energies

MDPI

Article

Multihousehold Load Forecasting Based on a Convolutional Neural Network Using Moment Information and Data Augmentation

Shree Krishna Acharya [1], Hwanuk Yu [2], Young-Min Wi [3] and Jaehee Lee [4,*]

[1] School of Business, University College Dublin, Dublin A94XF34, Ireland; shree.acharya@ucd.ie
[2] School of Electrical Engineering, Korea University, Seoul 02841, Republic of Korea; yhu6044@korea.ac.kr
[3] Department of Electrical Engineering, Sangmyung University, Seoul 03016, Republic of Korea; youngmin@smu.ac.kr
[4] Department of Electrical and Control Engineering, Mokpo National University, Muan 58554, Republic of Korea
* Correspondence: jaehee@mnu.ac.kr

Abstract: Deep learning (DL) networks are a popular choice for short-term load forecasting (STLF) in the residential sector. Hybrid DL methodologies based on convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) have a higher forecasting accuracy than conventional statistical STLF techniques for different types of single-household load series. However, existing load forecasting methodologies are often inefficient when a high load demand persists for a few hours in a day. Peak load consumption is explicitly depicted as a tail in the probability distribution function (PDF) of the load series. Due to the diverse and uncertain nature of peak load demands, DL methodologies have difficulty maintaining consistent forecasting results, particularly when the PDF of the load series has a longer tail. This paper proposes a multihousehold load forecasting strategy based on the collective moment measure (CMM) (which is obtained from the PDF of the load series), data augmentation, and a CNN. Each load series was compared and ordered through CMM indexing, which helped maintain a minimum or constant shifting variance in the dataset inputted to the CNN. Data augmentation was used to enlarge the input dataset and solve the existing data requirement issues of the CNN. With the ordered load series and data augmentation strategy, the simulation results demonstrated a significant improvement in the performance of both single-household and multihousehold load forecasting. The proposed method predicts day-ahead multihousehold loads simultaneously and compares the results based on a single household. The forecasting performance of the proposed method for six different household groups with 10, 20, 30, 50, 80, and 100 household load series was evaluated and compared with those of existing methodologies. The mean absolute percentage error of the prediction results for each multihousehold load series could be improved by more than 3%. This study can help advance the application of DL methods for household load prediction under high-load-demand conditions.

Keywords: multihousehold load forecasting; collective moment measure (CMM); convolutional neural network (CNN); data augmentation; shifting variance

Citation: Acharya, S.K.; Yu, H.; Wi, Y.-M.; Lee, J. Multihousehold Load Forecasting Based on a Convolutional Neural Network Using Moment Information and Data Augmentation. *Energies* **2024**, *17*, 902. https://doi.org/10.3390/en17040902

Academic Editors: Majdi Mansouri, Mohamed Trabelsi and Sertac Bayhan

Received: 18 January 2024
Revised: 11 February 2024
Accepted: 13 February 2024
Published: 15 February 2024

1. Introduction

Short-term load forecasting (STLF) facilitates the operation of power systems as an economical, secure, and reliable program. It is equally important in the residential sector. Recently, STLF methodologies based on statistical, machine learning, deep learning (DL), causal, econometric, and judgment-based algorithms have been developed [1]. For single households, these techniques must deal with granular load series collected from advanced metering infrastructure (AMI). Each series is composed of uncertain load consumption

values, which have a high degree of stochasticity, making them difficult to define using a physical measure [2–4].

Although people today live and spend time in grouped areas with distinct lifestyles, such as in apartment blocks, department stores, and companies, each household produces a unique load consumption series. The major exogenous factors that bring differences in household load series include the use of automatic thermal devices, installation of electrical vehicles (EVs), penetration of renewable energy resources, and evolution of social demographics. The impact of these factors is expected to increase in the future, and the corresponding load series will become more complex for analysis and prediction. Hence, a new approach that addresses two main issues should be developed: (1) The vast number of customer load series at any given time, and (2) the increasing amount of uncertainty in single-household load series.

The load series from a single household has properties similar to nonstationary signals and has been mainly analyzed based on zero-mean, trend, and seasonal modeling [1]. Many statistical approaches can be used to extract hidden information from unusual load consumptions with the help of consecutive load observations during peak load demand hours [5–7]. Relevant customer load series has been explored using a Gaussian process for forecasting the load profiles of target households [8]. With the help of the correlation and L2 norm-based input selection methods, an appropriate sub-load series was extracted for each type of special day [9]. Clustering- and classification-based forecasting methodologies reduce the uncertainty level by batching homogenous customers, days, or weather parameters [9–13]. DL methodologies, such as support vector machine (SVM) [14], multilinear perceptron (MLP) [15], convolutional neural networks (CNNs) [16–18], and recurrent neural networks (RNNs) [19], have been used to forecast single-household loads. Pooling [16] and data augmentation [17] are popular algorithms for preparing a training dataset for the CNN from historical load series. A multiple household load forecasting model using Bayesian networks has been presented with various input features such as past consumption, temperature, and socioeconomic and electricity usage aspects [20]. A forecasting model for multiple individual households' short-term electricity consumption has also been developed using two-level information extraction stages [21], where input datasets are structured in the low-level stage and a combined LSTM and CNN-based forecasting model is developed in the high-level stage. However, existing methodologies are associated with challenges in precisely batching and exploiting the diverse load series, and too weakly structured a dataset to achieve a high forecasting accuracy.

Advanced CNN-based architectures have been developed for image recognition, medical imaging, and natural language processing applications [22]. ImageNet [23] and VGGNet [24] are well-known CNN architectures in which a number of images can be taken as input for classifying multiple image-like objects. These architectures have been mainly developed for data classification problems that are closer to multi-series forecasting than to single-series forecasting. Many load series are recorded in ordered 1D architectures; however, CNN-based architectures operate by accepting three- or higher-dimensional datasets. Few have studied the local correlations between multiple time-series data. Convoluting multi-nonstationary signals can produce a better performance by exploiting the local minima, correlations, and dependencies [25]. Multisite PV profiles are predicted by developing a space–time matrix using a greedy adjoining algorithm [26]. For better accuracy, a large dataset is required for training with a minimum or constant shifting variance [27,28]. Existing pooling and augmentation techniques are popular for enlarging datasets and producing better forecasting results than their peers. However, the pooling strategy [16] has difficulty in delivering better forecasting results because of the large number of complex load series. Complex load series contain an abundance of tail information in their probability distribution functions (PDFs) because there may be higher load demand for a few hours in any given day. The data augmentation technique [17] has difficulty enlarging the appropriate dataset because of its shortcomings in generating a large number of similar residual load series. This makes it difficult to achieve a high forecasting accuracy for a single household;

improving the performance continues to be a prerequisite. Hence, exploring multiple household load series for simultaneous load prediction is another approach to improve the forecasting accuracy of existing models.

This paper proposes a novel day-ahead multihousehold forecasting method based on a CNN architecture in which forecasting results are obtained simultaneously. The proposed method was tested on six different household groups containing 10, 20, 30, 50, 80, and 100 households with different load series from the database. Each load series was identified as unique by evaluating its collective moment measure (CMM). The CMM quantifies the hidden integrated statistics of the load series using moment parameter information, such as the mean and variance. The quantitative measurement of the moment parameters assists in learning the daily few-hour peak load from the load series of the household groups. The load series discovered based on the CMM index are more likely and placed in order for data augmentation. The data augmentation methodology for the ordered multihousehold load series maintains a minor or minimum shifting variance in the input dataset for the CNN. Although high load-demand hours are present each day in each load series, CMM-based ordering balances the entire uncertain peak load consumption, and the CNN can learn to predict accurate day-ahead load profiles for all input households. The simulation results showed a significant improvement of more than 8% and 3% in the forecasting accuracy in terms of the mean absolute percentage error (MAPE) over existing CNN methods with pooling [16] and augmentation [17], respectively.

The remainder of this paper is organized as follows. Section 2 discusses multihousehold load forecasting, the significance of the minimum shifting variance in the CNN, and reordering load series based on the tail PDF. Section 3 describes the estimation of the tail length, the data augmentation methodology for single-household load series, and the procedure of the proposed multihousehold forecasting algorithm. Section 4 presents the obtained multihousehold load forecasting simulation results and comparisons. Section 5 discuss the potential results and limitations of the paper. Finally, Section 6 concludes the paper.

2. Problem Description

2.1. Multihousehold Load Forecasting Based on CNN

The multihousehold load forecasting methodology predicts the number of day-ahead load profiles of different households simultaneously by sharing load-consuming characteristics from the nearest neighborhood. Many machine- and DL-based forecasting strategies, such as pooling [16] and k-nearest neighbor [29], can perform forecasting operations for single-household load series using the load-series data of other households. In comparison, a data augmentation strategy can predict all the household load series with a higher forecasting accuracy. Figure 1 shows the overall procedure of the proposed CNN-based multihousehold load forecasting methodology. The data augmentation method meets the data requirement of the CNN for multihousehold load forecasting, even though each household has a peak load consumption for a few hours in its daily load profiles.

