

Article

A Deep Learning Approach for Peak Load Forecasting: A Case Study on Panama

Bibi Ibrahim * and Luis Rabelo

Industrial Engineering & Management Systems Department, University of Central Florida, Orlando, FL 32816, USA; luis.rabelo@ucf.edu

* Correspondence: bibi.ibrahim@knights.ucf.edu; Tel.: +1-407-810-1185

Abstract: Predicting the future peak demand growth becomes increasingly important as more consumer loads and electric vehicles (EVs) start connecting to the grid. Accurate forecasts will enable energy suppliers to meet demand more reliably. However, this is a challenging problem since the peak demand is very nonlinear. This study addresses the research question of how deep learning methods, such as convolutional neural networks (CNNs) and long-short term memory (LSTM) can provide better support to these areas. The goal is to build a suitable forecasting model that can accurately predict the peak demand. Several data from 2004 to 2019 was collected from Panama's power system to validate this study. Input features such as residential consumption and monthly economic index were considered for predicting peak demand. First, we introduced three different CNN architectures which were multivariate CNN, multivariate CNN-LSTM and multihead CNN. These were then benchmarked against LSTM. We found that the CNNs outperformed LSTM, with the multivariate CNN being the best performing model. To validate our initial findings, we then evaluated the robustness of the models against Gaussian noise. We demonstrated that CNNs were far more superior than LSTM and can support spatial-temporal time series data.

Keywords: peak load forecasting; convolutional neural networks; long-short term memory; deep learning



Citation: Ibrahim, B.; Rabelo, L. A. Deep Learning Approach for Peak Load Forecasting: A Case Study on Panama. *Energies* **2021**, *14*, 3039. <https://doi.org/10.3390/en14113039>

Academic Editor: Hongseok Kim

Received: 6 April 2021
Accepted: 21 May 2021
Published: 24 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electricity plays an important role in modern society. Within the past years, energy consumption has been continuously growing worldwide, due to steady increase in population, economic growth, and weather factors. This is becoming more evident in developing countries. As consumer patterns change throughout the day and year, power systems are facing more challenges to balance supply and demand in real time, especially during the peak periods when electricity demand is the highest. The peak demand is presenting a growing trend due to the load growth of different customer sectors such as residential, commercial, etc. A study conducted by Burillo et al., for example, found that the peak demand is projected to grow in Los Angeles, California, due to higher penetration of air conditioning systems in residential and commercial buildings [1]. In many countries, electric vehicles (EVs) are also starting to connect to the grid, which are driving more demand. Modern power systems typically use a balance of expensive peak power plants, small-scale energy storage systems (ESS), and demand response programs to mitigate the peak demand. ESS, especially, are being aggregated in a virtual power plant setting to support the grid during periods of high demand.

Forecasting the peak load has become a major concern for power system planners, as they must ensure that consumer demand is always met for the power grid to operate reliably. Accurate forecasts on the long term will allow planners to determine how much generation capacity is needed to meet future demand, while meeting environment policies.

However, forecasting the monthly peak demand is a challenging problem since it is highly nonlinear and noisy. The peak demand is often correlated with other variables

that are also nonlinear, with strong monthly patterns. Therefore, it is very sensitive to these variations. Conventional time series and regression-based models cannot learn the nonlinear relationship between the inputs and outputs and therefore they may not provide good performance. Nowadays, time series data is inherently large-scale, nonlinear, and noisy. The energy sector is breeding big data from the Internet of things devices (e.g., sensors, and smart meters) that contains valuable information to support decision making. However, analyzing this data can be challenging and requires more advanced algorithms for mining the data.

In the last decade, time series forecasting has undergone a paradigm shift from statistical driven models to a hybrid of statistical and machine learning driven models. As a result, machine learning and deep learning methods are rapidly emerging with better performance for time series problems. These consist of trained algorithms that can learn the nonlinear mapping between the inputs and outputs in big data. Deep learning, in particular, consists of advanced algorithms that apply a more powerful approach for learning complex patterns, by using deeper structures that can extract more meaningful information. Therefore, they are more effective at handling high dimensional data. Among the deep learning methods, convolutional neural networks (CNNs) are becoming more attractive for time series analysis since they can leverage spatial information in different dimensions. Long short-term memory (LSTM) is another popular method that has received increased attention since it can model sequential data very well.

The peak load forecasting is an important research topic within the energy sector that has not been fully explored from a deep learning perspective. Therefore, this leads to the main research question of how deep learning methods can provide better support to these areas that are very nonlinear with complex patterns. Thus, the goal of this study is to develop a forecasting model based on deep learning that can predict the monthly peak demand. A real case study on Panama's power system is presented to validate the model. Two deep learning methods based on CNN and LSTM are introduced. This paper is organized as follows: Section 2 presents the literature review related to peak demand forecasting and the contributions of this study. Section 3 provides a high-level overview of the framework. Section 4 provides a more detailed description of the case study. Section 5 describes the data exploration and transformation process. Section 6 outlines the architecture of the models. Section 7 describes the performance metrics used for model assessment. Section 8 provides the experimental setup and results. Section 9 discusses the results obtained. Section 10 presents the conclusions and future directions for this study.

2. Related Works

A wide variety of models have been proposed in the literature for forecasting peak demand. These models can be classified into four main categories which are (1) regression models, (2) time series models, (3) machine learning models, and (4) deep learning models. Regression models include multiple linear regression (MLR) and simple linear regression. Time series models include autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), and Holt–Winters. These models are well established and are still being used due to their simplicity. Machine learning models include artificial neural networks (ANN), support vector machine (SVM), support vector regression (SVR), k-means clustering, decision trees, and ensembles. These models use shallow networks and tedious feature engineering for learning patterns. Deep learning models, on the other hand, use deeper networks of more than one hidden layer that can automatically learn and extract the features. These include CNN, LSTM, gated recurrent unit (GRU), and recurrent neural networks (RNNs). In recent years, machine learning and deep learning models have become a growing trend for handling time series data, due to their robustness and superior performance. Therefore, they have rapidly become the state-of-the-art methods for time series applications.

Most of the research conducted to this date has been focused on predicting the peak demand for a particular country or smart grid building, to improve demand side management. Different forecasting horizons have been introduced which include daily, weekly, monthly, and yearly. Many of the peak demand models proposed in the literature incorporated weather, population, and economic variables as main indicators. Carvallo et al. conducted a review of the energy forecasting methodologies produced by twelve major utility companies in the United States, between 2003 and 2007 [2]. The study found that most models incorporated historical sales, weather variables such as Cooling Degree Days (CDD) and Heating Degree Days (HDD), as well as demographic projections of customer loads [2]. Another major finding of this study was that most utilities overestimated the forecasts for peak demand since they did not project slow economic growth [2].

Several time series regression models have been introduced in the literature. Rallapalli and Ghosh earlier on proposed a multiplicative SARIMA model to predict the monthly peak demand for five regions in India [3]. The model was compared with the official forecasts done by the Central Authority. SARIMA achieved the best performance, since it could capture the seasonality better. The mean absolute percentage error (MAPE) ranged between 0.93 and 3.94% for the five regions [3]. Sigauke and Chikobvu developed three time series regression models for forecasting daily peak electricity demand in South Africa, using historical data from 2000 to 2010. The best-selected model incorporated maximum temperature, minimum temperature, and peak day temperature as significant predictors [4]. Khorsheed developed four MLR models to predict monthly energy peak demand for the Kingdom of Bahrain, using eight years of historical data. The best model obtained had an R-squared value of 0.94, which incorporated the month and maximum temperature [5]. AL-Hamad and Qamber evaluated two models based on MLR and adaptive neuro-fuzzy inference systems (ANFIS) to predict long term peak demand for six countries in the Gulf Cooperation Council region up to the year 2024 [6]. Both models incorporated gross domestic product (GDP) and population as input variables. ANFIS had the highest prediction accuracy, with an average percentage error of 0.53% [6]. Almazrouee et al. recently evaluated two methods, Prophet (Facebook's forecasting tool) and Holt-Winters, for forecasting the long term daily peak demand for Kuwait up to the year 2030, using ten years of historical data [7]. The study found that the Prophet model performed better than Holt-Winters with MAPE of 1.75% [7].

In recent years, there has been an increasing number of studies based on machine learning and deep learning approaches for forecasting the peak load. Pannakkong et al. applied ensemble machine learning models based on ANN, SVM, and deep belief networks (DBN), for predicting monthly peak demand in Thailand [8]. The ensemble model of ANN and DBN performed the best for predicting demand one month ahead, with MAPE of 1.44% [8]. Kwon et al. compared a deep learning LSTM model with MLR for predicting weekly peak load in Korea [9]. Three input features were used: weekly temperature, GDP, and previous peak demand. The study found that LSTM performed better than MLR, with MAPE of 2.16% and 2.67% for the years 2017 and 2018, respectively [9].

Elkamel et al. conducted a unique study by comparing a MLR model with three different CNNs for predicting monthly electricity consumption for Florida, using data from 2010 to 2016 for training [10]. The CNN model outperformed MLR, with MAPE of 2.5%. The authors found that the model performed even better when the dataset was enhanced [10].

Son and Kim proposed a LSTM model to predict monthly residential demand, a primary driver of peak demand in South Korea [11]. The model incorporated mostly eight weather variables such as CDD, maximum temperature, and wind speed. The proposed model was compared with four benchmark methods: SVR, ANN, ARIMA, and MLR. LSTM performed the best, with MAPE of 0.07% [11]. Kim et al., on the other side, proposed various ensemble trees (bagging and boosting) for forecasting the peak load of small industrial facilities [12]. The model incorporated the previous hourly load patterns as the main feature. The isolation forecast algorithm was introduced to identify outliers. They found

that decision trees in overall were not suitable for predicting the peak demand [12]. Lai et al. [13] recently evaluated deep neural networks (DNN) with enhanced historical loads to predict daily peak demand for Austria, Czech, and Italy. The model was benchmarked against LSTM, ARIMA, and deep residual networks (DRN). The study concluded that deep neural networks outperformed all the models when tested on the three different datasets [13]. Mehdipour et al. [14] performed an empirical comparison of four machine learning methods that are widely used, i.e., SVR, gradient boosting regression trees, ANN, and LSTM for predicting the load demand of individual buildings. It was demonstrated that LSTM and gradient boosting methods were able to effectively detect the daily peak. Shirzadi et al. [15] proposed a high-level framework based on machine learning and deep learning methods, i.e., SVM, random forest (RF), non-linear autoregressive exogenous (NARX), and LSTM to predict daily load for Bruce County, Canada. Temperature and wind speed were the main input features used for training the models. Both LSTM and NARX provided the best accuracy for forecasting the peaks. RF, on the other side, overestimated the highest peaks but still was within the acceptable range, with MAPE of 10.25%. Atef and Eltawil [16] compared deep bidirectional-LSTM, unidirectional LSTM, and SVR for electricity forecasting and found that the bidirectional LSTM can stably predict demand without overfitting. SVR, on the other side, was not able to follow the demand pattern. They also found that increasing the network depth did not improve predictive performance.

