggesting a data-driven load-forecasting method. However, in the

aforementioned study, a significant performance improvement was not reported in the case of the cubic spline compare to the case of quadratic spline—instead, piecewise cubic interpolation required greater computational time during the training process than the quadratic spline. Therefore, piecewise quadratic-spline interpolation is used for data augmentation in this study.

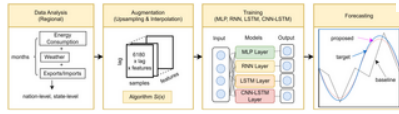


Fig 2. Visualization of the proposed methodology.
<https://doi.org/10.1371/journal.pone.0278071.g002>

The function, $S(x)$, interpolates each local data point piecewise to restrict Runge's phenomenon. Here,

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2, \quad (1)$$

where $i \in [0, 1, \dots, n]$ and $x \in R$. The parameters a_i , b_i , and c_i denote quadratic polynomial coefficients. A multi-dimensional array of size $6180 \times lag \times features$ defines a new daily polynomial coefficient structure. For each region, the length of the time series is taken to be 6180, lag denotes the length of the shifted time-frequency for single-step forecasting, and $features$ denote multivariable factors from the given datasets containing data regarding electricity consumption, weather information, and import/export indexes.

Deep learning models

1. Multilayer perceptron.

MLP was introduced to address the limitations of the perceptron. By stacking one or more hidden layers between the input and output layers, MLP maps the input and output nonlinearly. First, the input layer receives an input tensor of arbitrary shape. The weight of each neuron is calculated as the data transmitted from the input to the output layer. Then, the error is minimized by backpropagating the mini-batch gradient descent (for weight update) for every epoch. Notably, the performance of the feedforward network depends on the learning rate and the effectiveness of the optimizer of the backpropagation process. Finally, the output layer of the MLP yields the desired output, e.g., the classification or prediction result. The simplest form of MLP is based on linear regression and comprises a single input and output layer with no activation functions (i.e., a linear activation function). Eq.(2) presents the mathematical expression of MLP, where w denotes the vector of weights, x denotes the input tensor, b denotes the bias, and σ denotes the nonlinear activation function.

$$y = \sigma \left(\sum_{i=1}^n w_i x_i + b \right) = \sigma(W^T x + b) \quad (2)$$

Despite its nonlinearity, MLP experiences overfitting and vanishing gradient problems. Overfitted models fail to predict test data correctly, thereby diminishing their feasibility. In addition, in MLPs containing hundreds of hidden layers, the backpropagation process can diminish the gradient to 0, or close to 0, impeding further training.

2. Recurrent neural networks.

RNNs and RNN-based models exhibit superior performance on sequence data, such as text and time series data [25]. While feedforward neural networks remember only the current input, RNNs store information regarding the temporal order and consider previous inputs as well as the current state during decision-making. As in the case of a feedforward network, an RNN first calculates the loss function on a batch (forward propagation). Then, it updates the gradients based on the current state by calculating the state memory of earlier instances (backpropagation through time).

$$a^{(t)} = \tanh(W_{ax}x^{(t)} + W_{aa}a^{(t-1)} + b_a) \quad (3)$$

$$\hat{y}^{(t)} = \text{softmax}(W_{ya}a^{(t)} + b_y) \quad (4)$$

Eq.(3) presents a basic RNN cell, where $x^{(t)}$ denotes the current input, and $a^{(t-1)}$ denotes a previous hidden state with previous information. A single RNN unit takes $x^{(t)}$ and $a^{(t-1)}$ as inputs and outputs $a^{(t)}$ to the following RNN cell. Output $a^{(t)}$ is also used to predict $y^{(t)}$. However, RNNs are also prone to the vanishing gradient problem, as their cells have short-term memory.