2.2. Effectiveness of CNN with Low-Shifting Data

Shifting invariance is key to realizing a high prediction accuracy when using a CNN [30,31] and is useful for forecasting multihousehold load series. Shifting invariance means maintaining unbiased learning for small displacements in the input dataset [32], where the convolutional architecture is robust to learning the peak load difference in the input dataset. However, the CNN forecasting results vary significantly under high and low peak loads because of parameter sharing between the convolution layers and the partial effects from the pooling layers [28]. Hence, variance shifting should be minimized or set as a constant in the input dataset to achieve a high forecasting accuracy when using a CNN. To acknowledge the low shifting and evaluate the potential significant difference

in the input series $L = [l_{1,1}, l_{1,2}, l_{1,3}, \ldots\ldots l_{d,1}, \ldots l_{d,t}, \ldots l_{d,T}, \ldots l_{D,1} \ldots l_{D,T}]$, the average shifting variance $\bar{\delta}_{sf}$ is computed from n shifting variance δ_{sf}, given by:

$$\delta_{sf}(n) = \frac{1}{F} \sum_{i=n-F+1}^{n} \left(l_i - \bar{l}_F\right)^2 \quad (1)$$

$$\bar{\delta}_{sf} = \frac{1}{D \times T - F} \sum_{n=F}^{D \times T} \delta(n) \quad (2)$$

where l_i is the hourly electricity consumption with domain of $i = T \cdot (d-1) + t$, T represents the total hours in a day, t is the hour of the day d, D is the number of days, F is the specified length of the window series in l_d to calculate the variance, and $\bar{l}_F$ is the mean load consumption of the window series.

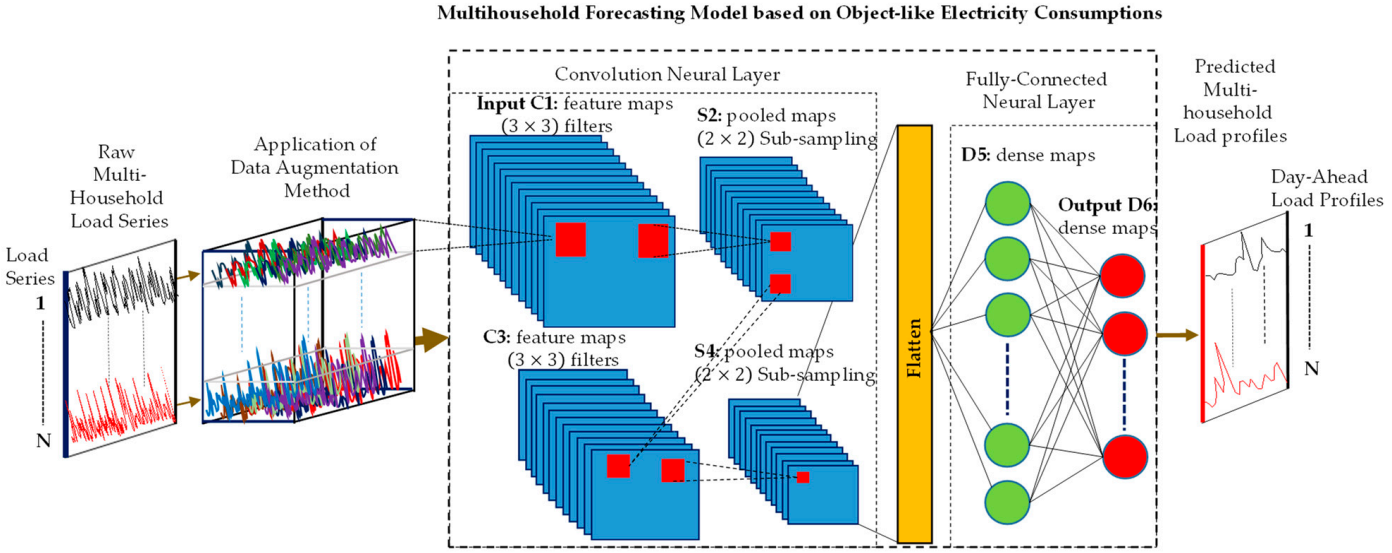

Figure 1. CNN-based multihousehold load forecasting methodology using data augmentation.

2.3. Reordering Load Series Based on Tail Length for Reducing Shifting Variance

The tail length of a load series of a single household can be explicitly observed when there is unusual load consumption. All the load series exhibited either long or short tails in their PDFs [33,34]. Figure 2 shows two different load series with their corresponding tail characteristics. Figure 2a,b shows the short-tail load series and the corresponding PDF; Figure 2c,d shows the long-tail load series and corresponding PDF, respectively. The main reason for the occurrence of either short- or long-tail PDFs in the load series is the high demand for a few hours on a daily basis.

For example, air conditioners, heaters, ovens, washing machines, and microwaves are higher load-consuming utilities and are mostly used for specific durations only. Other major causes for different peak loads are the type of day, number of family members in a household, working hours, and whether EVs are used or not. The same peak demand may not exist at the same hour on a daily basis. This leads to a high uncertainty in day-ahead load prediction and affects the forecasting accuracy.

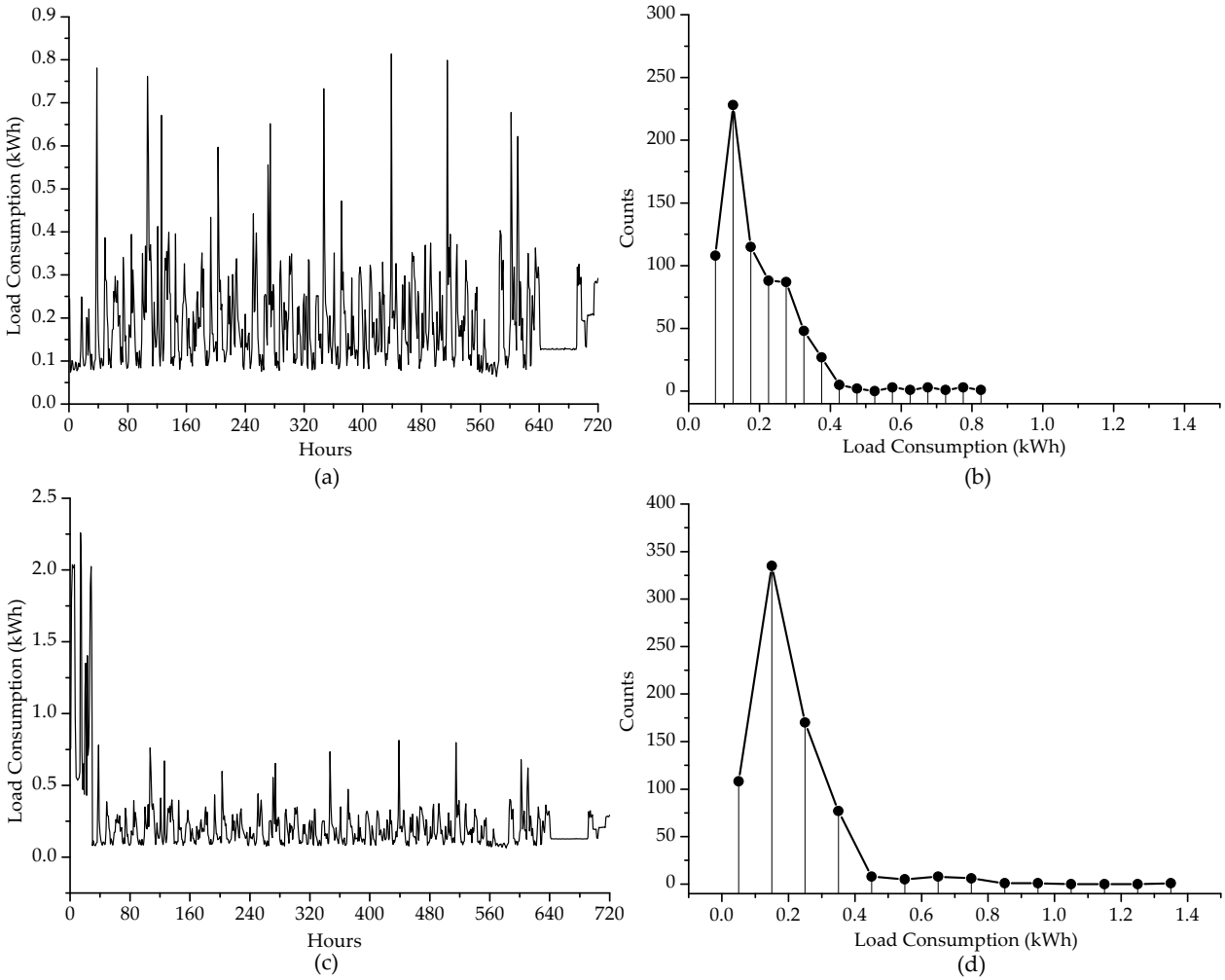

Figure 2. Single-household load series and their corresponding probability distribution functions (PDFs): (**a**) Short-tail load series, (**b**) PDF of the short-tail load series, (**c**) long-tail load series, and (**d**) PDF of the long-tail load series.