In addition to the works cited, hybrid models have also been proposed to improve forecasting accuracy. Firstly, Laouafi et al. [17] introduced a hybrid model consisting of fuzzy c-means clustering and ANFIS for forecasting daily peak demand, where clustering was used to classify the daily peaks according to temperature. Dai et al. [18] evaluated the robustness of a hybrid model based on ensemble mode decomposition with adaptive noise (EMDAN) and SVM for daily peak load forecasting. The grey wolf optimization was adopted to improve convergence. The model was found to be reliable and achieved superior performance than SVM, with MAPE of 0.25%. Park et al. [19] proposed a machine learning model that combined Holt-winters and decision trees, i.e., RF and bagging, to accurately predict the peak load, which was further utilized to improve ESS scheduling strategy. The hybrid model based on Holt-Winters and bagging trees was found to be superior, with MAPE of 6.7%. Nepal and Yamaha [20] recently presented a hybrid learning approach based on k-means clustering and neural networks to predict the peak demand of university buildings. The data was grouped into 6 clusters that presented similar demand behavior. The model performed far better than neural networks, with MAPE of 6.1%. Kazemzadeh et al. introduced a unique hybrid approach that combines ARIMA, ANN and SVR to forecast the yearly peak load in Iran, using historical data of twenty six years [21]. The study found that the hybrid method outperformed the individual ARIMA and ANN methods, with MAPE of 0.97% [21]. Wu et al. [22] recently evaluated a combined GRU-CNN model for short term load forecasting and found that it was able to effectively follow the nonlinear load pattern and predict the highest peaks.

Table 1 lists the accuracy of the best performing models obtained for the papers reviewed, in terms of MAPE.

Table 1. Comparison of papers reviewed for peak demand forecasting.

Authors and Ref.	Forecasting Resolution	Methods	Historical Data	MAPE (%)
Rallapalli and Ghosh [3]	Monthly	SARIMA	2005–2011	0.93–3.94%
Sigauke and Chikobvu [4]	Daily	Regression	2000–2010	1.78%
Khorsheed [5]	Monthly	MLR	2009–2016	Not reported
AL-Hamad and Qamber [6]	Yearly	MLR ANFIS	Not reported	Not reported

Table 1. Cont.

Authors and Ref.	Forecasting Resolution	Methods	Historical Data	MAPE (%)
Almazrouee et al. [7]	Daily	Prophet Holt–Winters	10 years	1.75%
Pannakkong et al. [8]	Monthly	ANN SVM DBN	2002–2017	1.44%
Kwon et al. [9]	Weekly	LSTM MLR	2007–2016	2.16–2.67%
Elkamel et al. [10]	Monthly	CNN MLR	2010–2016	2.50%
Son and Kim [11]	Monthly	SVR ANN ARIMA MLR LSTM	1991–2012	0.07%
Kim et al. [12]	Daily	Ensemble trees	2017–2018	44.05–53.08%
Lai et al. [13]	Daily	DNN DRN ARIMA LSTM	2006–2015	2.79–3.48%
Mehdipour et al. [14]	Daily	SVR Gradient boosting ANN LSTM	2011–2014	Not reported
Shirzadi et al. [15]	Hourly	SVM RF NARX LSTM	2010–2019	4.2%
Atef and Eltawil [16]	Hourly	Bidirectional LSTM LSTM SVR	2015–2018	0.22%
Laouafi et al. [17]	Daily	Fuzzy clustering ANFIS	2012–2014	2.79%
Dai et al. [18]	Daily	EMDAN SVM	92 days	0.25%
Park et al. [19]	Minutes	Holt–Winters + ensemble trees	2015–2017	6.70%
Nepal and Yamaha [20]	Hourly	k-means clustering ANN	2017–2018	6.1%
Kazemzadeh et al. [21]	Yearly	ARIMA ANN SVR	26 years	0.97%
Wu et al. [22]	Minute	GRU + CNN	August 2018	2.30%

Based on the discussion above, we found some missing gaps in the literature. Although deep learning is becoming more popular for predicting peak electricity demand, we found that CNNs have not been explored for this particular problem. Only one study [10]

attempted to evaluate CNN and that was for predicting energy consumption. Therefore, the contribution of this paper is significant, since no similar study has been done previously. Furthermore, it was found that no previous work has been conducted on Panama's power system, making it a potential case study.

3. Methodology

Figure 1 provides a high-level overview of the framework proposed. The framework consists of two deep learning methods based on CNNs and LSTM that were evaluated for forecasting the peak load one month ahead. Therefore, a potential case study on the Panama's power system was presented to validate the models. First, historical data on the peak demand from 2004 to 2019 was collected to analyze the current electricity demand trend for Panama and determine which methods were the most suitable for modeling this problem. In addition, this study collected monthly data for several input features to understand how they were related to the peak demand. Next, the data was preprocessed and transformed into the appropriate input shape prior to training the models. The data preprocessing consisted of mainly two steps: normalization and partitioning the data into a training and test set. First, the training data is fed to the corresponding CNN and LSTM models, which enabled feature extraction and sequential learning, respectively. The CNNs consisted of three different architectures based on multivariate CNN, multivariate CNN-LSTM and multihead CNN, which are explained more in-depth in Section 6. The models were built in Keras, an open source tool for neural networks and deep learning. Section 6 also provides a more detailed description of the architecture of the models. For the CNN models, it was important to define the hyperparameters such as the number of convolutional layers and pooling layers, along with the number of filters and filter size. In the case of LSTM, we had to define the number of LSTM layers and hidden units.

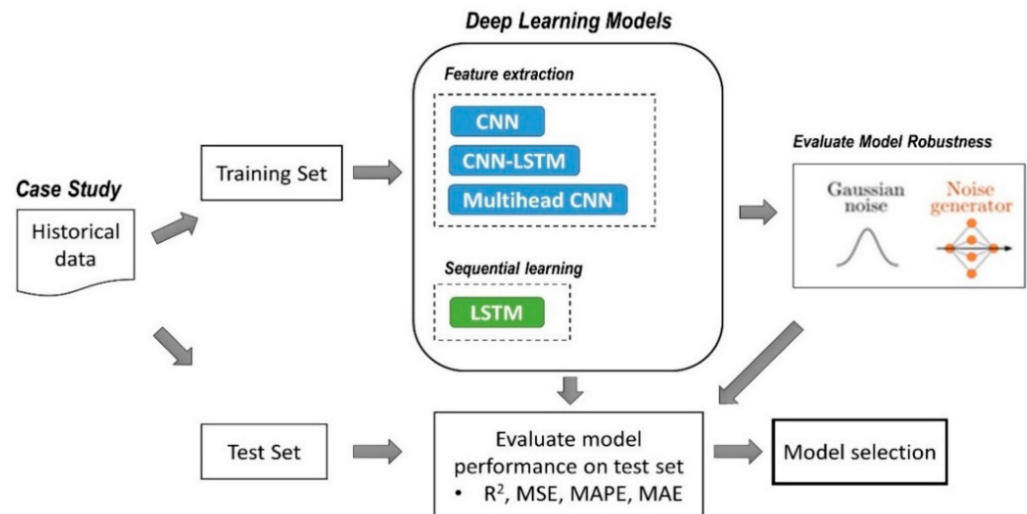


Figure 1. Deep learning framework proposed for forecasting peak demand.

In order to provide a fair comparison, the assessment of the models was rigorously conducted in two steps: without and with Gaussian noise. Therefore, the models were first trained without Gaussian noise to gain a better insight of the model performance. The models were evaluated on the test set, using four performance metrics that are explained in Section 7. Next, we proceeded to evaluate the robustness of the CNN and LSTM models against Gaussian noise. Finally, we selected the best model for forecasting the peak demand.

3.1. Convolutional Neural Networks

Convolutional neural networks are a special type of deep learning networks that have rapidly transformed the computer vision field, achieving state of the art performance in image classification and object detection tasks. Currently, it is one of the most established

methods for spatial pattern recognition. Their success was first demonstrated in 1989 by LeCun, who developed the LeNet model for classifying handwritten digits [23]. In the last decade, CNNs have been actively researched in terms of representation learning, depth, and pooling operations [24], which have contributed to their remarkable progress. Recently, remote sensing and medical images are becoming more increasingly available in big data repositories, which brings new requirements for novel methods that can leverage data structures with increased spatial–spectral dimensions. Therefore, CNNs are being more widely introduced and challenged to support image segmentation and classification.

CNNs are also becoming more attractive for solving time series forecasting problems such as wind speed prediction. Time series data collected from energy systems are very nonlinear and granular. They can exhibit not only sequential but also spatial patterns, which cannot be captured by RNNs. Therefore, CNNs have been extensively investigated to understand how they can better handle univariate and multivariate time series data. Depending on the feature extraction complexity, different CNN structures can be applied such as 1D CNN, two-dimensional (2D) CNN, and three-dimensional (3D) CNN. The 1D CNN, in particular, has been mainly studied for time series analysis [25–27], human motion recognition, and signal processing applications, in which the filter moves only in one direction. Many authors also developed hybrid CNN-LSTM models to perform deeper feature extraction.

CNN uses a hierarchical learning approach, which starts by extracting the lower-level features in the first layers followed by more abstract features in the last layers. Two main attributes that distinguish CNNs from traditional neural networks are local connectivity and parameter sharing. One problem that originates in fully connected networks is that more parameters need to be learned. Therefore, CNNs take basic feed-forward ANN one step forward, by connecting the neurons to only local regions in the spatial input data. The weights of the neurons are shared across the space, which can help to segment the regions that contain similar local features. Therefore, locally connected layers can significantly reduce the number of learnable parameters and the training time.

There are four main components and operations that are core to a CNN which are: convolutional layers, nonlinearity, pooling layer, and fully connected layers. Figure 2 demonstrates the overview of the CNN architecture.

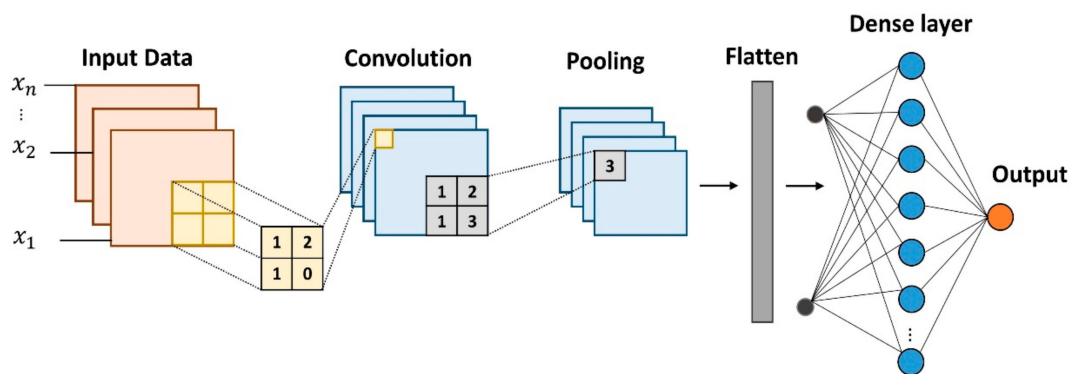


Figure 2. CNN deep learning architecture.