3. Long-short term memory.

To counteract the vanishing gradient problem of RNNs, three gates that forget, update, and output the hidden states of current and previous instances are used in LSTM [26]. In Eq.(5) the forget gate decides whether to delete or retain stored memory values of the input state. If one of the values of $\Gamma_f^{(t)}$ is 0, or close to 0, the LSTM cell removes information from the corresponding component of $c^{(t-1)}$. If one of the values is 1, the LSTM cell retains the information. Similarly, the update gate reflects the information from the forget gate, whereas the output gate determines the outputs to be used. All these decisions are then added to the cell state, and the memory variable is subsequently delivered to the subsequent LSTM cell with a hidden state. Thus, each LSTM-cell tracks and updates a memory variable $c^{(t)}$ (i.e. a cell state) at each time step, which can be different from $a^{(t)}$.

$$\Gamma_f^{(t)} = \sigma(W_f[a^{(t-1)}, x^{(t)}] + b_f) \quad (5)$$

$$\Gamma_u^{(t)} = \sigma(W_u[a^{(t-1)}, x^{(t)}] + b_u) \quad (6)$$

$$\hat{c}^{(t)} = \tanh(W_c[a^{(t-1)}, x^{(t)}] + b_c) \quad (7)$$

$$c^{(t)} = \Gamma_f^{(t)} \circ c^{(t-1)} + \Gamma_u^{(t)} \circ \hat{c}^{(t)} \quad (8)$$

$$\Gamma_o^{(t)} = \sigma(W_o[a^{(t-1)}, x^{(t)}] + b_o) \quad (9)$$

$$a^{(t)} = \Gamma_o^{(t)} \circ \tanh(c^{(t)}) \quad (10)$$

4. Convolutional neural network and long-short-term memory.

CNN-LSTM architectures extract detailed features from multiple variables to forecast EEC and can remember complex irregular trends [Z]. The CNN-LSTM model comprises two stages—feature extraction and forecasting [27], as illustrated in Fig 3. Feature extraction comprises a 1D CNN consisting of convolution layers, an activation function, and a pooling layer. The 1D CNN extracts features from an input layer that receives multivariable macroeconomic and observational time-series sequences. Then, the CNN layer transmits the results to the LSTM layer, and the dense layer forecasts the future electricity consumption of each region.

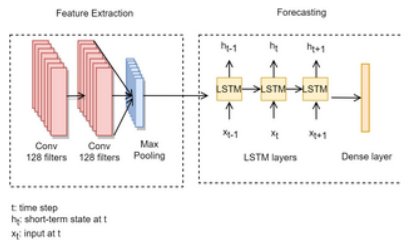


Fig 3. CNN-LSTM model architecture.

<https://doi.org/10.1371/journal.pone.0278071.g003>

In Eq (11), h_{ij}^1 denotes the output vector of the first convolution layer, where $i \in [0, 1, \dots, n]$, $x_i^0 = x_1, x_2, \dots, x_n$ denotes the input energy consumption input vector, n denotes the length of time-series input sequence, j denotes the index value of the feature map corresponding to each lag, M denotes the number of filters, and W denotes the weight vector of the kernel. h_{ij}^l denotes the result of the l th convolution layer and the i th value of the layer. To reduce the network computation costs and the number of trainable parameters, the CNN layer uses a pooling layer to reduce the spatial size of the representation. The pooling layer p_{ij}^l integrates a neuron filter on a previous layer into a scalar in the subsequent layer, where T denotes the size of the stride and R denotes the pooling size. Max-pooling is used in this paper, which selects the maximum value from each filter during feature extraction.

$$h_{ij}^1 = \sigma \left(\sum_{m=1}^M W_{m,j}^1 x_{i+m-1,j}^0 + b_j^1 \right) \quad (11)$$

$$h_{ij}^l = \sigma \left(\sum_{m=1}^M W_{m,j}^l x_{i+m-1,j}^{l-1} + b_j^l \right) \quad (12)$$

$$p_{ij}^l = \max_{r \in R} h_{i \times T + r, j}^{l-1} \quad (13)$$

The kernel of the 1D CNN moves vertically across a feature map; hence, it receives a single integer as the filter height. In contrast, a 2D CNN has a two-dimensional kernel, e.g., a 2×1 filter [Z]. A 1D CNN is selected in this study instead of a 2D CNN because our input data consists of a one-dimensional time series sequence. In addition, 1D CNN is more competitive in terms of computational cost than 2D CNN [28].