3. Proposed Multihousehold Load Forecasting Method

3.1. Tail Length Estimation of the PDF of Electricity Load

Electricity demand data from households with uncertain peaks can result in a long tail in the PDF. In this study, the occurrence of these uncertain peaks was estimated using the CMM index. Statistical moment metrics, such as the mean and variance, can be used to extract the appropriate tail information from each load series and help identify the uniqueness and learn how to predict the uncertain daily few-hour peak loads for each load series.

The overall uncertainty in the PDF of the load-series data is an ensemble of moments from different types of daily load profiles. Let the daily load profile $l_d = [l_{d,1},\ l_{d,2}, l_{d,3},\ \ldots\ldots\ l_{d,t}, \ldots . l_{d,\mathrm{T}}]$, where t represents the hourly load consumption on day d. The following statistical results can be obtained:

$$s_{d,\,1} = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} l_{d,t}$$

$$s_{d,\,2} = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} (l_{d,t} - s_{d,\,1})^2$$

$$s_{d,\,3} = \sqrt{\frac{\sum_{t=1}^{\mathrm{T}} (l_{d,t} - \mathrm{s_{d,1}})^2}{T}}$$
$$s_{d,\,4} = \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \left[\frac{(l_{d,t} - s_{d,\,1})}{s_{d,\,3}} \right]^3 \tag{3}$$
$$s_{d,\,5} = \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \left[\frac{(l_{d,t} - s_{d,\,1})}{s_{d,\,3}} \right]^4$$
$$s_{d,\,6} = \sum_{t=2}^{\mathrm{T}} |l_{d,t} - l_{d,t-1}|$$

where $l_{d,t}$ represents the single-household load consumption value at time t for day d, and $s_{d,1}$, $s_{d,\,2}$, $s_{d,3}$, $s_{d,4}$, $s_{d,5}$, and $s_{d,6}$ are the statistical moment measuring tools for day d, such as the mean, variance, standard deviation, skewness, kurtosis, and cumulative difference, respectively. The cumulative difference is a newly introduced metric that measures the degree to which the daily load profile contains the sum of the hourly differences in the daily load consumption. Each statistical moment measures vary and represent the shift in the uncertainty on a daily basis. Hence, the overall metric difference vector I_d can be computed on a daily basis as follows:

$$I_d = |s_d - s_{d-1}| \; \forall d \tag{4}$$

where $s_d = [s_{d,1}, s_{d,2}, s_{d,3}, s_{d,4}, s_{d,5}, s_{d,6}]$ represents the overall change in the uncertainty values between the load profiles for days d and $d-1$.

The CMM can serve as a valuable tool for defining characteristics such as the long-tail PDF of a single-household load series. Consequently, the CMM vector $\mathrm{I}_m = [I_1, I_2, I_3, \ldots, I_d, \ldots, I_D]$ for a single household m can be obtained, where d represents D historical load profile days. Each load profile has a unique tail property, and the compositions of the different types of load profiles make each household load series unique. In this study, the L2 norm was used to calculate the overall magnitude of the CMM vector, which comprised the entire uncertainty in the load series. The comprehensive uncertainty presented in a CMM vector of the load series m is calculated as follows:

$$I(m) = \sqrt{\sum_{d=1}^{D} |I_d|^2} \tag{5}$$

where $I(m)$ is the L2-norm magnitude value of the CMM vector. Each load series L_m can be rearranged based on the increasing values obtained from Equation (5).

3.2. Reordering Input Load Series

For each household group, each quantitative uncertainty measure $I(m)$ in a single-household load series can be calculated as follows:

$$I = [I(1), I(2), I(3), \ldots\ldots I(m), \ldots\ldots I(N)] \tag{6}$$

where N is the total number of load series in the input household group. The difference between $I(m)$ is measured by the integrated uncertainty between the two load series, and can help place each load series in the appropriate order. The proper placement of the load series may preserve the ordering of the shifting variance in the input load-series dataset. The lowest $I(m)$ indicates that the m-th load series has the lowest number of uncertain load profiles, and its position is calculated as follows:

$$I(m) = \underset{m}{\operatorname{argmin}} \| I(m) \| \tag{7}$$

Equation (7) is applied for all households where N is replaced by $N-1$ each time. All the $N \times (N-1)/2$ iterations are eventually performed, and the ordered CMM values are finalized. The load series are placed according to the lowest $I(m)$ position given by Equation (7), and all the load series are placed in this order. The ordered dataset L is expressed as follows:

$$L = \begin{bmatrix} l_{1,1} & l_{1,2} & \dots & l_{1,d} & \dots & l_{1,D} \\ l_{2,1} & l_{2,2} & \dots & l_{2,d} & \dots & l_{2,D} \\ \vdots & \dots & \ddots & \dots & \vdots & \\ l_{m,1} & l_{m,2} & \dots & l_{m,d} & \dots & l_{m,D} \\ \vdots & \dots & \vdots & \dots & \vdots & \\ l_{N,1} & l_{N,2} & \dots & l_{N,d} & \dots & l_{N,D} \end{bmatrix}^T \tag{8}$$

where $L = [L'_1, L'_2, L'_3, \dots L'_m, \dots L'_N]^T$, and each $L'_m = (l_{m,1} \ l_{m,2} \ \dots \ l_{m,d} \ \dots \ l_{m,D})$ is the m-th-ordered single-household load series containing daily load profiles of D days.

3.3. Data Augmentation for Single-Household Load Series

The main purpose of using data augmentation in the residential sector is to learn new features from a single load series that is less volatile but still shows all the possible characteristics of the selected series [17]. Each L'_m is passed to the data-augmentation procedure, which creates $\mathcal{L}$ number of homogenous load series. Using multiple k-means clustering, each L'_m generates $\mathcal{L}$ different centroid load profiles $C^m_{\updownarrow}$ that are less volatile and uncertain relative to each other. Each load profile l^m_d of L'_m is decomposed into its centroid load profile $C^m_{\updownarrow}$ and daily residual load profile $r^m_d(\updownarrow)$ as follows:

$$l_{m,d} = C^m_{\updownarrow} + r^m_d(\updownarrow) \tag{9}$$

The identified $C^m_{\updownarrow}$ extract completes the seasonality of the load series, which is repetitive and easy to estimate using forecasting methods. By contrast, $r^m_{\updownarrow,d}$ comprises a large amount of noisy information that implicitly contains the tail PDF property of the load series and is difficult to estimate. With $C^m_{\updownarrow}$, a new residual load series is generated that contains all the uncertain information of the load series, as given by:

$$R^m(\updownarrow) = L'_m - C^m_{\updownarrow} \tag{10}$$

where $R^m(\updownarrow) = \left(r^m_1(\updownarrow), r^m_2(\updownarrow), r^m_3(\updownarrow), \dots th\ r^m_d(\updownarrow), \dots . r^m_d(\updownarrow)\right)$ comprises new residual load profiles $r^m_d(\updownarrow)$ generated by $C^m_{\updownarrow}$ for all input days d. The efficacy of $R^m(\updownarrow)$ can be enhanced by maintaining k-input to the multiple k-means clustering. However, unwanted residual series are generated when the input k is greater than six. To explore only a homogenous residual load series and collect them in a single matrix R_{in}, the following equation was derived [17]:

$$R^m_{in} = \left\{ r_{\updownarrow} \middle| \varnothing_{\updownarrow} \leq \frac{1}{\mathcal{L}} \sum_{\updownarrow=1}^{\mathcal{L}} \varnothing_{\updownarrow} \right\} \tag{11}$$

where $\varnothing_{\updownarrow}$ is the Frobenius norm of $R^m(\updownarrow)$, which selects only less volatile residual series. Each residual series shows similar properties to the inputted m-th load series; when their corresponding $C^m_{\updownarrow}$ is added, an augmented load series L^m_{in} is generated as follows:

$$L^m_{in} = \left[L'_m(1), L'_m(2), L'_m(3), \dots . L'_m(\updownarrow) \dots L'_m(\mathcal{L})\right] \tag{12}$$

where $L'_m(\updownarrow) = \left[l^m_{\updownarrow 1}, l^m_{\updownarrow,2}, l^m_{\updownarrow,3} \dots l^m_{\updownarrow,d} \dots . l^m_{\updownarrow,D}\right]$ represents the extracted $\updownarrow$-th homogenous load series for household m. L^m_{in} includes $(\mathcal{L} \times D)$ load profiles with $\mathcal{L}$ number of augmented

load series for household m. Figure 3 shows the extracted $\mathcal{L}$ homogenous load series L_{in}^{m} from the augmentation CNN method for the two tested single-household load series.