The first layer of the CNN is the convolutional layer, which consists of constructive filters that are convolved with the inputs as they slide over the dataset, generating a set of hierarchical feature maps of reduced size (Figure 3). These maps provide a more detailed representation of the different types of features extracted that are relevant for predicting the output.

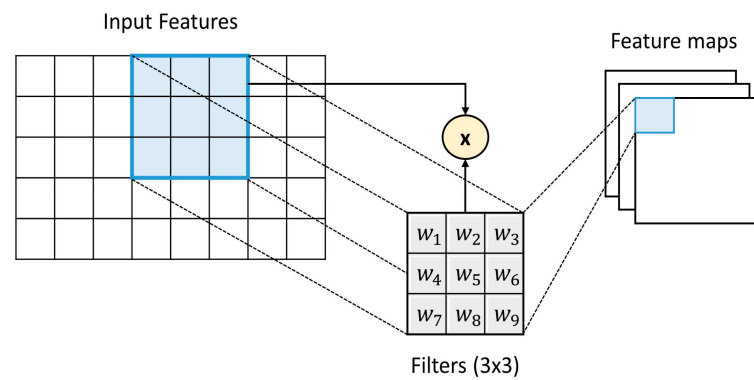


Figure 3. Feature maps generated from the convolutional operation.

The filters consist of weights that are learned during training, which help them determine what type of features need to be extracted. Next, the feature maps are passed through an activation function to introduce nonlinearity and learn the input to output nonlinear mapping. Rectified linear unit (ReLU) is an activation function that is commonly used because it can handle the vanishing gradient problem and can converge faster. The output feature maps are then further downsized by applying a second pooling layer, which keeps only the dominant features. The max pooling layer, for example, takes the maximum value of each patch in the feature map. Finally, they pass through fully connected layers to predict the final output. The output of the convolutional layer is provided in Equation (1).

$$c_i^{l,j} = \sigma \left(\sum_{m=1}^M w_m^j x_{i+m-1}^{0,j} + b_j \right) \quad (1)$$

where l is the convolutional layer index, σ the sigmoid activation function, w_m^j is the weight for the j th feature map, m is the filter size, x_i^0 is the input vector convolved and b_j is the bias.

3.2. Long Short-Term Memory Networks

The second model proposed is based on long short-term memory network (LSTM), which is benchmarked against the CNN models. LSTM consists of memory blocks that have a special feature known as the cell state c_t , which helps them to remember long term information, as demonstrated in Figure 4.

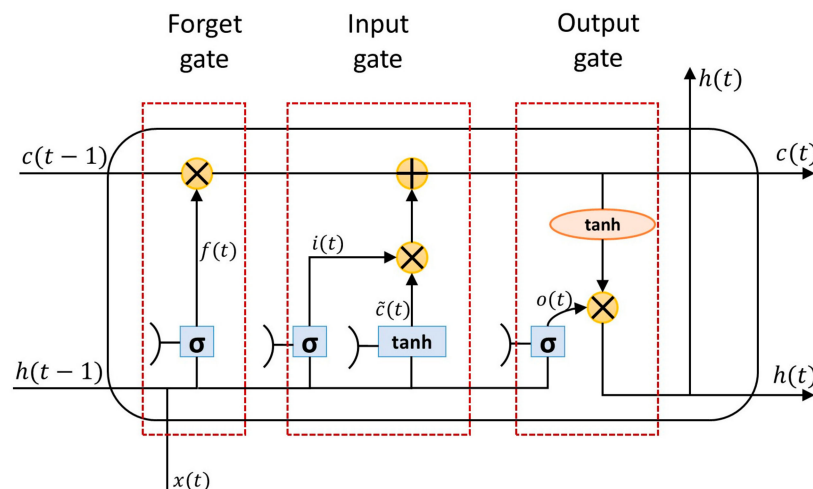


Figure 4. LSTM architecture.

The cell state runs through the top, which stores information learned from previous time steps. This information is updated at each time step using three main gates that

control what information should be added or removed from the cell state. These gates are the forget gate, input gate, and output gate. At each time step, the LSTM block takes two entries which are the previous hidden state h_{t-1} and the current input x_t . These are processed by three gates that decide what information is useful to update the cell state.

The forget gate is the first gate in the LSTM that decides on what information to forget or keep from the previous cell state Equation (2).

$$f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \quad (2)$$

where f_t is the output of the forget gate; W_{fh} , W_{fx} and b_f represent the parameters (weights and bias); and σ the sigmoid activation function.

The input gate decides what new information will be added to the cell state. This is accomplished in two simultaneous steps, each one with a different activation function. Therefore, the entries are passed through the sigmoid function Equation (3), which outputs the values between 0 and 1. These entries are also passed through the tanh function Equation (4) that creates a vector of new candidate values to update the cell state. The results of both activation functions are multiplied and then added to update the cell state Equation (5). Lastly, the output gate determines what information of the cell state to output based on Equations (6) and (7).

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \quad (3)$$

where i_t is the output of the input gate. W_{ih} , W_{ix} , and b_i are the parameters of the input gate.

$$\tilde{c}_t = \tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}) \quad (4)$$

where \tilde{c}_t are the new candidate values and tanh the activation function

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (5)$$

where c_t is the updated cell state

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \quad (6)$$

where W_{oh} , W_{ox} and b_o are the weights and bias of the output gate.

$$h_t = o_t \cdot \tanh(c_t) \quad (7)$$

where h_t represents the output vector of the memory cell.

4. Case Study

Many countries worldwide are facing problems to operate the power grid reliably and sustainably due to dynamic changes in supply and demand patterns. Panama is a relatively small country with a reliable power system. In 2016, the population surpassed 4 million. Within the past years, Panama has been experiencing strong economic growth accompanied by the development of new businesses and commercial centers, which have contributed to the increase in electricity demand. The peak demand, especially, has exhibited a growing trend from the years 2004 to 2019, as observed in Figure 5.

The data was provided by the National Secretary of Energy of Panama. Power grid planners recently became concerned as the peak demand increased drastically from 1665 megawatts (MW) in 2018 to 1961 MW in 2019. Currently, Panama has enough generation capacity to meet demand. As shown in Figure 5, demand is met by using a diverse portfolio of generation sources such as hydro, thermal, wind, and solar power plants. However, thermal power plants present a disadvantage that they have higher emissions, and therefore the grid cannot fully depend on these sources. Hydro power plants also create a problem during the dry season since there is not enough water to fill the reservoirs.

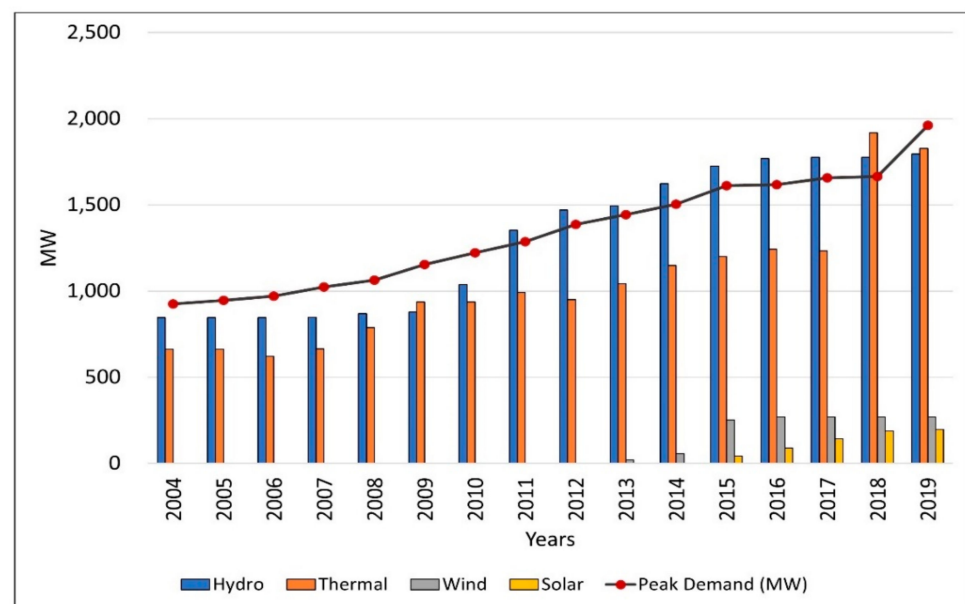


Figure 5. Peak demand growth of Panama by year from 2004 to 2019.

Panama has set the goal of meeting the future electricity demand with cleaner energy and has started investing in more wind and solar plants over the recent years. In 2019, wind and solar accounted for 468 MW of renewable energy capacity installed. Therefore, forecasting the peak load by month can help planners to manage existing and future power plant investments for meeting electricity demand. At the present, no model has been developed for forecasting the peak demand for Panama. Furthermore, it is still not clear which sectors are driving more demand.

5. Materials

5.1. Data Exploration and Preprocessing

This study was conducted on Panama's power system to determine what are the main challenges they are facing, in terms of managing peak demand growth. First, we proceeded to collect monthly peak demand data for the time period 2004 to 2019 from the National Dispatch Center (Spanish: Centro Nacional de Despacho) [28], in order to have a better perspective of the historical trend. The data is available to the public on the website [28]. The peak demand time series can be observed in Figure 6. It is evident that the peak demand is highly nonlinear and noisy. The highest demand was 1907 MW, which occurred in August 2019.

A boxplot was also constructed to determine if the peak demand presented monthly patterns (Figure 7). Based on the boxplot, it is clear that the peak demand significantly varied across the months, particularly for May, August, and December which had the highest demand recorded.

The next step of our data exploration consisted of identifying which variables had the most effect on predicting peak demand. A list of eight variables were identified for building the model, which included economic indicators, weather related variables, and energy consumption of different sectors in megawatt hours (MWh). These are summarized in Table 2.

Data for each variable was collected monthly for the period 2004 to 2019 from different sources. Weather variables such as minimum temperature and maximum temperature were provided by the Panama Canal Authority since the data is not publicly available. Weather data was collected specifically for Panama City, which is a densely populated area that has experienced fast business and commercial growth over the past ten years. Businesses and large commercial centers have air conditioning systems installed which

tend to drive more electricity demand. Therefore, the Panama City was found to be a representative region for conducting the study.

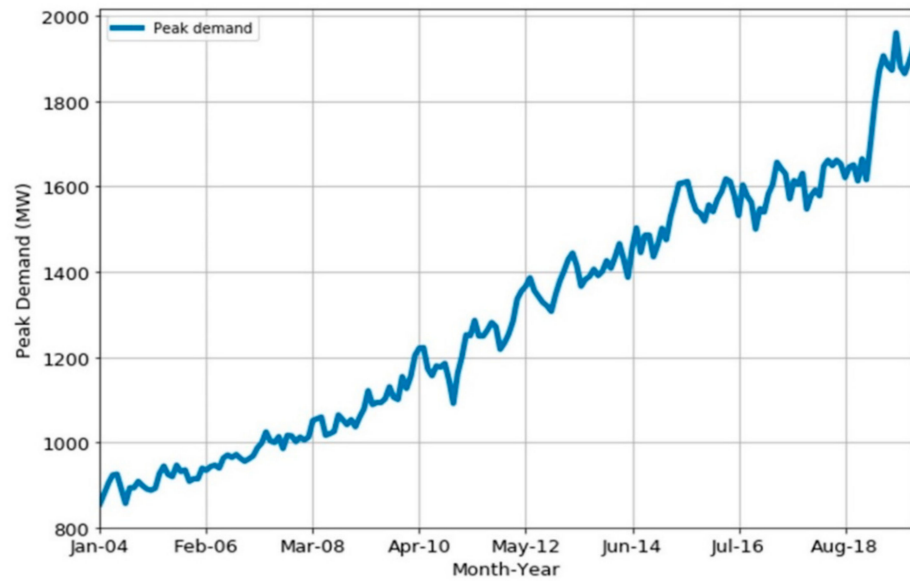


Figure 6. Monthly peak demand growth of Panama from 2004 to 2019.