5. Evaluation metrics.

The proposed model was compared with existing ones in terms of RMSE and MAPE. RMSE calculates prediction errors and indicates how far the residuals are from the line of best fit. RMSE is useful to configure models for a specific variable and compare prediction errors between different models. RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (C_t - A_t)^2} \quad (14)$$

where n denotes the number of observations, C_t denotes the forecasted values of consumption, and A_t denotes the observed actual values at time stamp t .

MAPE is used to compare the forecasting accuracy of different time-series models. For each forecasted point in period t , the prediction error is given by $e_t = A_t - C_t$, and the absolute value of the percentage error, $e_t = (A_t - C_t)/A_t$, is summed and divided by the number of fitted points, n . MAPE corresponding to a period t is given by:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{(A_t - C_t)}{A_t} \right| \times 100 \quad (15)$$

Results

Data pre-processing

For data pre-processing, the multivariable time-series data are rescaled from monthly to daily periods. Each feature contains 204 rows of monthly values collected between January 2004 and December 2020 (204 months). 204 months are upsampled into 6180 days for 16 South Korean regions. The upsampled dates contain NaN values for all 3 features of historical electricity usage, export value, and import value. Historical electricity consumption is defined as the sum of residential, industrial, educational, and other electricity usages. Export and import values denote the total price of exported and imported merchandise, respectively. As mentioned in Extraction of Important Features, because the climate variable has little correlation with historical electricity usage, climatic data is excluded from experiments. The empty values are estimated using spline interpolation. The total length of the preprocessed data is taken to be $6180 \times 16 \times 4$, without any missing values. Following previous energy forecasting studies [6], MLP is adopted as our baseline model. The baseline method is taken to be univariate single-step forecasting based on a previous consumption time series. To confirm the effect of augmentation, multivariable forecasting is also included in this experiment. The unaugmented univariate and multivariable methods are used to predict the monthly consumption based on the data of the previous two months. The proposed multivariable augmented model uses the data of the interpolated past 30 days to predict the consumption figure for each day. To compare the daily forecasts with the monthly forecasts, we extract consumption predictions for specific dates corresponding to the time intervals before interpolation.

Model hyperparameters

To focus on the impact of data augmentation, the architecture of the DL models is simplified. The MLP consists of a single dense layer with one node and one linear activation function. Both the RNN and LSTM comprise 128 units, a ReLU (Rectified Linear Unit) activation function and a dropout layer are included following a single hidden layer. The CNN-LSTM contains 2 layers of 1D CNN with 128 filters, kernel size of 1, stride of 1, pooling size of 2, and a ReLU activation function, followed by the LSTM layer with identical hyperparameter as the LSTM-only model. All models accept a 2D array of (length of dataset, number of features) as the input and use the Adam optimizer for compilation. The final dense layer of the augmented model forecasts the power consumption for the subsequent day and those of the other predicts the consumption for the subsequent month. All models split the training and test datasets following a 70:30 ratio. The training is repeated for each model, region, and method to ensure a fair comparison.

Comparison of national performance

The performance of the proposed model is compared with those of other DL algorithms in terms of model fitness on different examples. Notably, LSTM does not always yield superior results compared to RNN [29], even though the former improves the exploding/vanishing gradient problems of RNNs. The performance of each algorithm is estimated on a basic dataset, that is univariate and without augmentation. Fig 4 compares the national electricity usage predictions obtained using deep learning models with those of the baseline monthly dataset. This figure confirms that the time-series outputs of the RNN, LSTM, and CNN-LSTM are almost identical. Table 3 summarizes the evaluation metrics used for further comparison of the models. The RMSE and MAPE scores are observed to decrease significantly with the addition of recurrent layers, and the inclusion of a 1D convolution layer is observed to improve performance additionally. Also, to further demonstrate the robustness of CNN-LSTM, we add the forecasting comparison of the USA's electricity consumption. Retrieved from [30], the USA dataset consists of the national electricity usage and the price from January 1990 to July 2022. Same with the South Korea dataset, spline interpolation is applied to the monthly USA dataset.