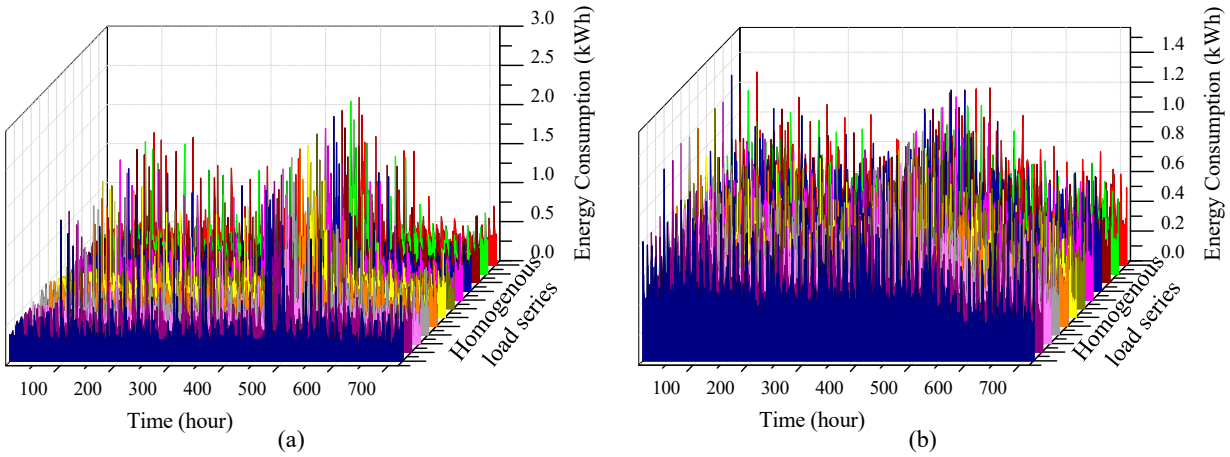

Figure 3. Homogenous load series obtained from the augmentation technique: (**a**) Household load series 1 with a short-tail PDF, (**b**) household load series 2 having a long-tail PDF.

3.4. Proposed Multihousehold Load Forecasting

Many CNN architectures are available in pretrained versions for image classification, segmentation, localization, and object detection [22–24]. The CNN architecture passes the input data to different layers: (a) input, (b) convolution layer, (c) pooling, and (d) fully connected layers. However, there are no precise rules for designing a CNN architecture. All the proposed CNN-based architectures require a large amount of input data to ensure better accuracy.

To perform multihousehold load forecasting operations, the proposed CNN forecasting model was designed in a manner similar to the existing CNN architecture [19–21]. Let all the augmented multihousehold load series collected in order using Equation (12) be given by:

$$X = \begin{bmatrix} L'_1(1), L'_1(2), L'_1(3), \ldots . L'_1(\updownarrow) \ldots L'_1(\mathcal{L}) \\ L'_2(1), L'_2(2), L'_2(3), \ldots . L'_2(\updownarrow) \ldots L'_2(\mathcal{L}) \\ L'_3(1), L'_3(2), L'_3(3), \ldots . L'_3(\updownarrow) \ldots L'_3(\mathcal{L}) \\ \vdots \ \vdots \ \vdots \ \ddots \ \ddots \ \ddots \ \vdots \ \ddots \ \vdots \\ L'_m(1), L'_m(2), L'_m(3), \ldots . L'_m(\updownarrow) \ldots L'_m(\mathcal{L}) \\ \vdots \ \vdots \ \vdots \ \ddots \ \ddots \ \ddots \ \vdots \ \ddots \ \vdots \\ L'_N(1), L'_N(2), L'_N(3), \ldots . L'_N(\updownarrow) \ldots L'_N(\mathcal{L}) \end{bmatrix} \quad (13)$$

where X contains $N \times (\mathcal{L} \times D)$ load profiles with N input load series, $\mathcal{L}$ is the count of the augmented series, and D units of daily load profiles are presented in the load series.

The newly generated homogenous load series was treated as an image-like input and was convolutionally operated using CNN filters. Although CNN filters are resistant to learning over noisy information, X provides the necessary and sufficient information to predict the high demand for a few hours. The convolved data are passed to the pooling layer, which is vital for selecting the maximum pointing information to update the weight and bias in the CNN layers. The ending of the CNN layers is attached to a fully connected layer that receives all the higher-dimensional learning weights in a one-dimensional space. Eventually, this layer performs the forecasting operation for a multihousehold load in one time-step. To perform the training operation, the proposed model provides multihousehold

load profiles $\hat{Y}_t$, which receive the actual multihousehold load profiles Y_t. The two profiles can be expressed as:

$$\hat{Y}_t = \begin{bmatrix} \hat{l}_{1,D,\,1} & \hat{l}_{1,D,\,2} & \cdots & \hat{l}_{1,D,\,t} & \cdots & \hat{l}_{1,D,\,T} \\ \hat{l}_{2,D,\,1} & \hat{l}_{2,D,\,2} & \cdots & \hat{l}_{2,D,\,t} & \cdots & \hat{l}_{2,D,\,T} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \hat{l}_{m,D,\,1} & \hat{l}_{m,D,\,2} & \cdots & \hat{l}_{m,D,\,t} & \cdots & \hat{l}_{m,D,\,T} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \hat{l}_{N,D,\,1} & \hat{l}_{N,D,\,2} & \cdots & \hat{l}_{N,D,\,t} & \cdots & \hat{l}_{N,D,\,T} \end{bmatrix} \tag{14}$$

$$Y_t = \begin{bmatrix} l_{1,D,\,1} & l_{1,D,\,2} & \cdots & l_{1,D,\,t} & \cdots & l_{1,D,\,T} \\ l_{2,D,\,1} & l_{2,D,\,2} & \cdots & l_{2,D,\,t} & \cdots & l_{2,D,\,T} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ l_{m,D,\,1} & l_{m,D,\,2} & \cdots & l_{m,D,\,t} & \cdots & l_{m,D,\,T} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ l_{N,D,\,1} & l_{N,D,\,2} & \cdots & l_{N,D,\,t} & \cdots & l_{N,D,\,T} \end{bmatrix} \tag{15}$$

where $l^m_{D+1,\,t}$ and $\hat{l}^m_{D+1,\,t}$ are the actual and predicted hourly load consumptions at time t. The optimizing equation developed for the proposed multihousehold load forecasting algorithm can be written as:

$$Minimize \sqrt{\sum_{j} \sum_{m=1}^{N} \sum_{t=1}^{T} \left(l_{m,D,\,t} - \hat{l}_{m,D,\,t} \right)^2} \tag{16}$$

where j denotes the number of predefined training epochs in the proposed CNN model.

Figure 4 shows the overall block diagram of the proposed multihousehold load forecasting method. The peak consumed loads in each load series generated a significant difference between the load profiles. The reordering of each load series based on the CMM was effective in maintaining a low shifting variance in the input multihousehold dataset. Maintaining this variance is useful for achieving a high forecasting accuracy for the multiple load series, where every household forecasting result within a different household group can be obtained simultaneously. The N-load series is maintained by ordering based on the tail PDFs for multihousehold load forecasting. With each ordered load series as the input, an augmented load series was generated and attached by maintaining the minimum shifting variance in the dataset. Consequently, the consecutive household load series from the ordering process presented the minimum difference in its tail characteristics.

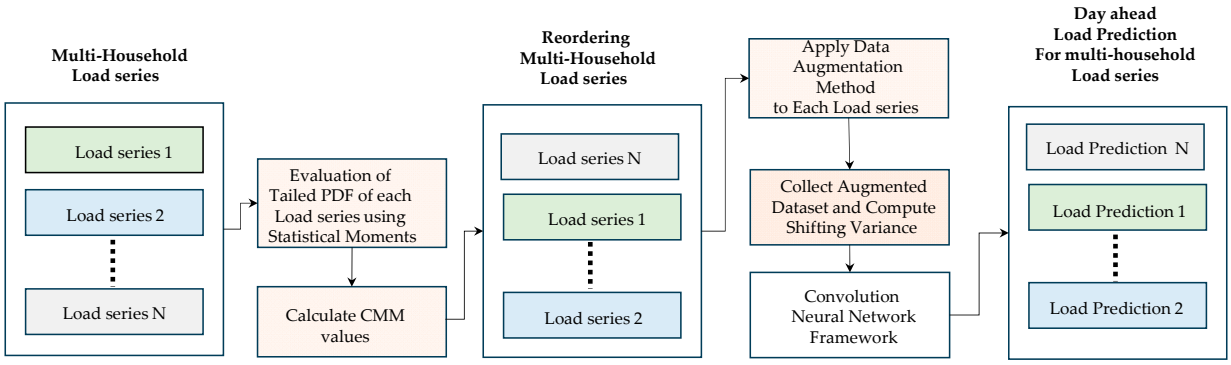

Figure 4. Block diagram of the proposed multihousehold load forecasting method.