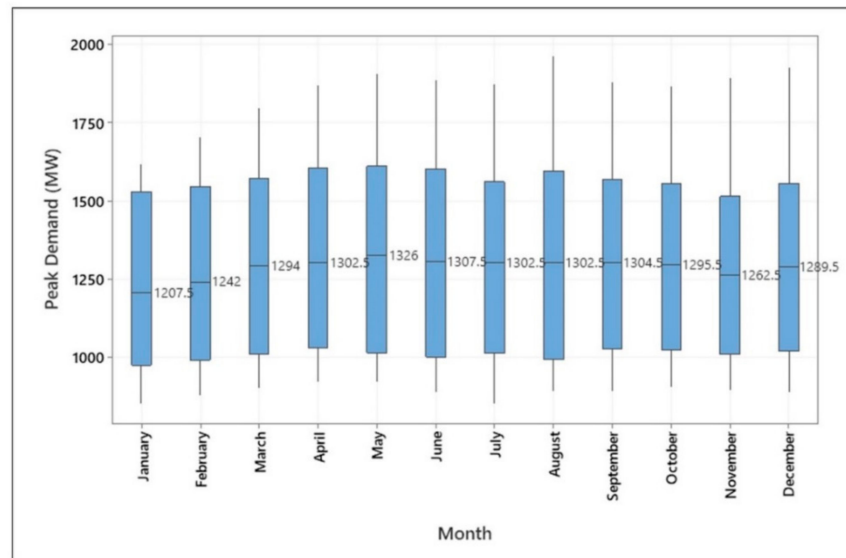


Figure 7. Boxplot of the peak demand by month.

Table 2. Variables considered for predicting the peak demand.

Variable	Description
Peak demand (Response variable)	Maximum electricity demand of Panama in MW
Month	Month 1–12 (January to December)
Residential Consumption	Total residential consumption in MWh
Big Clients Consumption	Total consumption of big clients in MWh
Commercial Consumption	Total commercial consumption in MWh
Monthly Economic Activity Index	Monthly economic indicator of Panama
Minimum Temperature	Minimum Temperature of Panama City in °C
Maximum Temperature	Maximum Temperature of Panama City in °C
Previous peak demand	Peak demand of previous months

The rest of the indicators such as residential consumption, big clients consumption, commercial consumption, and monthly economic activity index were collected from the National Institute of Statistics and Census of Panama. The big clients represent the consumers that have a demand superior to 100 kilowatts and can purchase energy from the wholesale electricity market. Therefore, this variable was considered to be significant for predicting peak demand. The monthly economic activity index is also an important indicator which estimates the activity of 14 different economic sectors in Panama such as agriculture, electricity, and water supply, etc. In addition, a categorical month indicator was used to capture the monthly variations. The peak demand from previous months were also used as important lags to predict the future demand.

Each variable was plotted to determine which ones presented a similar trend to the peak demand (Figure 8). It was clear that residential consumption (Figure 8a), commercial consumption (Figure 8c), and monthly economic index (Figure 8d) exhibited a similar long-term upward pattern as the peak demand. However, commercial consumption remained relatively stable between the period 2015 to 2019. On the other hand, the variable “big clients consumption” (Figure 8b) followed a different pattern, where there was an increased spike in demand between the period 2017 and 2019 that could have contributed to the sudden peak demand growth. With respect to the weather variables, the minimum temperature (Figure 8e) and maximum temperature (Figure 8f) presented highly nonlinear and fluctuating patterns over the time. The maximum temperature, in particular, did not increase over the years and therefore could not account for the upward trend pattern of the peak demand.

5.2. Feature Importance

After the data were collected and analyzed, the next step consisted of identifying which variables were more significant for predicting peak demand through the feature selection process. Currently, there is a wide range of feature selection methods available for machine learning. For this study, the SelectKBest method was used from the Scikit-learn library in Python. It has the advantage that it can be used for both classification and regression problems. This method retains the best K number of features with the highest score. The features were scored using the f regression function. In order to determine the feature importance, a total of seven features/variables were selected and ranked according to the SelectKBest method. The results are provided in Table 3.

Table 3. Feature Ranking According to the *SelectKBest* Method.

Rank	Features	Score
1	Commercial Consumption	5672.69
2	Monthly Economic Activity Index	4081.79
3	Residential Consumption	3566.18
4	Big Clients Consumption	357.35
5	Minimum Temperature	7.53
6	Maximum Temperature	7.39
7	Month	0.01

Based on Table 3, the top four features were: commercial consumption, monthly economic activity index, residential consumption, and big clients consumption. Variables such as minimum temperature, maximum temperature, and month were not considered to be significant.

5.3. Data Preprocessing

5.3.1. Partitioning

Data preprocessing was necessary prior to building and training the models. First, all the variables were merged into one excel file. The file was uploaded in Python. Then, the entire dataset was split into the training set (80%) and test set (20%), while preserving

the temporal order of the data. Data from January 2004 to September 2016 was used as the training set (153 data points) and data from October 2016 to December 2019 was used as the test set (39 data points). This process was applied to all the models.

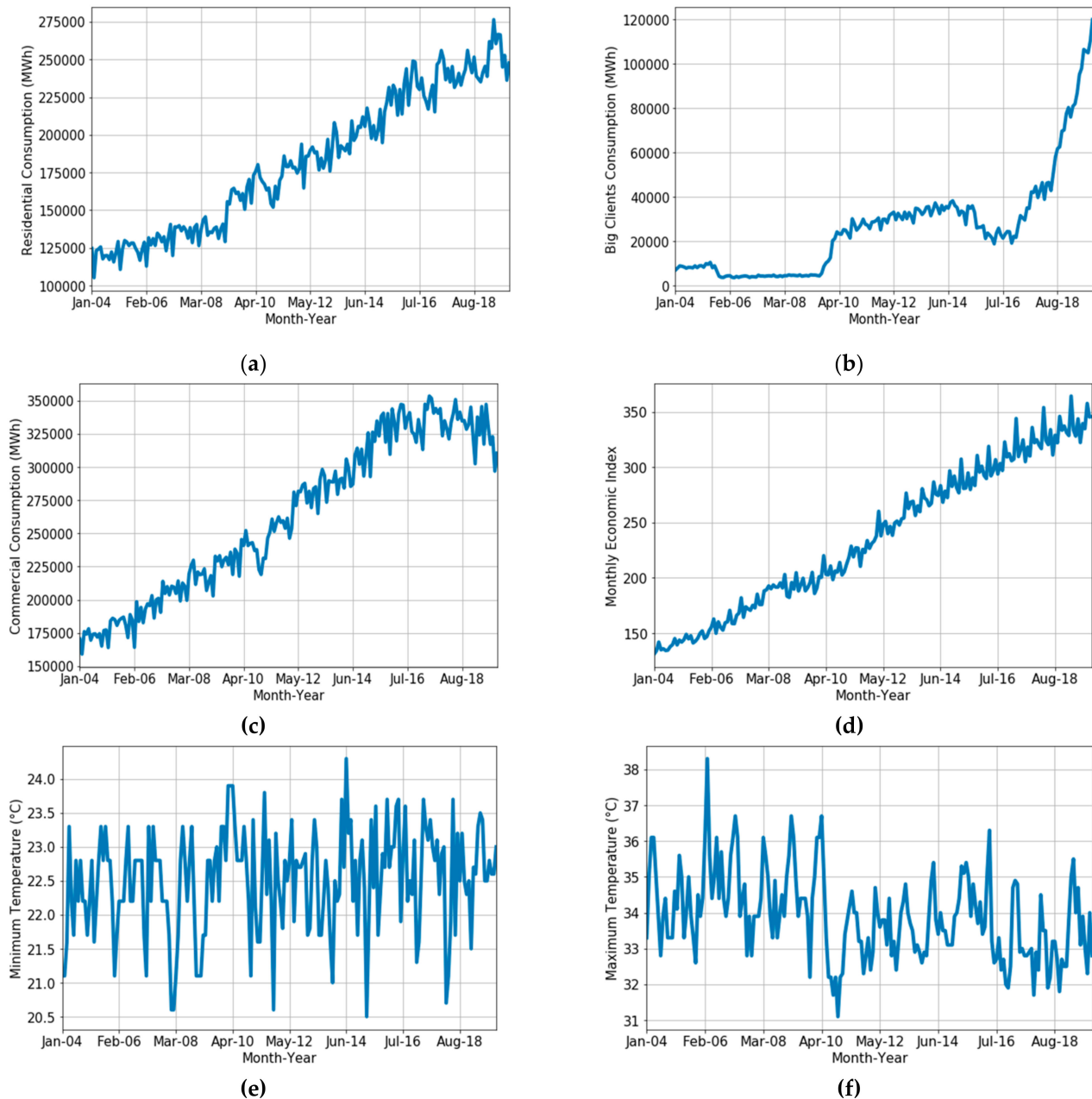


Figure 8. Comparison of the input variables trend for predicting peak demand (a) Residential consumption; (b) Big clients consumption; (c) Commercial consumption; (d) Monthly economic index; (e) Minimum temperature, and (f) Maximum temperature.

5.3.2. Normalization

The second step consisted of normalizing the data, since each input feature has a different scaling range, which can affect the learning process. Therefore, the data was normalized between 0 and 1 values, using the min-max scaler from the Scikit-learn package in Python. First, the training set was normalized. Then, the test set was normalized according to the parameters learned from the training set.

5.3.3. Data Transformation

The third step involved defining the number of previous timesteps and features that will be used for predicting the next month's peak demand, by using the sliding window approach. This process was necessary since the one-dimensional (1D) CNN and LSTM methods receive a three-dimensional tensor (samples, timesteps, and features). To accomplish this, the training dataset was divided into a list of input and output sequences. An example of how the data was transformed is provided below. Equation (8) shows the matrix containing the original input features. Then, the first input and output are generated, which are represented by Equations (9) and (10), respectively. The input contains the features with the specified timesteps to look back, while the output contains the variable to be predicted.

$$\text{Input Data} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ x_{13} & x_{23} & \dots & x_{n3} \\ x_{14} & x_{24} & \dots & x_{n4} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \dots & x_{nn} \end{bmatrix} \quad (8)$$

$$\text{Input 1} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ x_{13} & x_{23} & \dots & x_{n3} \end{bmatrix} \quad (9)$$

$$\text{Output 1} = [x_{n4}] \quad (10)$$

6. Model Architecture

6.1. Multivariate CNN

The first approach considered was a multivariate CNN model, which has the advantage that it can handle multiple time series for predicting the output. Each time series data is processed as a separate channel by one CNN model that learns the spatial mapping between the inputs and output to extract meaningful local to global information. As a result, one single feature map is generated that contains all the main features extracted from each time series. First, we started building the models with some of the top selected features that were obtained in Section 5.2 (Table 3). These were monthly economic activity index, residential consumption, and big clients consumption. The peak demand of the previous months was also used as a significant input feature. The month indicator was incorporated to capture the monthly patterns, although it was not found to be important. Next, we proceeded to incorporate the variables that had the lowest feature importance which were minimum temperature and maximum temperature to evaluate their effect. The multivariate CNN structure can be observed in Figure 9.