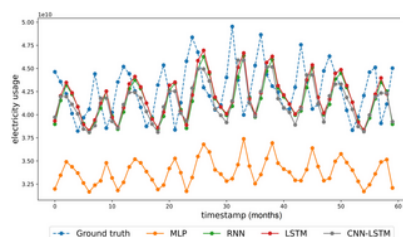


Fig 4. Comparison of national forecasting performance.
<https://doi.org/10.1371/journal.pone.0278071.g004>

South Korea	MLP	RNN	LSTM	CNN-LSTM
RMSE	0.4553	0.1766	0.175	0.169
MAPE	0.4737	0.1402	0.1389	0.1381
USA	MLP	RNN	LSTM	CNN-LSTM
RMSE	1.6907	0.3111	0.3085	0.2761
MAPE	1.8008	0.1868	0.1874	0.1641

<https://doi.org/10.1371/journal.pone.0278071.t003>

Table 3. Forecasting performance evaluation.
<https://doi.org/10.1371/journal.pone.0278071.t003>

Regional performance comparison

We verified the effect of multivariable augmentation on the performance of the CNN-LSTM model. The state-wise monthly electricity consumption prediction obtained using each method on the test dataset of 60 months is illustrated in Fig 5. The proposed method effectively minimizes the error in all regions compared to other methods, and the multivariable augmented flows overlap almost perfectly with the actual power demand data. Only Gyeongnam, Gyeongbuk, and Ulsan exhibit observable differences between the ground truth and multivariable augmentation. Table 4 summarizes the evaluation results of the CNN-LSTM model for 16 regions using different data processing methods. The proposed augmented multivariable model is observed to yield useful results in all states, exhibiting the lowest RMSE and MAPE scores.

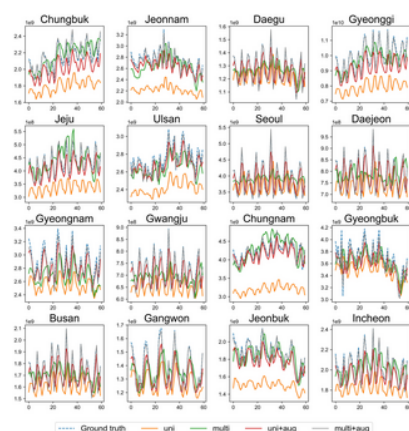


Fig 5. State-level forecasting performance comparison of CNN-LSTM.
<https://doi.org/10.1371/journal.pone.0278071.g005>

Region	RMSE		MAPE	
	Univariable	Multivariable	Univariable	Multivariable
Busan	0.2503	0.1232	0.0313	0.0156
Chungbuk	0.3609	0.1232	0.0313	0.0156
Chungnam	0.2738	0.1232	0.0313	0.0156
Daegu	0.2539	0.1232	0.0313	0.0156
Daejeon	0.2503	0.1232	0.0313	0.0156
Gangwon	0.2609	0.1232	0.0313	0.0156
Gyeonggi	0.2609	0.1232	0.0313	0.0156
Gyeongbuk	0.2609	0.1232	0.0313	0.0156
Gyeongnam	0.2609	0.1232	0.0313	0.0156
Gwangju	0.2609	0.1232	0.0313	0.0156
Incheon	0.2609	0.1232	0.0313	0.0156
Jeonbuk	0.2609	0.1232	0.0313	0.0156
Jeonnam	0.2609	0.1232	0.0313	0.0156
Seoul	0.2609	0.1232	0.0313	0.0156
Ulsan	0.2609	0.1232	0.0313	0.0156

Table 4. Evaluation of forecasting performance of CNN-LSTM.
<https://doi.org/10.1371/journal.pone.0278071.t004>