Figure 5 shows the flowchart of the proposed methodology for multihousehold load forecasting, comprising three stages: (a) CMM load-series ordering, (b) data augmentation, and (c) the CNN framework. In the first stage, multiple load series are ordered using the lowest CMM values. Data augmentation is applied in the second stage to maintain a constant or minimum shifting variance in the input dataset. Finally, in the third stage, *N* load profiles are predicted for N multiple household load series using the CNN hyperparameters.

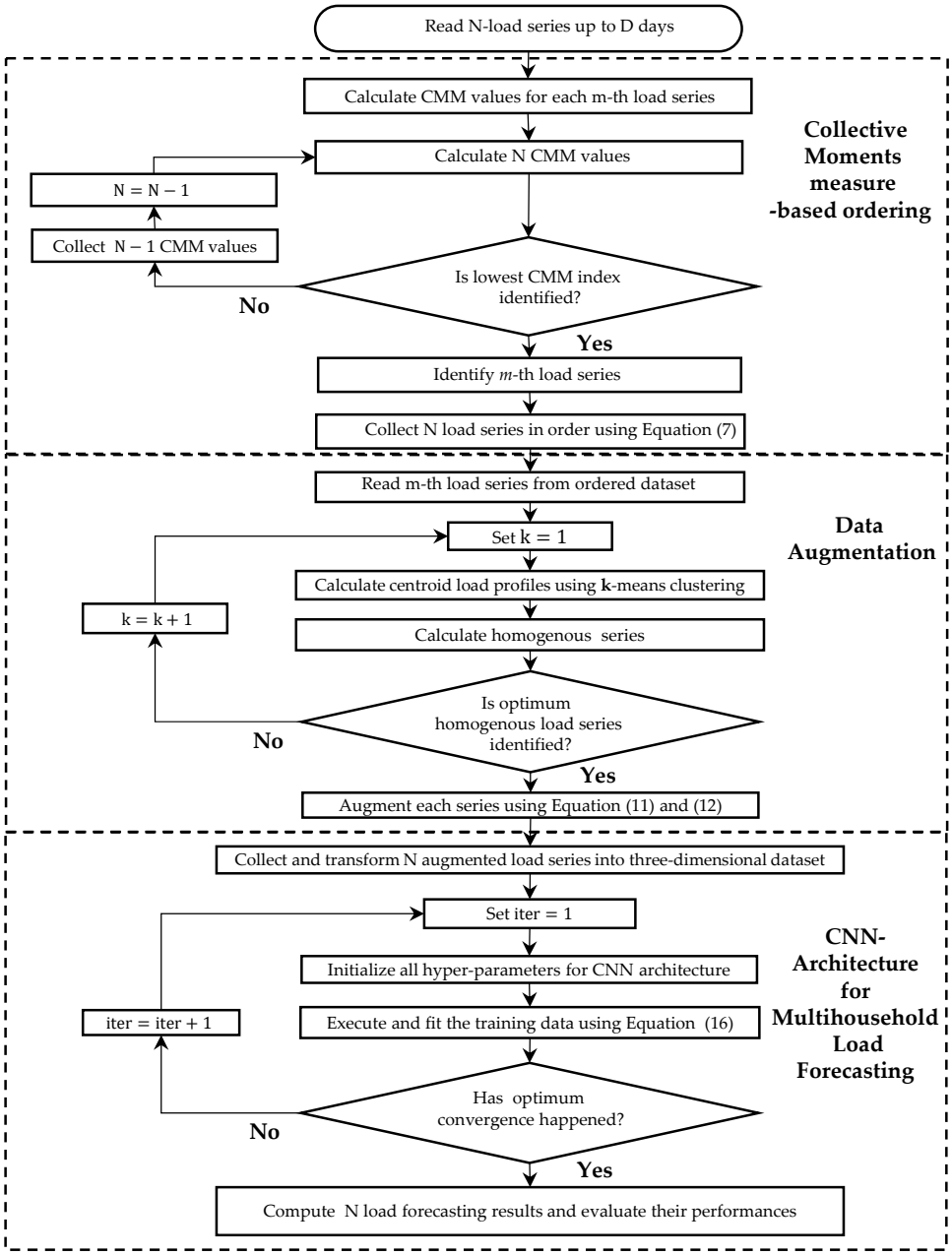

Figure 5. Flowchart of the proposed multihousehold day-ahead load forecasting method.

4. Simulation Results

4.1. Data Description and CNN Architecture Modeling

A large amount of multiload series data was collected from 1181 residential households with AMI installed, located in Seoul, South Korea, for the experiment. The AMI recording began in September 2016 and ended in August 2017.

Figure 6 shows the model of the proposed CNN architecture for the prepared dataset containing an input *N* household load series. The training dataset contained each load

series data up to day D, whereas the testing dataset contained only multiload profiles of the next day, $D + 1$.

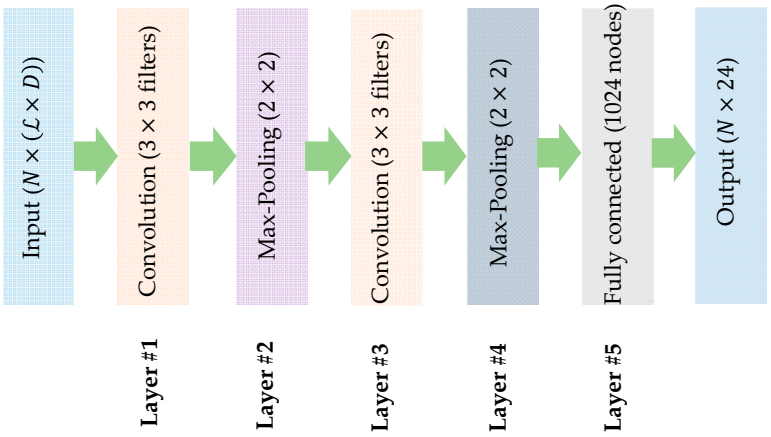

Figure 6. Modeling of the proposed CNN-based architecture for multihousehold load forecasting.

Table 1 lists the hyperparameters of the tested DL framework for performing multihousehold forecasting operations. The program was written in Python using the Keras and TensorFlow libraries [35–37]. The program was tested on an i7 CPU-based computer, and numerous forecasting results were obtained using Google Colab, which is a GPU-based cloud application. Three core DL architectures, namely CNN, LSTM, and CNN–LSTM, were employed at the start of the proposed method to determine which DL benchmark produced the best forecasting results.

Table 1. Hyper-parameter modeling for the tested core DL framework.

Method	Hidden Layers	Hidden Nodes/Filters	No. of Epochs(Iteration)	Activation Function	Optimizer
CNN	2	96	150	ReLU	RMSprop
LSTM	2	72	300	Sigmoid, tanh	RMSprop
CNN–LSTM	2	96	150	ReLU	RMSprop

4.2. Performance Evaluation Metrics

To generalize the proposed forecasting methodology, each input dataset containing 10, 20, 50, 80, and 100 household load series was used for performance evaluation. To compute the load forecasting results for the m-th household in each multiple-household group, the following performance metrics were used for each DL methodology for comparison:

$$\mathrm{MAPE_m} = \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\left| l^{m\prime}_{D+1,\,t} - \hat{l}^{m\prime}_{D+1,\,t} \right|}{l^{m\prime}_{D+1,\,t}} \times 100\% \right)_m \tag{17}$$

$$\mathrm{RMSE_m} = \left(\sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\left(l^{m\prime}_{D+1,\,t} - \hat{l}^{m\prime}_{D+1,\,t} \right) \right)^2} \right)_m \tag{18}$$

$$\mathrm{MAPE_{avg}} = \frac{1}{N} \sum_{m=1}^{N} \mathrm{MAPE_m} \tag{19}$$

$$\mathrm{RMSE_{avg}} = \frac{1}{N} \sum_{m=1}^{N} \mathrm{RMSE_m} \tag{20}$$

where $l^{m\prime}_{D+1,\,t}$ and $\hat{l}^{m\prime}_{D+1,\,t}$ are the actual and predicted load consumptions of the next day $D + 1$ for hour t; MAPE_m and RMSE_m are the MAPE and root-mean-square error (RMSE) for testing household m. MAPE_{avg} and RMSE_{avg} represent the average MAPE and RMSE results, respectively, to generalize the forecasting results of the proposed multihousehold methodology.