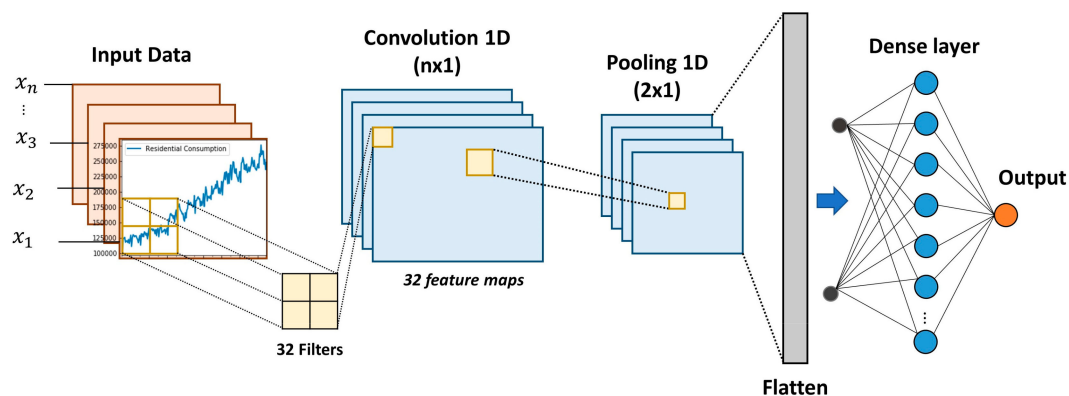


Figure 9. Multivariate CNN architecture.

As observed in Figure 9, the input data is provided to the CNN model so it can extract the features. A sliding window is used to predict the peak demand for the next month, which is explained in Section 5.3.3. The first layer used is a 1D convolutional layer that consists of 32 filters of different sizes, generating a set of 32 feature maps as output. Next, a 1D max pooling layer of size 2×1 is used to reduce the dimensions of the feature maps, while keeping only the dominant features for predicting the peak demand. A flatten layer was then used to output the information as one long vector. Lastly, it is passed through a fully connected layer of 50 hidden neurons to predict the final output. The Adam optimizer was used for training the model. Table 4 summarizes the model architecture.

Table 4. Layer configuration of the multivariate CNN model.

Layer	No. of Convolution Filters/Neurons	Kernel Size	Stride	Activation Function
Convolution 1D	32	$n^1 \times 1$	1	ReLU
Max Pooling 1D		2×1	1	
Dense	50			ReLU
Dense (Output)	1			

¹ Filter sizes of 2 to 11 were used.

6.2. Multivariate CNN-LSTM

The second model proposed is based on CNN-LSTM structure, which combines the advantages of both methods to model time series data that exhibit both spatial and temporal patterns. The model only included four input features: month, residential consumption, big clients consumption and the peak demand from previous months. A sliding window approach was used, where the previous timesteps of three months is used to predict the peak demand for the next month. The model architecture is demonstrated in Figure 10.

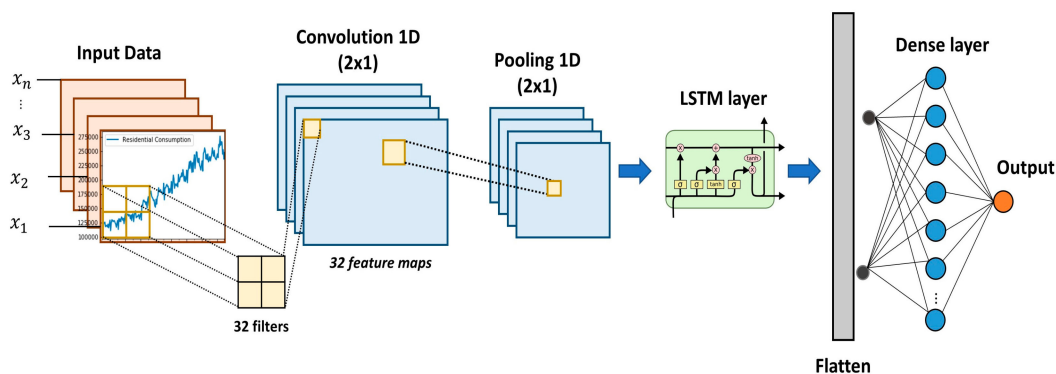


Figure 10. Multivariate CNN-LSTM architecture.

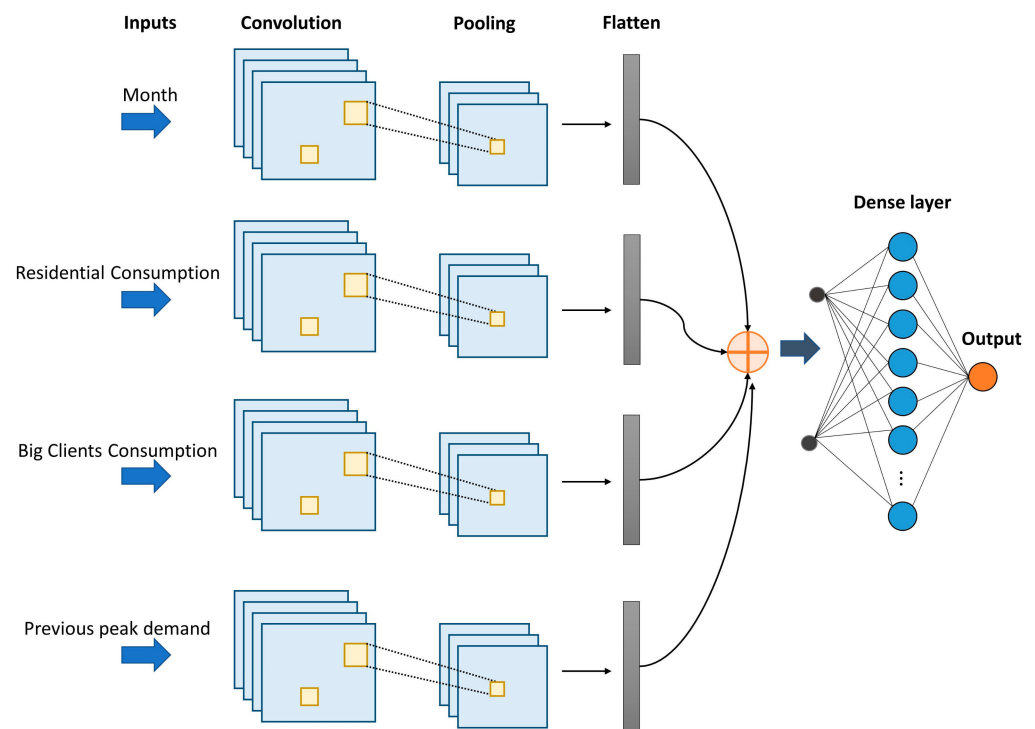
The CNN-LSTM model involved two main processes which are feature extraction and sequential learning that are performed by the CNN and LSTM, respectively. First, the CNN input layer receives the four input features. The convolutional layer applies 32 filters of size 2×1 to extract the features. A max pooling layer of size 2×1 is then applied to keep only the relevant features. The extracted features are further transformed and passed through a LSTM layer of 50 hidden units, to learn the sequential patterns that could not be captured by the CNN. Next, a flatten layer is used followed by a fully connected layer of 50 hidden neurons. Table 5 summarizes the model architecture.

Table 5. Layer configuration of the multivariate CNN-LSTM model.

Layer	No. of Convolution Filters/Neurons	Kernel Size	Stride	Activation Function
Convolution 1D	32	2×1	1	ReLU
Max Pooling 1D		2×1	1	
LSTM	50			ReLU
Dense	50			ReLU
Dense (Output)	1			

6.3. Multihead CNN

The third model evaluated is based on the multihead CNN structure that consists of independent CNN models, known as convolution heads, which process each time series separately to learn the different feature representations. This model requires more data preprocessing since each time series needs to be transformed into input and output sequences. Figure 11 provides an overview of the multihead CNN structure.

**Figure 11.** Multihead CNN architecture.

The model incorporates four input features: month, residential consumption, big clients consumption, and previous peak demand. Therefore, four sub CNN models were needed to process each time series. The CNN model first receives the input time series. Next, the features are extracted by two layers: 1D convolutional layer that consists of 32 filters of size 2×1 and 1D max pooling layer of size 2×1 , with a stride of 1. A separate feature map is produced for each time series, where the information is passed through a flatten layer. The results of each CNN model are merged and then go through a fully connected layer of 50 hidden neurons to predict the final output. Table 6 summarizes the model structure for each CNN model.

Table 6. Layer configuration of the multihead CNN model.

Layer	No. of Convolution Filters/Neurons	Kernel Size	Stride	Activation Function
Convolution 1D	32	2×1	1	ReLU
Max Pooling 1D		2×1	1	
Dense	50			ReLU
Dense (Output)	1			

6.4. LSTM

The LSTM model proposed consists of three layers: input layer, hidden layer, and output layer. First, the input layer receives four input features for predicting the peak demand: month, residential consumption, big clients consumption, and previous peak demand. A sliding window of three timesteps was used. Next, two LSTM layers of 50 hidden units were used to process the input data, in order to learn the temporal patterns. Multiple LSTM hidden layers were considered to extract richer information. The transformed data was passed on to a flatten layer followed by a fully connected layer of 50 hidden neurons. The LSTM structure is summarized in Table 7.

Table 7. Layer configuration of the LSTM model.

Layer	No. of Neurons	Activation Function
LSTM	50	ReLU
LSTM	50	ReLU
Dense	50	ReLU
Dense (Output)	1	

7. Performance Metrics

Four performance metrics were used to evaluate the performance of the models on the test set. These include the coefficient of determination (R^2), mean squared error (MSE), MAPE, and mean absolute error (MAE).