Ablation studies

To establish the superiority of the proposed model, two experiments are conducted as ablation studies. First, following the 1D CNN kernel visualization technique presented in [7, 31], Fig 6 illustrates the noise-reducing power of each 1D CNN layer. The intermediate output obtained from the second convolution network is observed to be less spiky than the first kernel. This is because each CNN filter reduces the noise of the input dataset. This selectivity over input information also allows cost-effective computation and is, hence, useful for mid-term time series forecasting. The illustration aids the analysis of the number of layers that benefit from smoothing the noise of raw data and selectively extracting important information. Moreover, the loss landscapes of LSTM and the proposed model are compared in Fig 7. In [32], the loss landscape was introduced as a method to represent loss convergence of a model throughout the training process. From the figure, it is evident that the proposed CNN-LSTM hybrid model converges the loss to the minima satisfactorily, and the graph is smooth and concave, indicating facile training. On the other hand, the LSTM-only model exhibits convex loss surfaces and the minima are not visible, indicating that the training of the model is more difficult. This proves that the inclusion of 1D CNN layers before the LSTM layer simplifies training.

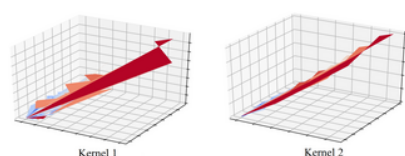


Fig 6. Kernel output comparison of 1D CNN layers.
<https://doi.org/10.1371/journal.pone.0278071.g006>

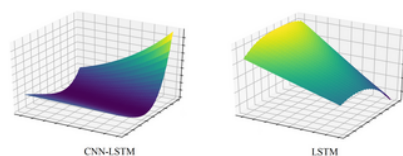


Fig 7. Comparison of loss landscapes of LSTM and CNN-LSTM.
<https://doi.org/10.1371/journal.pone.0278071.g007>

Discussion

Electricity is the backbone of modern society. As such, the accurate prediction of electricity demand and consumption is more significant now than ever before. However, many factors complicate the forecasting accuracy of energy consumption, necessitating the development of advanced forecasting models. In this paper, we presented a new hybrid CNN-LSTM multivariable EEC forecasting techniques that integrates the advantages of CNNs and LSTMs. To incorporate economic insights into energy usage, the proposed technique was expanded using import/export values. The proposed approach enables the prediction of regional monthly electricity consumption. We tested the proposed techniques (components and algorithm) using baseline techniques on univariable historical energy consumption data and compared them to conventional deep learning models. Comprehensive experiment results proved that the proposed technique extracts useful information between international trade values and electricity usage, thereby improving the accuracy of national/regional predictions. We also expanded the proposed technique to consider data interpolation. The proposed mechanism can be further improved using data augmentation.

Conclusions

In this study, we investigated the effects of CNN-LSTM on augmented multivariable time series datasets. We concluded that dimension reduction using the pooling layer of 1D CNN reduces noise and thereby reduces the RMSE and MAPE scores. The LSTM layer was also observed to be well suited to process time series data as it receives inputs for each time step. Extensive experiments and ablation studies were performed, establishing the benefits afforded by the proposed CNN-LSTM architecture paired with multivariable augmentation to provincial time series forecasting for EEC.