4.3. Single-Household Load Forecasting

To evaluate the performance results for an individual household, the first ten household load series from the first group were used for forecasting using the CNN, LSTM, and CNN–LSTM methodologies. The input dataset was ordered, augmented, and prepared using the proposed methodology. Figure 7 shows the actual and predicted daily load profiles of the nine tested load series for the next day $D + 1$. In Figure 7, the forecasting results of the tested load profiles are arranged from lower-tail PDF load series to the longer-tail PDF load series. The CNN produced the most accurate prediction results, as shown in Figure 7g–i, which had a longer-tail PDF. Hence, the proposed CNN architecture predicted accurate results for the peak load consumption, despite the unusually high load demand observed for a few hours in a day.

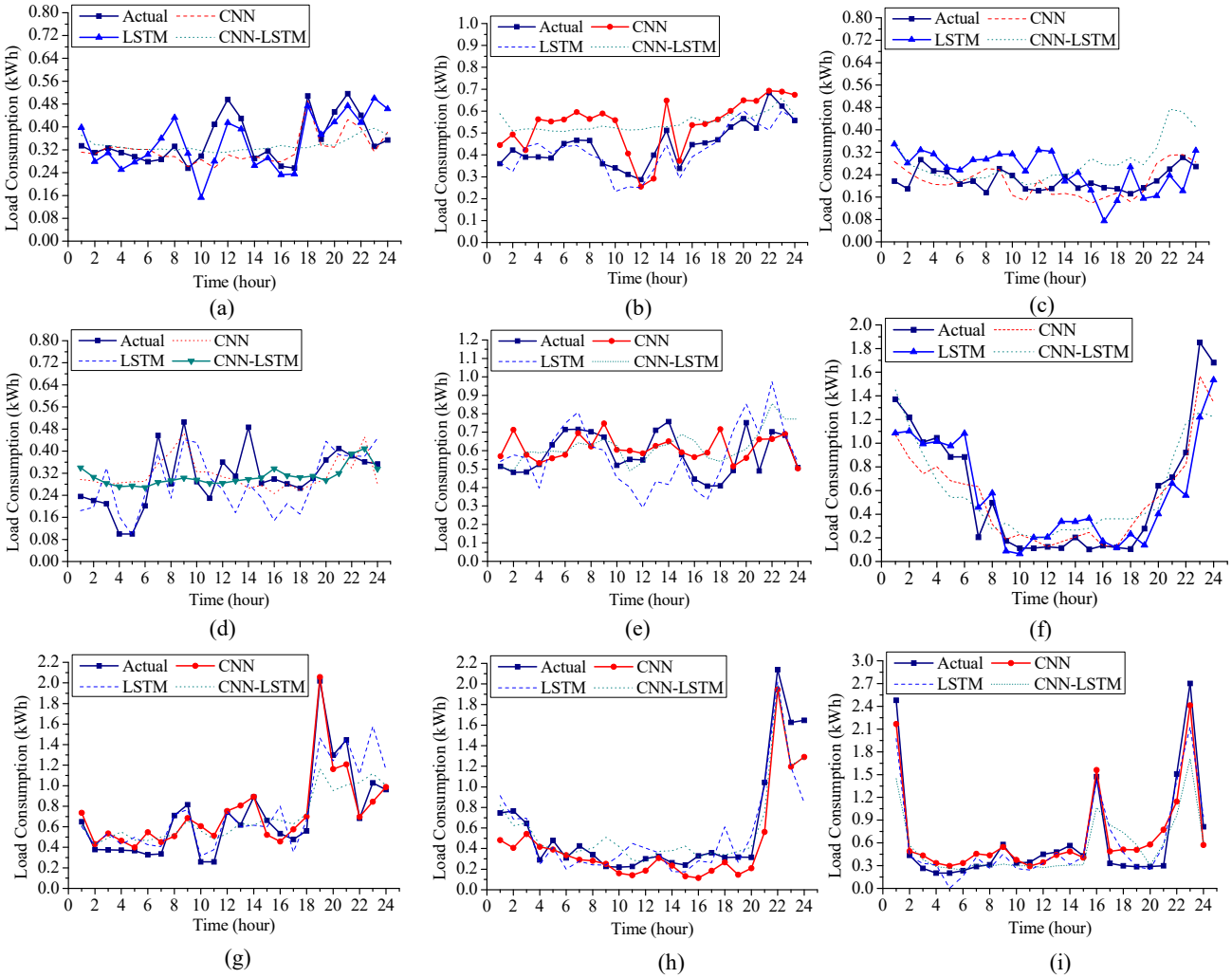

Figure 7. Predicted next-day load profiles of nine different households from the first household group: (**a**) ordered short-tail household 1, (**b**) ordered short-tail household 2, (**c**) ordered short-tail household 3, (**d**) ordered medium-tail household 4, (**e**) ordered medium-tail household 5, (**f**) ordered medium tail household 6, (**g**) ordered long-tail household 7, (**h**) ordered long-tail household 8, and (**i**) ordered long-tail household 9.

Table 2 presents the forecasting MAPE and RMSE results for the ten households. Among the ten load series, the CNN outperformed for six of them that had longer tail characteristics in their PDFs. In terms of the average forecasting results, the CNN enhanced the performance results, with the forecasting $MAPE_{avg}$ being higher than 2.3% and forecasting $RMSE_{avg}$ being higher than 0.032 kWh compared with the LSTM and CNN–LSTM algorithms, respectively.

Table 2. Comparison of the forecasting performance in terms of the MAPE (%) and RMSE (kWh) for the first household group.

Household	MAPE (%)			RMSE (kWh)		
	LSTM	CNN	CNN–LSTM	LSTM	CNN	CNN–LSTM
1	18.45	22.55	23.18	0.071	0.093	0.090
2	17.43	16.49	18.32	0.106	0.105	0.122
3	21.11	27.72	27.14	0.078	0.104	0.103
4	32.18	23.67	26.10	0.109	0.108	0.142
5	32.81	20.66	28.10	0.164	0.136	0.162
6	18.56	26.10	25.74	0.125	0.155	0.152
7	45.10	32.82	37.85	0.130	0.111	0.116
8	29.37	27.53	28.23	0.137	0.135	0.179
9	35.27	31.15	32.06	0.256	0.208	0.213
10	22.79	22.63	22.83	0.208	0.207	0.208

4.4. Average Single-Household Load Forecasting

The average single-household load forecasting results are effective for generalizing the performance of the proposed multihousehold load forecasting model. Figure 8 shows the box-plot results of the average forecasting performance result of four different household groups for a week. Each box plot represents the average MAPE results from the household groups containing 30, 50, 80, and 100 household load series. The box plot reveals that the performance of the CNN varies much less than those of LSTM and CNN–LSTM, despite the increase in the number of load series.

Table 3 shows the average MAPE performance results of all the multihousehold groups in July 2017. The month of July was considered for the observation because it contained mostly long-tail PDF load profiles in the load series compared with the other months in South Korea. Each result can be generalized to each household in the household group. As a result, the CNN-based model demonstrated a better performance, with the forecasting MAPE being higher than 3.5% and the forecasting RMSE being higher than 0.035 kWh for the general case.

Table 3. Average forecasting performance results in terms of the MAPE (%) and RMSE (kWh) for the different multi-household groups in July 2017.

Multihousehold Groups	MAPE (%)			RMSE (kWh)		
	LSTM	CNN	CNN–LSTM	LSTM	CNN	CNN–LSTM
$N = 10$	25.64	22.33	26.95	0.145	0.137	0.147
$N = 20$	31.83	27.32	30.27	0.151	0.157	0.177
$N = 30$	34.49	27.67	30.37	0.176	0.148	0.161
$N = 50$	37.74	31.16	35.69	0.237	0.172	0.209
$N = 80$	36.75	28.39	34.82	0.202	0.155	0.201
$N = 100$	38.36	32.43	36.16	0.171	0.131	0.162