The R^2 , also known as the coefficient of determination, is an important statistical measure used for regression analysis. It measures how well the regression model fits the data and it ranges between 0 and 1. It can be interpreted as the proportion of variance in the response that is explained by the predictors. The R^2 is calculated based on the sum of squares of residuals (SSR) and sum of squares total (SST). SSR measures the deviation of the predicted values from the observed values while SST represents the squared differences between the observed values and its mean. The equations for SSR and SST are given as:

$$SSR = \sum_i (y_i - \hat{y}_i)^2 \quad (11)$$

$$SST = \sum_i (y_i - \bar{y})^2 \quad (12)$$

Therefore, the R^2 is useful to test the model significance. The higher the R^2 value, the more explanatory power it has. The equation for R^2 is provided below:

$$R^2 = 1 - \frac{SSR}{SST} \quad (13)$$

The MSE measures the average squared difference between the predicted values and the actual values. It is given as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

The MAPE measures the forecasting accuracy of the model in terms of percentage, which makes it easier to interpret. It calculates the absolute difference between the predicted and actual values. It is given as:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (15)$$

The MAE calculates the absolute difference between the actual values and predicted values. It is more robust to outliers since it takes the absolute difference of the values. It is provided below:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

where n is the number of data points, y_i is the observed data, \hat{y}_i the predicted values, and \bar{y} the mean of the observed data.

8. Experiments and Results

This study evaluated and compared the performance of three CNN models with an LSTM model for predicting monthly peak demand. The experiments were conducted using the Dell Inspiron 15 7000 laptop with Intel® Core™ i7-8565U CPU@1.80 GHz, 64-bit Windows 10 operating system, and 8 GB memory. All the models were built in Python 3.7.6 and Keras with Tensorflow as backend. In order to build the CNN models, part of the source code was accessed from Machine Learning Mastery [29].

The entire dataset was divided into a training set (80%) and test set (20%). We first proceeded to normalize the training data between 0 and 1 values using the *min-max* scaler from Python. Then, the test set was normalized according to the training normalization parameters. This process was applied to all the models. In order to train the model, the number of timesteps for predicting the peak demand had to be defined. Therefore, the training set was transformed into a set of input and output sequence vectors. Each model was trained for 1000 epochs. The models were optimized using the Adam optimizer with the default learning rate of 0.001. The MSE loss function was used to train the models and the batch size was set to 10. The results of each model are provided below.

8.1. CNN Models Performance

8.1.1. Multivariate CNN

A total of 27 multivariate CNN models with different input feature combinations were built for predicting the peak demand. Table 8 shows the number of input features and timesteps used to build each model, as well as the model performance on the test dataset.

Table 8. Performance of the different multivariate CNN models proposed.

Model No.	No. of Features	No. of Timesteps	Kernel Size	Test Dataset			
				R ²	MSE	MAPE (%)	MAE
1	5	3	2 × 1	0.81	2893.71	2.44	42.27
2		4	3 × 1	0.72	4279.46	3.06	52.35
3		5	4 × 1	0.59	6149.42	3.70	63.76
4		6	5 × 1	0.69	4677.33	3.43	59.21
5		8	7 × 1	0.80	3167.02	2.88	48.55
6		10	9 × 1	0.81	3184.28	2.46	43.27
7		12	10 × 1	0.81	3218.95	2.82	48.85

Table 8. Cont.

Model No.	No. of Features	No. of Timesteps	Kernel Size	Test Dataset			
				R ²	MSE	MAPE (%)	MAE
8	4	3	2 × 1	0.92	1271.65	1.62	27.86
9		4	2 × 1	0.30	10,583.14	4.66	80.48
10		5	3 × 1	−0.47	22,447.11	6.04	107.59
11		6	5 × 1	−0.02	15,833.36	4.98	89.12
12		8	7 × 1	0.45	8920.29	4.67	80.41
13		10	7 × 1	0.87	2199.62	2.40	40.88
14		12	10 × 1	0.75	4208.58	3.49	59.38
15	6	3	2 × 1	0.61	6022.84	3.40	59.43
16		4	3 × 1	0.73	4140.59	2.91	50.51
17		5	3 × 1	0.48	7837.17	4.26	73.01
18		6	5 × 1	0.70	4642.65	3.00	52.19
19		8	7 × 1	0.69	5052.66	3.60	60.75
20		10	9 × 1	0.52	7803.71	4.36	76.33
21		12	11 × 1	0.60	6603.69	4.36	73.65
22	5	3	2 × 1	0.62	5922.29	3.80	63.84
23		5	4 × 1	0.21	12,087.59	4.79	84.59
24		6	5 × 1	0.61	5938.94	3.51	60.66
25		8	7 × 1	0.72	4531.35	2.91	51.75
26		10	6 × 1	0.64	5895.48	3.97	67.69
27		12	11 × 1	0.82	3041.57	2.89	49.82

Model 1 thru Model 7 used five input features: month, residential consumption, big clients consumption, monthly economic index, and previous peak demand. Model 8 thru Model 14 used four input features: month, residential consumption, big clients consumption, and previous peak demand. Model 15 thru Model 21 incorporated six input features: month, residential consumption, big clients consumption, minimum temperature, maximum temperature, and previous peak demand. Model 22 thru Model 27 included five input features: month, residential consumption, big clients consumption, maximum temperature, and previous peak demand. Different number of timesteps were evaluated to determine significant lags for predicting the next month peak demand.

Based on Table 8, it can be observed that Model 8 had the best performance with R-squared value of 0.92, MSE of 1271.65, MAPE of 1.62%, and MAE of 27.86. This model received four input features consisting of three previous timesteps for predicting the month ahead peak demand. These features are discussed above. It was also evident that the performance of the models worsened when the minimum and maximum temperature were included.

Figure 12 shows how the multivariate CNN model performed on the test dataset. In overall, the model was able to follow the peak demand pattern. The model performed fairly well between the time period January 2019 to December 2019, in which the peak demand increased drastically. However, it could not stably predict the highest peak points which occurred in August and December 2019. For the month of December, the model underestimated demand by 118 MW.

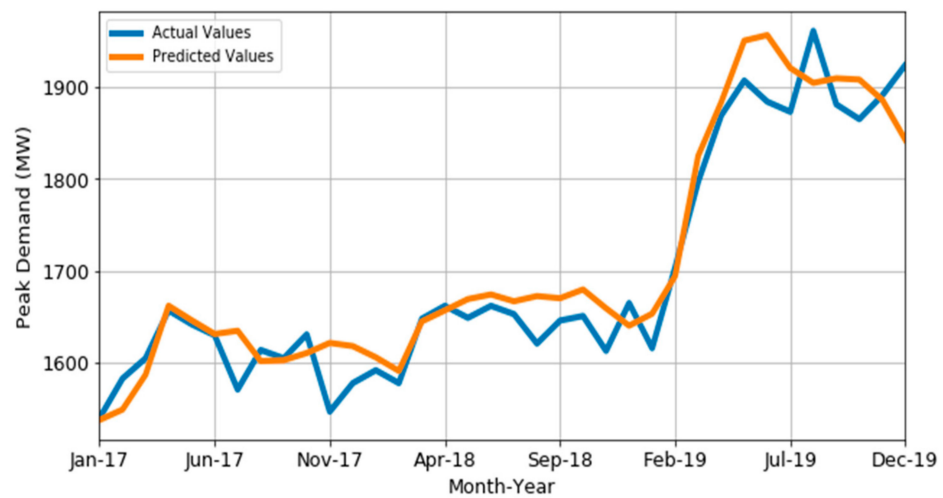


Figure 12. Performance of the multivariate CNN model on the test set.

8.1.2. Multivariate CNN-LSTM

The multivariate CNN-LSTM model performed poorly compared to the multivariate CNN. The R-squared value was 0.50, the MSE was 7731.80, the MAPE was 3.76%, and MAE was 65.54. The results are provided in Table 9. Figure 13 demonstrates the performance of the model on the test dataset. It can be observed that the model overpredicted the demand by far, especially for the time period July 2019 to December 2019. In overall, the CNN-LSTM model did not provide stable performance.

Table 9. Results of the multivariate CNN-LSTM model on the test set.

Model	No. of Features	No. of Timesteps	Test Dataset			
			R ²	MSE	MAPE (%)	MAE
CNN-LSTM	4	3	0.50	7731.80	3.76	65.54

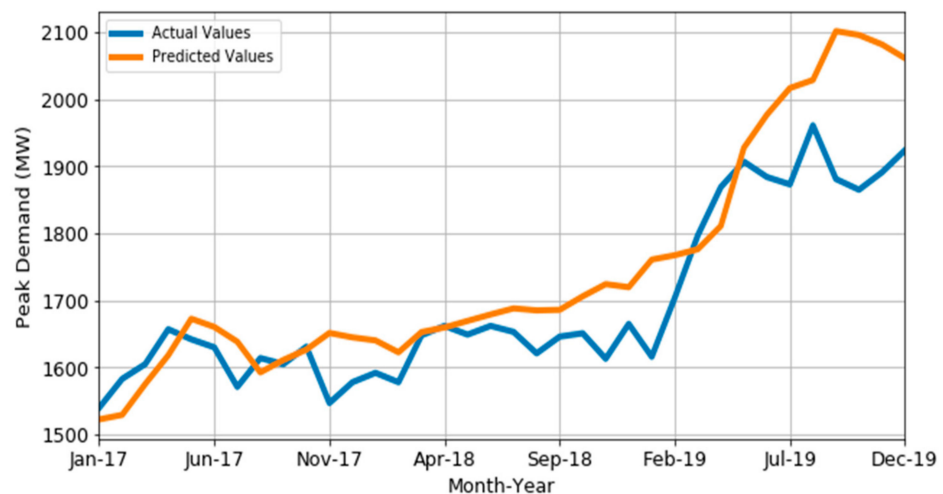


Figure 13. Performance of the multivariate CNN-LSTM model on the test set.

8.1.3. Multihead CNN

The multihead CNN performed much better than the multivariate CNN-LSTM but was still inferior to the multivariate CNN model. The R-squared value was 0.81, the MSE was 2940.98, the MAPE was 2.30%, and MAE was 40.17. The results are summarized in Table 10. Although the model was able to follow the peak demand pattern, it underesti-

mated the demand for many months and was not able to predict the highest peak occurred in August 2019 (Figure 14).

Table 10. Results of the multihead CNN model on the test set.

Model	No. of Features	No. of Timesteps	Test Dataset			
			R ²	MSE	MAPE (%)	MAE
Multihead CNN	4	3	0.81	2940.98	2.30	40.17

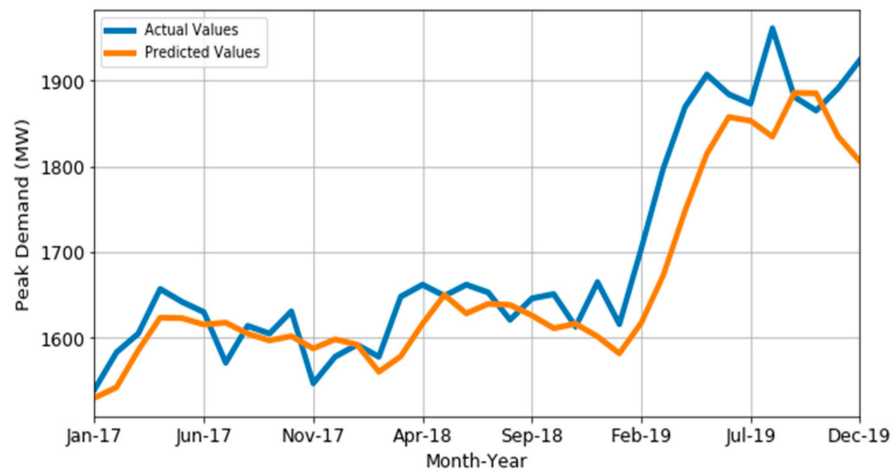


Figure 14. Performance of the multihead CNN model on the test set.