References

1. Wang Q, Su M, Li R, Ponce P. The effects of energy prices, urbanization and economic growth on energy consumption per capita in 186 countries. *Journal of cleaner production*. 2019;225:1017–1032.
[View Article](#) • [Google Scholar](#)
2. Arora V, Lieskovsky J. Electricity use as an indicator of US economic activity; 2016.
3. Haes Alhelou H, Hamedani-Golshan ME, Njenda TC, Siano P. A survey on power system blackout and cascading events: Research motivations and challenges. *Energies*. 2019;12(4):682.
[View Article](#) • [Google Scholar](#)
4. Cissokho L, Seck A. Electric power outages and the productivity of small and medium enterprises in Senegal. *Investment climate and business environment research fund Report*. 2013;77:13.
[View Article](#) • [Google Scholar](#)
5. Mujeeb S, Javaid N, Ilahi M, Wadud Z, Ishmanov F, Afzal MK. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. *Sustainability*. 2019;11(4):987.
[View Article](#) • [Google Scholar](#)
6. Kim J, Moon J, Hwang E, Kang P. Recurrent inception convolution neural network for multi short-term load forecasting. *Energy and buildings*. 2019;194:328–341.
[View Article](#) • [Google Scholar](#)
7. Kim TY, Cho SB. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*. 2019;182:72–81.
[View Article](#) • [Google Scholar](#)
8. Amara F, Agbossou K, Dubé Y, Kelouwani S, Cardenas A, Hosseini SS. A residual load modeling approach for household short-term load forecasting application. *Energy and Buildings*. 2019;187:132–143.
[View Article](#) • [Google Scholar](#)
9. Wang L, Lv SX, Zeng YR. Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. *Energy*. 2018;155:1013–1031.
[View Article](#) • [Google Scholar](#)
10. Yu L, Zhao Y, Tang L, Yang Z. Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting*. 2019;35(1):213–223.
[View Article](#) • [Google Scholar](#)
11. Hu H, Wang L, Peng L, Zeng YR. Effective energy consumption forecasting using enhanced bagged echo state network. *Energy*. 2020;193:116778.
[View Article](#) • [Google Scholar](#)

12. Kaytez F. A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. *Energy*. 2020;197:117200.
[View Article](#) • [Google Scholar](#)
13. Aslam S, Herodotou H, Mohsin SM, Javaid N, Ashraf N, Aslam S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*. 2021;144:110992.
[View Article](#) • [Google Scholar](#)
14. Xiong Xin, Hu Xi, Guo Huan. A hybrid optimized grey seasonal variation index model improved by whale optimization algorithm for forecasting the residential electricity consumption. *Energy*. 2021
[View Article](#) • [Google Scholar](#)
15. Hadjout D, Torres JF, Troncoso A, Sebaa A, Martínez-Álvarez F. Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market. *Energy*. 2022
[View Article](#) • [Google Scholar](#)
16. Jana RK, Ghosh I, Sanyal MK. A granular deep learning approach for predicting energy consumption. *Applied Soft Computing*. 2020;89:106091.
[View Article](#) • [Google Scholar](#)
17. Liu C, Sun B, Zhang C, Li F. A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine. *Applied energy*. 2020;275:115383.
[View Article](#) • [Google Scholar](#)
18. K Indicator. Exports and Imports Ratio (GDP); 2022.
19. KOSIS. Electric Power Consumption by Use; 2022.
20. OPEN DATA PORTAL. KEPCO Electricity Usage Status by Region Application; 2022.
21. Open MET Data Portal. Statistic Division; 2022.
22. Korea Customs Service. Trade Statistics for Import/Export by Region; 2022.
23. Wang G, Jia L. Short-term wind power forecasting based on BOMLS K-means similar hours Clustering method. In: 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). IEEE; 2019. p. 1–5.
24. Williams S, Short M. Electricity demand forecasting for decentralised energy management. *Energy and Built Environment*. 2020;1(2):178–186.
[View Article](#) • [Google Scholar](#)
25. Lipton ZC, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:150600019*. 2015.
26. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*. 1997;9(8):1735–1780. pmid:9377276
[View Article](#) • [PubMed/NCBI](#) • [Google Scholar](#)
27. Livieris IE, Pintelas E, Pintelas P. A CNN–LSTM model for gold price time-series forecasting. *Neural computing and applications*. 2020;32(23):17351–17360.
[View Article](#) • [Google Scholar](#)
28. Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: A survey. *Mechanical systems and signal processing*. 2021;151:107398.
[View Article](#) • [Google Scholar](#)
29. Xiao D, Huang Y, Wang H, Shi H, Liu C. Health assessment for piston pump using LSTM neural network. In: 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). IEEE; 2018. p. 131–137.
30. EIA. The U.S.A Energy Information Administration; 2022.
31. Tae-Young Kim, Sung-Bae Cho. Web traffic anomaly detection using C-LSTM neural networks. *Expert Systems with Applications*. 2018.
32. Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, Tom Goldstein. Visualizing the loss landscape of neural nets. *Advances in neural information processing systems*. 2018.