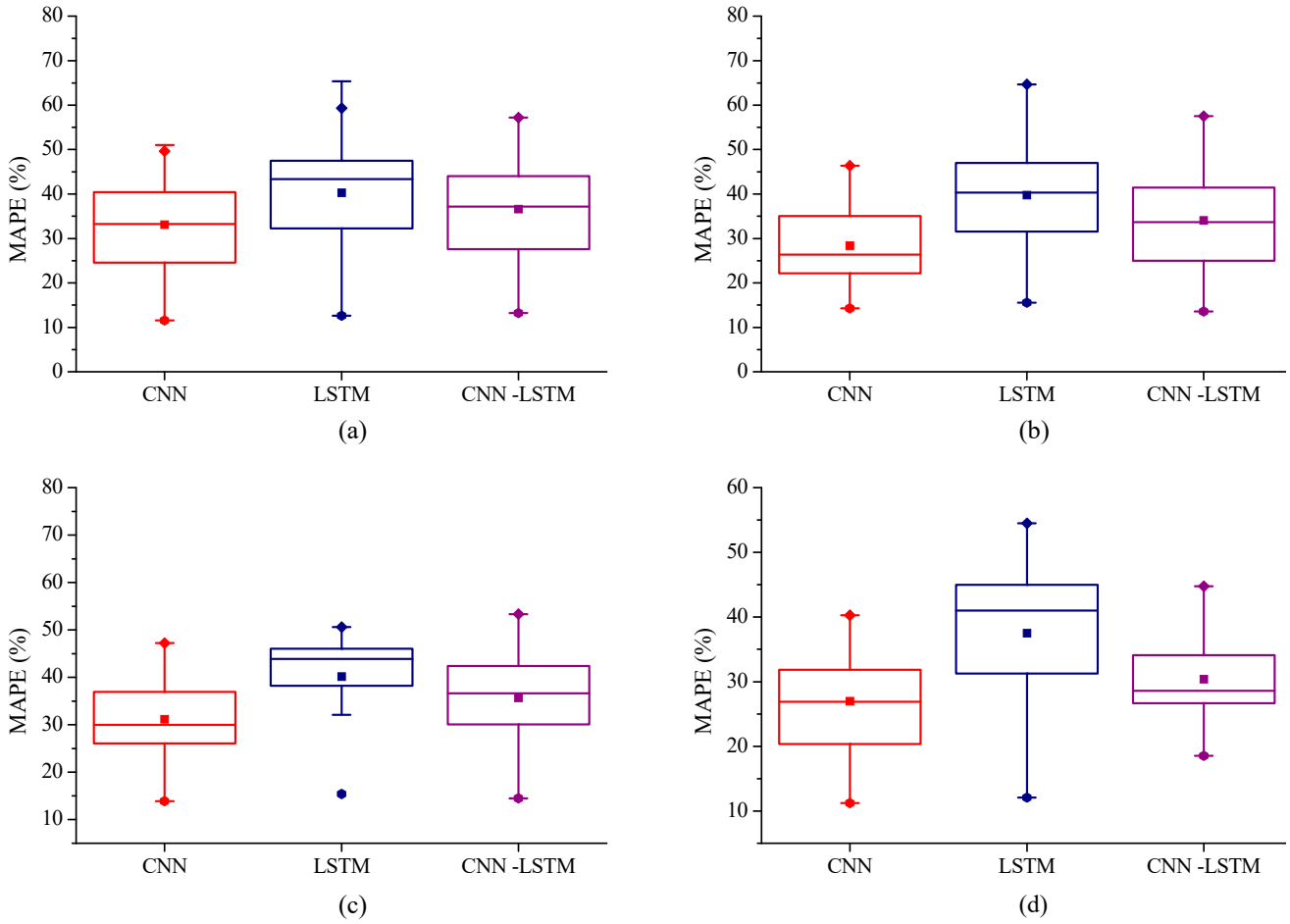

Figure 8. Box-plot results of the average MAPE for the different household groups: (**a**) 30 households, (**b**) 50 households, (**c**) 80 households, and (**d**) 100 households.

4.5. Comparison with Pooled and Data Augmentation Methodologies

Pooled and data augmentation methodologies are efficient forecasting strategies for the electricity load of a single household. Therefore, they were compared with the proposed methodology for the same amount of input load series. For a robust comparison, four weeks from each season were selected to evaluate the forecasting performance of the 20, 50, 80, and 100 household load series. Table 4 presents a comparison of the forecasting performances of the proposed methodology, pooled CNN, and augmented CNN in terms of $MAPE_{avg}$ and $RMSE_{avg}$. The proposed methodology shows noteworthy improvements of more than 8% and 0.108 kWh in $MAPE_{avg}$ and $RMSE_{avg}$, respectively. These results are significant for single-household day-ahead forecasting.

4.6. Augmentation vs. Ordered Strategies

Because the proposed method uses both augmentation techniques and ordered strategies, its forecasting evaluations were studied separately. Both the methods were found to be suitable for multihousehold load forecasting. Table 5 presents the average performance results in terms of $MAPE_{avg}$ and $RMSE_{avg}$ for the augmented, ordered, and proposed methods. The comparison shows that creating a hybrid configuration based on augmentation and reordering using the CMM for the CNN brings about an improvement of more than 2% in the MAPE. Hence, both data enlargement from augmentation and ordering based on the tail PDF of multiple load series are useful for producing robust forecasting results of the multihousehold load.

Table 4. Comparison of the average forecasting MAPE (%) and RMSE (kWh) performance results between existing and proposed forecasting methodologies for all household groups across all seasons in 2017.

Multihousehold Groups	Season	MAPE (%)			RMSE (kWh)		
		Pooled CNN	Augmented CNN	Proposed	Pooled CNN	Augmented CNN	Proposed
$N = 20$	Spring	36.23	29.28	25.46	0.236	0.198	0.122
	Summer	39.25	31.26	28.13	0.245	0.204	0.111
	Autumn	41.25	35.36	27.32	0.275	0.207	0.157
	Winter	42.63	33.23	26.63	0.277	0.213	0.118
	Average	39.84	32.28	26.82	0.258	0.205	0.127
$N = 50$	Spring	34.23	29.12	23.38	0.244	0.191	0.092
	Summer	36.15	30.93	27.92	0.251	0.201	0.107
	Autumn	39.37	33.72	31.16	0.297	0.218	0.172
	Winter	38.42	29.82	23.69	0.281	0.209	0.095
	Average	37.04	30.89	26.53	0.268	0.204	0.116
$N = 80$	Spring	33.15	28.12	22.26	0.256	0.181	0.075
	Summer	36.52	33.85	28.39	0.274	0.193	0.155
	Autumn	35.26	30.09	25.15	0.269	0.212	0.095
	Winter	34.29	27.06	22.15	0.264	0.195	0.088
	Average	34.80	29.78	24.48	0.265	0.195	0.103
$N = 100$	Spring	32.25	29.65	27.75	0.233	0.171	0.082
	Summer	34.68	31.17	27.99	0.241	0.182	0.081
	Autumn	36.16	33.46	31.11	0.278	0.196	0.121
	Winter	35.17	31.16	28.30	0.271	0.193	0.098
	Average	34.56	31.36	28.78	0.255	0.185	0.095

Table 5. Comparison of the forecasting performance results in terms of MAPE (%) and RMSE (kWh) between augmentation, ordered, and proposed methodologies.

Method	MAPE (%)				RMSE (kWh)			
	$N = 20$	$N = 50$	$N = 80$	$N = 100$	$N = 20$	$N = 50$	$N = 80$	$N = 100$
Augmentation	30.09	28.92	26.15	29.31	0.217	0.117	0.096	0.087
Only ordered	33.25	32.48	31.36	34.21	0.184	0.164	0.144	0.126
Proposed	28.13	27.16	24.29	26.99	0.111	0.110	0.095	0.082

5. Discussion

In this study, we developed a methodology for identifying synergies in energy consumption patterns between households and presented simulation results demonstrating the effectiveness of the proposed approach.

As each household's electricity consumption differed from others', especially in peak load values for a few hours in a day, the difference was explicitly illustrated when analyzing the tail PDF of the load series. The statistics-based CMM index was able to order the load series by computing the overall moment measure of the PDF generated from the load series. The electricity demand data used consisted of data from groups of households with ensured similarity. The ordered dataset was enlarged by using a data augmentation methodology that is effective in fulfilling the data requirements of the CNN. As a result, the proposed approach showed a significant improvement of more than 8% and 3% in forecasting accuracy in terms of the MAPE over existing CNN methods for the typical individual household.

Despite efforts to deliver higher forecasting results, the proposed model exhibits some unaddressed aspects and limitations. While the proposed method demonstrated utility in a case study involving 100 households over a one-year period, it is limited by the availability of only one year's worth of data. Implementing the proposed methodology requires access to comprehensive data, which may not always be possible. In cases where access

is restricted or data availability is limited, applying the proposed approach to groups of households with slightly different characteristics remains unexplored territory. Moreover, alternative methods for identifying synergies could yield better results. Exploring these methods could enhance the robustness of the proposed approach.

In conclusion, the performance of small-scale electricity demand prediction could be improved by aligning input data and adjusting the learning sequence with CMM, along with the combination of data augmentation. However, the underlying principles behind the synergy of this combination have not been analyzed. This study did not explore alternative methods for data augmentation techniques for better synergy with rearranging input data; therefore, further research in this area will be necessary in the future.

6. Conclusions

Residential load series contain granular load consumption values. Therefore, their performance results are comparatively higher, particularly in terms of the MAPE. Existing DL-based forecasting methodologies for load series are often inefficient due to the high load demand observed for a few hours in a given day, which is explicitly depicted as a long tail in the PDF of the load series. A multihousehold forecasting method based on load-series ordering, data augmentation, and CNN seems to be most suitable.