8.2. LSTM Performance

The LSTM model performed the worse. The R-squared value was -0.52 , the MSE was 23,654.8, the MAPE was 5.7%, and MAE was 103.27. The results are provided in Table 11. When evaluated on the test set, the model was not able to predict the upward peak demand trend across the period January 2019 to December 2019 (Figure 15).

Table 11. Results of the LSTM model on the test set.

Model	Number of Features	Number of Timesteps	Test Dataset			
			R ²	MSE	MAPE (%)	MAE
LSTM	4	3	-0.52	23654.8	5.7	103.27

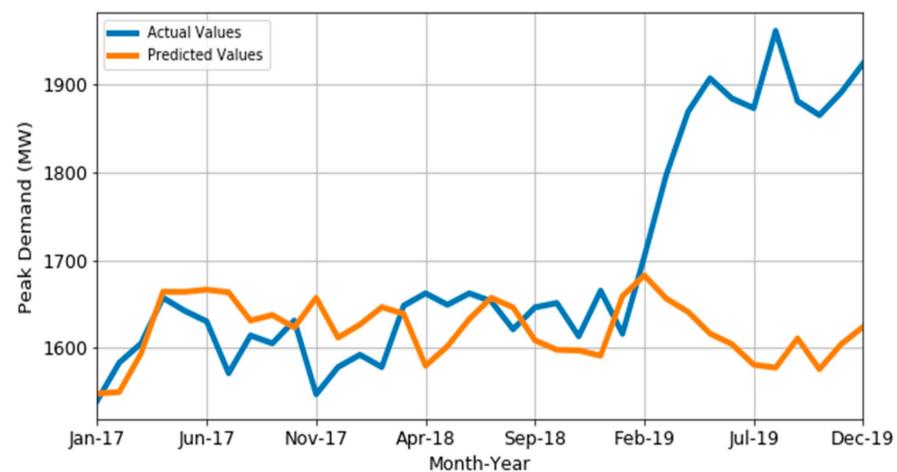


Figure 15. Performance of the LSTM model on the test set.

8.3. Comparison of Model Performance

This section provides a general assessment of each model. The three CNN models and the LSTM model are compared in terms of computational speed and four performance metrics, which are provided in Table 12. The multivariate CNN performed the best out of all the CNNs and was superior to the LSTM. The model provided an R-squared value of 0.92, MSE of 1271.65, MAPE of 1.62%, and MAE of 27.86. The LSTM had poor performance, although it was trained for 1000 epochs. The multivariate CNN model also had the fastest processing time of 15.85 s, while the LSTM required more time of 88.08 s. This was due to the fact that the CNN had a smaller number of 1989 trainable parameters while the LSTM had 38801 parameters. Figure 16 provides an overall comparison of the models on the test set.

Table 12. Comparison of CNNs and LSTM in terms of training time and performance measures.

Model	No. of Trainable Parameters	Training Time (s)	Test Dataset			
			R ²	MSE	MAPE (%)	MAE
Multivariate CNN	1989	15.85	0.92	1271.65	1.62	27.86
Multivariate CNN-LSTM	19,489	34.81	0.50	7731.80	3.76	65.54
Multihead CNN	6885	19.63	0.81	2940.98	2.30	40.17
LSTM	38,801	88.08	−0.52	23,654.8	5.7	103.27

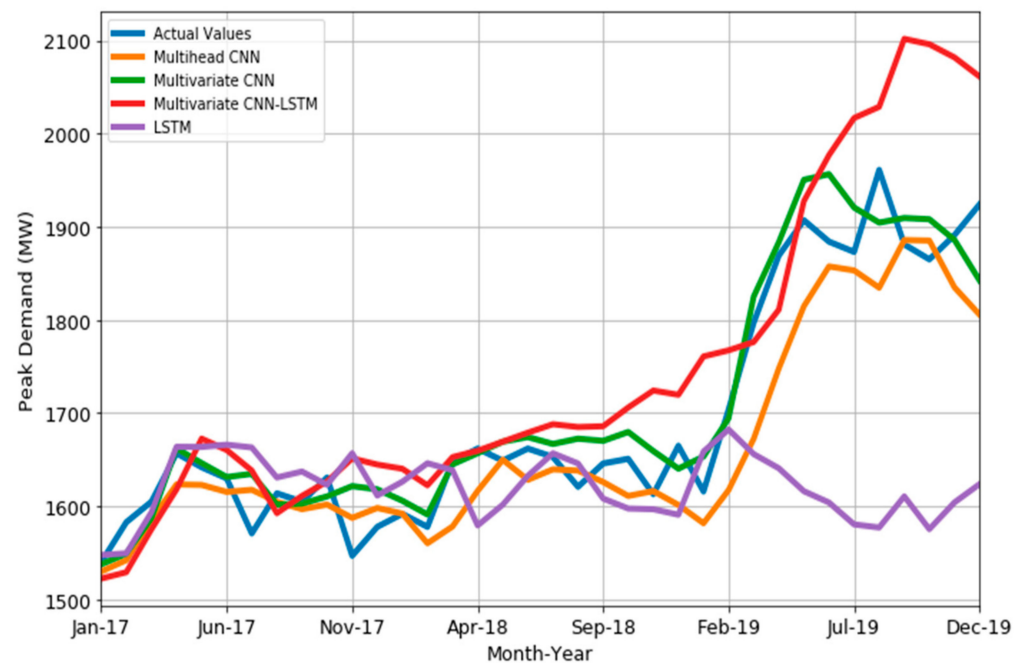


Figure 16. Comparison of CNNs and LSTM on the test set.

8.4. Model Robustness

The next step of this study consisted in evaluating the robustness of the CNNs and LSTM models against noise. It is well known that the effectiveness of deep learning networks depends largely on the quality of the data. According to the literature, one problem that can arise from training deep networks on noisy small-scale datasets is overfitting. Since the network is trained on limited data, it will be constrained to learn only certain types of patterns. Therefore, it will tend to memorize the high frequency signals that are not relevant [30]. As a result, this will cause the model to overfit and not generalize well on

the test set. There are many strategies that have been applied to reduce overfitting such as adding noise to the training data, the dropout method and weight regularization. A recent study found that when adversarial noise was carefully added to the CNN, the network was able to perform smoother mapping and generalize better [31].

Training neural networks with Gaussian noise has also attracted considerable attention. The Gaussian noise, consisting of the same frequency distribution, is intentionally added during training to perturb the input signals, which makes the learning process more robust. For this study, the Gaussian noise with standard distribution of 0.3 was added to the training data as an intermediate layer. The models were run five times to provide a fair comparison and the results are provided in Table 13. The R^2 was used to measure model robustness. It is clearly evident that the CNNs demonstrated to be more robust than the LSTM. The multivariate CNN and multihead CNN had similar performance, with average R^2 value of 0.90 and 0.89 respectively. The LSTM still performed poorly and was not found to be robust. In overall, the models generalized well on the test set.

Table 13. Evaluation of CNNs and LSTM robustness against Gaussian noise.

Method	Performance Metrics	No. of Simulations					Average
		1	2	3	4	5	
Multivariate CNN	R^2	0.89	0.89	0.90	0.90	0.91	0.90
	MSE	1677.16	1764.12	1528.08	1480.61	1455.50	1581.09
	MAPE (%)	2.03	2.05	1.96	1.86	1.88	1.95
	MAE	34.60	34.84	33.38	31.76	32.04	33.32
Multivariate CNN-LSTM	R^2	0.86	0.86	0.88	0.87	0.84	0.86
	MSE	2199.99	2218.77	1710.25	1958.56	2433.16	2104.15
	MAPE (%)	1.94	1.99	1.96	1.84	2.05	1.95
	MAE	34.086	34.81	33.44	32.236	35.99	34.11
Multihead CNN	R^2	0.89	0.90	0.90	0.89	0.89	0.89
	MSE	1628.03	1609.41	1603.30	1702.54	1657.71	1640.20
	MAPE (%)	1.91	1.87	1.91	1.92	1.87	1.90
	MAE	32.598	31.98	32.65	33.05	32.19	32.49
LSTM	R^2	-0.22	0.46	0.67	-0.49	-0.45	-0.01
	MSE	18,922.44	8346.65	5155.88	23,116.06	22,405.75	15,589.36
	MAPE (%)	5.73	4.17	3.21	5.44	6.02	4.91
	MAE	101.28	72.08	56.32	99.15	106.57	87.08

9. Discussion

This study proposed a deep learning framework to predict monthly peak demand of Panama. Therefore, we compared the performance of two deep learning methods, CNNs and LSTM, to address this problem. First, a dataset on monthly peak demand was collected for the time period 2004 to 2019. It was clearly evident that the data exhibited highly nonlinear and noisy patterns. We then investigated a list of input features that could account for the upward nonlinear peak demand trend (Table 2). The objective of this study was to implement 1D CNNs and LSTM to learn the historical patterns in the data and use this trend to project the future peak demand growth for Panama.

We first introduced a CNN approach consisting of three different architectures which were multivariate CNN, multivariate CNN-LSTM, and multihead CNN. The CNNs were then benchmarked against the LSTM, a current state of the art method for time series forecasting. Through the multivariate CNN, we found that the input sequence of three previous timesteps consisting of four input features were significant for predicting the month ahead peak demand. These features were month, big clients consumption, residential consumption, and the previous peak demand. Therefore, we used the same features

and time lag to build the rest of the models. All the models were evaluated on a test set from the time period October 2016 to December 2019 that was very unstable. The model assessment was conducted using two steps: with and without noise.

First, the models were built and trained without noise to determine how well they generalized on the test set. Our initial finding was that the CNNs performed far better than the LSTM. The multivariate CNN outperformed all the models with R^2 of 0.92, MSE of 1271.65, MAPE of 1.62%, and MAE of 27.86. The feature extraction process was guided by a convolutional layer that consisted of 32 filters (2×1), followed by a max pooling layer (2×1). The model followed the peak demand pattern fairly well, but it underestimated the highest peaks that occurred in August and December 2019. Therefore, the features used were not able to fully capture these variations. The year 2019, especially, presented a drastic increase in electricity demand, which raised concerns. Some experts believe that the addition of new projects such as the Panama Metro Line as well as the growth of electricity clients are contributing towards the high demand. Therefore, it is important to consider other features that can account for this growth. One of the limitations of this study was that the data on monthly number of clients was only available from March 2013 and onwards. Therefore, we were not able to incorporate this feature into the models.

With respect to the other models proposed, the multivariate CNN-LSTM overestimated the peaks by far. LSTM consistently struggled to follow the peak demand pattern between the period February 2019 and December 2019, underestimating the demand.

In order to validate our initial assumptions about the performance of the models, we decided to evaluate the robustness of the models against noise. Therefore, the Gaussian noise of 30% intensity level was carefully introduced to the training set. After simulating the models five times, we were able to confirm that CNNs were far more robust than the LSTM. The reason behind this could be that the data exhibited spatial patterns, which the CNN was able to effectively extract. LSTM, on the other side, is more suitable for sequential learning. The multivariate CNN and multihead CNN had similar performance, with average R^2 value of 0.9 and 0.89, respectively. The multivariate CNN-LSTM performance was slightly inferior with average R^2 of 0.86. We also demonstrated that the CNNs required less training time with an average of 23.4 s for the three models.

We benchmarked our best performing model with respect to the proposed models in the literature. For monthly forecast, the MAPE ranged between 0.07% and 3.94% (Table 1). Therefore, we are confident that our model achieved state-of-the-art performance, with MAPE of 1.62%. Furthermore, while many authors found LSTM to be powerful, this study demonstrated that LSTM was not effective for this particular problem and it performed better when it was combined in a hybrid CNN-LSTM architecture.

10. Conclusions and Future Work

As modern power systems continue to face increasing peak demand due to many factors such as larger consumer loads, rapid acceleration of EVs and weather variations, the need for more accurate forecasting methods has never been greater. The peak demand exhibits a nonlinear and noisy trend, making it a very complex problem that needs to be attended faster. Deep learning is emerging with better performance to support these areas that are very nonlinear, noisy, and present diverse patterns. CNNs, especially, have received increased attention from the computer vision community, due to their outstanding performance in image classification and object recognition. This trend has favored their implementation in other areas such as time series forecasting for leveraging spatial-temporal data. LSTM, on the other side, has quickly become the state-of-the-art method for time series applications due to their capability for processing sequential information.

This paper introduced two deep learning methods, based on 1D CNNs and LSTM, to predict the monthly peak demand for Panama.

The main finding of this study was that CNNs achieved superior performance in both accuracy and robustness compared to LSTM, with short computational time. The reason behind this was that CNNs were able to extract and learn the complex patterns from the

input features. Therefore, the implications of this study are significant since we proved that CNNs can effectively support time series forecasting problems, especially when the data exhibits spatial patterns.

The multivariate CNN was the best performing model that provided state-of-the-art accuracy with R^2 of 0.92, MSE of 1271.65, MAPE of 1.62%, and MAE of 27.86. Although many studies have demonstrated LSTM to be powerful for predicting the peak demand pattern, we did not find it to be effective for this particular problem. However, LSTM performed better when it was combined in a hybrid CNN-LSTM structure, to decode the sequential patterns of the features extracted by the CNN.

The results of the multivariate CNN model are favorable for predicting the peak demand. However, we believe that there is still room for improvement. Our future work has several directions. First, we propose to improve the model by incorporating other variables that can better account for the highest peaks that occurred in the months of August and December 2019. This study suggests adding variables such as CDD and monthly number of tourists traveling to Panama. Many models have incorporated CDD to better capture the effect of temperature on peak demand. It is also important to investigate other sectors that could be contributing to the peak demand growth, such as the Panama Metro, which started operation in April 2019. Another interesting direction for this study will be to benchmark these models against ensemble machine learning methods such as gradient boosting and ANN-SVR, which are being more widely adopted in the literature to enhance forecasting accuracy. Lastly, we will consider CNNs to support other areas of the energy sector, such as short-term load forecasting (STLF) and renewable prediction which are becoming more prominent. STLF, in particular, has become an active area of research to help improve load smoothing and peak demand shaving of buildings.

Author Contributions: Conceptualization, L.R.; methodology, B.I. and L.R.; software, B.I.; validation, B.I. and L.R.; formal analysis, B.I. and L.R.; investigation, B.I.; resources, B.I.; data curation, B.I.; writing—original draft preparation, B.I.; writing—review and editing, B.I. and L.R.; visualization, L.R.; supervision, L.R.; project administration, L.R.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. Data on monthly peak demand can be found in the National Dispatch Center website [28].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Burillo, D.; Chester, M.V.; Pincetl, S.; Fournier, E.D.; Reyna, J. Forecasting peak electricity demand for Los Angeles considering higher air temperatures due to climate change. *Appl. Energy* **2019**, *236*, 1–9. [[CrossRef](#)]
2. Carvallo, J.P.; Larsen, P.H.; Sanstad, A.H.; Goldman, C.A. Long term load forecasting accuracy in electric utility integrated resource planning. *Energy Policy* **2018**, *119*, 410–422. [[CrossRef](#)]
3. Rallapalli, S.R.; Ghosh, S. Forecasting monthly peak demand of electricity in India—A critique. *Energy Policy* **2012**, *45*, 516–520. [[CrossRef](#)]
4. Sigauke, C.; Chikobvu, D. Peak electricity demand forecasting using time series regression models: An application to South African data. *J. Stat. Manag. Syst.* **2016**, *19*, 567–586. [[CrossRef](#)]
5. Khorsheed, E. Long-Term Energy Peak Load Forecasting Models: A Hybrid Statistical Approach. In Proceedings of the 2018 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 6 February–5 April 2018; pp. 1–6.
6. Al-Hamad, M.Y.; Qamber, I.S. GCC electrical long-term peak load forecasting modeling using ANFIS and MLR methods. *Arab J. Basic Appl. Sci.* **2019**, *26*, 269–282. [[CrossRef](#)]
7. Almazrouee, A.I.; Almeshal, A.M.; Almutairi, A.S.; Alenezi, M.R.; Alhajeri, S.N. Long-term forecasting of electrical loads in kuwait using prophet and holt-winters models. *Appl. Sci.* **2020**, *10*, 5627. [[CrossRef](#)]

8. Pannakkong, W.; Aswanuwath, L.; Buddhakulsomsiri, J.; Jeenanunta, C.; Parthanadee, P. Forecasting medium-term electricity demand in Thailand: Comparison of ANN, SVM, DBN, and their ensembles. In Proceedings of the 17th International Conference on ICT and Knowledge Engineering, Bangkok, Thailand, 20–22 November 2019; pp. 1–6. [CrossRef]
9. Kwon, B.-S.; Park, R.-J.; Song, K.-B. Weekly Peak Load Forecasting for 104 Weeks Using Deep Learning Algorithm. In Proceedings of the 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Macao, China, 1–4 December 2019; pp. 1–4.
10. Elkamel, M.; Schleider, L.; Pasiliao, E.L.; Diabat, A.; Zheng, Q.P. Long-term electricity demand prediction via socioeconomic factors—A machine learning approach with florida as a case study. *Energies* **2020**, *13*, 3996. [CrossRef]
11. Son, H.; Kim, C. A deep learning approach to forecasting monthly demand for residential-sector electricity. *Sustainability* **2020**, *12*, 3103. [CrossRef]
12. Kim, D.-H.; Lee, E.-K.; Qureshi, N.B.S. Peak-load forecasting for small industries: A machine learning approach. *Sustainability* **2020**, *12*, 6539. [CrossRef]
13. Lai, C.S.; Mo, Z.; Wang, T.; Yuan, H.; Ng, W.W.Y.; Lai, L.L. Load forecasting based on deep neural network and historical data augmentation. *IET Gen. Trans. Dist.* **2020**, *14*, 5927–5934. [CrossRef]
14. Mehdipour Pirbazari, A.; Farmanbar, M.; Chakravorty, A.; Rong, C. Short-term load forecasting using smart meter data: A generalization analysis. *Processes* **2020**, *8*, 484. [CrossRef]
15. Shirzadi, N.; Nizami, A.; Khazen, M.; Nik-Bakht, M. Medium-term regional electricity load forecasting through machine learning and deep learning. *Designs* **2021**, *5*, 27. [CrossRef]
16. Atef, S.; Eltawil, A.B. Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. *Electr. Power Syst. Res.* **2020**, *187*. [CrossRef]
17. Laouafi, A.; Mordjaoui, M.; Laouafi, F.; Boukelia, T.E. Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology. *Int. J. Electr. Power Energy Syst.* **2016**, *77*, 136–144. [CrossRef]
18. Dai, S.; Niu, D.; Li, Y. Daily peak load forecasting based on complete ensemble empirical mode decomposition with adaptive noise and support vector machine optimized by modified grey wolf optimization algorithm. *Energies* **2018**, *11*, 163. [CrossRef]
19. Park, M.; Kim, J.; Won, D.; Kim, J. Development of a Two-Stage ESS-Scheduling model for cost minimization using machine learning-based load prediction techniques. *Processes* **2019**, *7*, 370. [CrossRef]
20. Nepal, B.; Yamaha, M. A Hybrid Model of Clustering and Neural Network Using Weather Conditions for Energy Management in Buildings. In Proceedings of the 4th International Conference on Compute and Data Analysis, San Jose, CA, USA, 9–12 March 2020; pp. 89–96.
21. Kazemzadeh, M.-R.; Amjadian, A.; Amraee, T. A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. *Energy* **2020**, *204*. [CrossRef]
22. Wu, L.; Kong, C.; Hao, X.; Chen, W. A short-term load forecasting method based on gru-cnn hybrid neural network model. *Math. Probl. Eng.* **2020**, *2020*. [CrossRef]
23. Khan, S.; Rahmani, H.; Shah, S.A.A.; Bennamoun, M. Convolutional Neural Network. In *A Guide to Convolutional Neural Networks for Computer Vision*, 1st ed.; Medioni, G., Dickinson, S., Eds.; Morgan & Claypool: San Rafael, CA, USA, 2018; Volume 8, pp. 43–68.
24. Kamath, U.; Liu, J.; Whitaker, J. Convolutional Neural Networks. In *Deep Learning for NLP and Speech Recognition*, 1st ed.; Drougas, P., Ed.; Springer: Cham, Switzerland, 2019; Volume 84, pp. 263–314.
25. Sadaei, H.J.; de Lima e Silva, P.C.; Guimarães, F.G.; Lee, M.H. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy* **2019**, *175*, 365–377. [CrossRef]
26. Lara-Benitez, P.; Carranza-García, M.; Luna-Romera, J.M.; Riquelme, J.C. Temporal convolutional networks applied to energy-related time series forecasting. *Appl. Sci.* **2020**, *10*, 2322. [CrossRef]
27. Tovar, M.; Robles, M.; Rashid, F. PV power prediction, using CNN-LSTM hybrid neural network model. Case of study: Temixco-morelos, México. *Energies* **2020**, *13*, 6512. [CrossRef]
28. Centro Nacional de Despacho. Available online: <https://www.cnd.com.pa/index.php/estadisticas> (accessed on 1 April 2021).
29. Brownlee, J. How to Develop Convolutional Neural Network Models for Time Series Forecasting. Available online: <https://machinelearningmastery.com/how-to-develop-convolutional-neural-network-models-for-time-series-forecasting/> (accessed on 25 February 2021).
30. Choubey, S.; Choubey, A.; Vishwakarma, A.; Dwivedi, P.; Vishwakarma, A.; Seth, A. Identification of face along with configuration beneath unobstructed ambient via reflective deep cascaded neural networks. In *Cognitive Informatics, Computer Modelling, and Cognitive Science*, 1st ed.; Sinha, G.R., Suri, J.S., Eds.; Academic Press: Cambridge, MA, USA, 2020; Volume 1, pp. 177–194.
31. You, Z.; Ye, J.; Li, K.; Xu, Z.; Wang, P. Adversarial Noise Layer: Regularize Neural Network by Adding Noise. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 909–913.