The main results of this study can be summarized as follows:

- Multihousehold forecasting research is significant for residential living areas such as apartment blocks, community buildings, or even small towns where each household's electricity consumption is different from others';
- The electricity consumption characteristics of each household are explicitly demonstrated in the load series PDF. Based on tail information in each load series' PDF, CMM ordering was developed;
- The proposed multihousehold load forecasting technique uses the CMM to index the order of the load series. The CMM produces an ordered load series by computing the overall moment measure of the PDF generated from the load series;
- The ordered multiple series maintain a minimum or constant shifting variance in the load-series dataset inputted to the CNN. A data augmentation method is used for each load series of the input multihousehold group to solve the existing data requirement issues of the CNN;
- A comparison of the average forecasting results for single households helped validate the efficacy of the proposed method;
- A higher load forecasting accuracy for both single and multiple households could be realized by utilizing the CMM, data augmentation, and CNN, which solves the peak load demand issue encountered in the residential sector.

In this manuscript, we have discussed multihousehold forecasting using moment information and data augmentation. In the future, we would like to explore deep statistical methodologies and incremental learning for the multihousehold electricity consumptions.

Author Contributions: S.K.A. developed the augmentation strategies for the CNN-based forecasting framework, performed the simulation operations, and wrote the paper. H.Y. and Y.-M.W. analyzed the simulation results and commented on the manuscript. J.L. provided guidance for research and revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1C1C1012408).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to the conduct of further research based on some of the data in this article.

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

References

1. Box, E.P.G.; Jenkins, G.M.; Gwilym, R.C.G.; Ljung, G.M. *Time Series Analysis Forecasting and Control*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015.
2. Gram-Hanssen, K. Standby consumption in households analyzed with a practice theory approach. *J. Ind. Ecol.* **2010**, *14*, 150–165. [CrossRef]
3. Richardson, I.; Thomson, M.; Infield, D.; Clifford, C. Domestic electricity use: A high resolution energy demand model. *Energy Build.* **2010**, *42*, 1878–1887. [CrossRef]
4. Stephen, B.; Tang, X.; Harvey, P.R.; Galloway, S.; Jennett, K.I. Incorporating Practice Theory in Sub-Profile Models for Short term aggregated Residential Load Forecasting. *IEEE Trans. Smart Grid* **2017**, *8*, 1591–1598. [CrossRef]
5. Abbas, F.; Feng, D.; Habib, S.; Rahman, U.; Rasool, A.; Yan, Z. Short term residential load forecasting: An improved optimal nonlinear auto regressive (NARX) method with exponential weight decay function. *Electronics* **2018**, *7*, 432. [CrossRef]
6. Welikala, S.; Dinesh, C.; Ekanayake, M.P.B.; Godaliyadda, R.I.; Ekanayake, J. Incorporating appliance usage patterns for non-Intursive load monitoring and load forecasting. *IEEE Trans. Smart Grid* **2017**, *10*, 448–461. [CrossRef]
7. Cheng, C.; Sa-Ngasoongsong, A.; Beyca, O.; Le, T.; Yang, H.; Kong, Z.J.; Bukkapatnam, S.T.S. Time series forecasting for nonlinear and non-stationary processes: A review and comparative study. *IIE Trans.* **2015**, *47*, 1053–1071. [CrossRef]
8. Xie, G.; Chen, X.; Weng, Y. An integrated Gaussian process modelling framework for residential load prediction. *IEEE Trans. Power Syst.* **2018**, *33*, 7238–7248. [CrossRef]
9. Ghofrani, M.; Ghayekhloo, M.; Arabali, A.; Ghayekhloo, A. A hybrid short-term load forecasting with a new input selection framework. *Energy* **2015**, *81*, 777–786. [CrossRef]
10. Wang, Y.; Chen, Q.; Gan, D.; Yang, J.; Kirschen, D.S.; Kang, C. Deep learning-based socio-demographic information identification from smart meter data. *IEEE Trans. Smart Grid* **2018**, *10*, 2593–2602. [CrossRef]
11. Zhang, Y.; Chen, W.; Xu, R.; Black, J. A cluster-based method for calculating baselines for residential loads. *IEEE Trans. Smart Grid* **2015**, *7*, 2368–2377. [CrossRef]
12. Zhong, S.; Tam, K.S. Hierarchical classification of load Profies based on their characteristics attributes in Frequecy domain. *IEEE Trans. Power Syst.* **2015**, *30*, 1–8. [CrossRef]
13. Quilumba, F.L.; Lee, W.J.; Huang, H.; Wang, D.Y.; Szabados, R.L. Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Trans. Smart Grid* **2014**, *6*, 911–918. [CrossRef]
14. Cecati, C.; Kolbusz, J.; Rozycki, P.; Siano, P.; Wilamowski, B.M. A novel RBP training algorithm for Shor-term electric load forecasting and comparative studies. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6519–6529. [CrossRef]
15. Ding, N.; Benoit, C.; Foggia, G.; Besanger, Y.; Wurtz, F. Neural network-based model design for short-term load forecast in distributed systems. *IEEE Trans. Power Syst.* **2016**, *31*, 72–81. [CrossRef]
16. Shi, H.; Xu, M.; Li, R. Deep learning for household load forecasting-A novel pooling deep RNN. *IEEE Trans. Smart Grid* **2018**, *9*, 5271–5280. [CrossRef]
17. Acharya, S.K.; Wi, Y.-M.; Lee, J. Short-term load forecasting for a single household based on convolution neural networks using data augmentation. *Energies* **2019**, *12*, 3560. [CrossRef]
18. Claessencs, B.J.; Varancx, P.; Ruelens, F. Convolution Neural Network for Automatic State-Time Feature Extraction in Reinforcement Learning Applied to Residential Load Control. *Trans. Smart Grid* **2018**, *9*, 3259–3269. [CrossRef]
19. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2017**, *10*, 841–851. [CrossRef]
20. Bessani, M.; Massignan, J.A.; Santos, T.M.; London, J.B., Jr.; Maciel, C.D. Multiple households very short-term load forecasting using bayesian networks. *Electr. Power Syst. Res.* **2020**, *189*, 106733. [CrossRef]
21. Jiang, L.; Wang, X.; Li, W.; Wang, L.; Yin, X.; Jia, L. Hybrid Multitask Multi-Information Fusion Deep Learning for Household Short-Term Load Forecasting. *IEEE Trans. Smart Grid* **2021**, *12*, 5362–5372. [CrossRef]
22. LeCun, Y.; Bengio, Y. *Convolution Neural Networks for Images, Speech and Time Series*; MIT Press: Cambridge, MA, USA, 1988.
23. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Stateline, NV, USA, 3–6 December 2012; pp. 1097–1105.
24. Pedregosa, F.; Varoquax, G.; Gramfort, A. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
25. Badeau, R.; Plumbley, M.D. Multichannel HR-NMF for modelling convolute mixtures of nonstationary signals in the time-frequency domain. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 20–23 October 2013; Volume 2013, pp. 1–4. [CrossRef]
26. Jeong, J.; Kim, H. Multi-site photovoltaic forecasting exploiting space-time convolutional neural network. *Energies* **2019**, *12*, 4490. [CrossRef]
27. Chen, W.; Xie, D.; Zhang, Y.; Pu, S. All you need is a few shifts: Designing efficient convolutional neural networks for image classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7234–7243. [CrossRef]
28. Chaman, A.; Dokmanic, I. Truly shift-invariant convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3772–3782. [CrossRef]
29. Fan, G.-F.; Guo, Y.-H.; Zheng, J.-M.; Hong, W.-C. Application of the weighted K-nearest neighbor algorithm for short-term load forecasting. *Energies* **2019**, *12*, 916. [CrossRef]

30. Ioffe, X.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
31. Huang, Q.; Li, J.; Zhu, M. An Improved convolution neural network with load range discretization for probabilistic load forecasting. *Energy* **2020**, *203*, 117902. [CrossRef]
32. Zhang, R. Making convolutional networks shift-invariant again. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 7324–7334.
33. Kwac, J.; Flora, J.; Rajagopal, R. Household energy consumption segmentation using hourly data. *IEEE Trans. Smart Grid* **2015**, *5*, 420–430. [CrossRef]
34. Marini, L.; Gutierrez-Polo, I.; Kooper, R.; Satheesan, S.P.; Burnette, M.; Lee, J.; Nicholson, T.; Zhao, Y.; McHenry, K.; Zhao, Y.; et al. Clowder: Open source data management for long tail data. In Proceedings of the Practice and Experience on Advanced Research Computing, Pittsburgh, PA, USA, 22–26 July 2018; pp. 1–8. [CrossRef]
35. Tensor Flow: Large Scale Machine Learning on Heterogeneous Distrusted Systems. *arXiv* **2016**, arXiv:1603.04467.
36. Acharya, S.K.; Wi, Y.-M.; Lee, J. Weather data mixing models for day-ahead PV forecasting in small-scale PV plants. *Energies* **2021**, *14*, 2998. [CrossRef]
37. Pattanayek, S. *Pro Deep Learning with TensorFlow*; Apress: Bangalore, India, 2017.