IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 21, 2020, accepted September 27, 2020, date of publication October 1, 2020, date of current version October 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028281

Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting

MUSAED ALHUSSEIN, KHURSHEED AURANGZEB, (Senior Member, IEEE), AND SYED IRTAZA HAIDER

Computer Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Khursheed Aurangzeb (kaurangzeb@ksu.edu.sa)

The authors extend their appreciation to the Deputyship for Research & Innovation, ''Ministry of Education'' in Saudi Arabia for funding this research work through the project number IFKSURG-1438-034.

ABSTRACT Power grids are transforming into flexible, smart, and cooperative systems with greater dissemination of distributed energy resources, advanced metering infrastructure, and advanced communication technologies. Short-term electric load forecasting for individual residential customers plays a progressively crucial role in the operation and planning of future grids. Compared to the aggregated electrical load at the community level, the prediction of individual household electric loads is legitimately challenging because of the high uncertainty and volatility involved. Results from previous studies show that prediction using machine learning and deep learning models is far from accurate, and there is still room for improvement. We herein propose a deep learning framework based on a combination of a convolutional neural network (CNN) and long short-term memory (LSTM). The proposed hybrid CNN-LSTM model uses CNN layers for feature extraction from the input data with LSTM layers for sequence learning. The performance of our developed framework is comprehensively compared to state-of-the-art systems currently in use for short-term individual household electric load forecasting. The proposed model achieved significantly better results compared to other competing techniques. We evaluated our proposed model with the recently explored LSTM-based deep learning model on a publicly available electrical load data of individual household customers from the Smart Grid Smart City (SGSC) project. We obtained an average mean absolute percentage error (MAPE) of 40.38% for individual household electric load forecasts in comparison with the LSTM-based model that obtained an average MAPE of 44.06%. Furthermore, we evaluated the effectiveness of the proposed model on different time horizons (up to 3 h ahead). Compared to the recently developed LSTM-based model tested on the same dataset, we obtained 4.01%, 4.76%, and 5.98% improvement for one, two, and six look-forward time steps, respectively (with 2 lookback time steps). Additionally, we have performed clustering analysis based on the power consumption behavior of the energy users, which indicate that prediction accuracy could be improved by grouping and training the representative model using large amount of data. The results indicated that the proposed model outperforms the LSTM-based model for both 1 h ahead and 3 h ahead in forecasting individual household electric loads.

INDEX TERMS CNN, deep learning framework, energy consumption, energy consumption forecasting, individual household, LSTM.

I. INTRODUCTION

Short-term electric load forecasting is a vital part of the energy sector since it concerns the forecasting of power consumption in the subsequent few hours. The accurate prediction of the load can significantly support the operations, maintenance, and management of the power system. Energy cannot be stored in considerable quantity, which means that there must be a fair balance between the generation and demand [1]. For the accurate and effective scheduling of power operations, a precise prediction of the power generated and the power load is critical. Subsequently, it is crucial to design and develop scheduling plans for efficient usage of the power available. Moreover, forecasting errors have

The associate editor coordinating the review of this manuscript and approving it for publication was Hayder Al-Hraishawi.

a considerable influence on the safety check, the dynamic state estimation, and power load dispatching of the power grid [2], [3]. To support both system operability and planning, power distribution companies rely on accurate forecasts of generation and consumption with different time horizons.

The integration of the information and communication technology (ICT) and advanced metering infrastructure (AMI) in the traditional power grid (TPG) results in its transformation into a smart grid (SG) that enables bi-directional communication between consumers and the utility. By integrating the ICT in power grids, it is possible to monitor and optimize power generation, power distribution, and power consumption. Due to intelligent techniques and ICT, the SGs empower its consumers with reliable, economical, sustainable, secure, and efficient energy. Demand-side management (DSM) technology applied in SGs enables efficient load utilization by shifting end customer load from peak hours to off-peak hours, helping both in cost reduction and energy management of the power grids. The opportunity of a two-way communication flow between utility and consumers empowers the optimization of energy consumption, which helps in refining the management and operation of the power system [4]–[8].

The forecast for energy consumption over time allows individual customers to assess their consumption habits and, whenever possible, to shift their energy use to off-peak periods. Accurate prediction of energy consumption provides energy users with the opportunity of relating their current usage pattern with the future expense of their energy. Consequently, these users might take advantage of the forecasting algorithms through awareness of their energy consumption and future projections, and they might be able to manage the expenses of their energy usage more efficiently.

Energy customers play an important role in smart grid demand response and can be divided into three categories: residential, business, and industrial sectors. The residential sector consumes a significant quantity of the total generated energy. The AMI installed in the residential sector is highly helpful in forecasting the short-term power load of the end energy customers [9].

In the past, statistical and machine learning models have been developed for predicting energy generation through renewable resources as well as aggregated load forecasting. These learning models are established on time-series analyses and can be termed data-driven models [10], [11].

Several approaches have been reported in the literature to address short-term electric load forecasting. Very few of them, however, addressed individual households. Recently, a deep learning model based on long short term memory (LSTM) has been developed for short-term individual electric load forecasting [12]. Their proposed model outperforms some well-known machine-learning methods.

LSTM networks and CNNs are probably the most widely used techniques of deep learning. The main idea of utilizing such models on time-series data is that the LSTM networks are able to capture the sequence pattern information while CNN models are useful in extracting the valuable features and may filter out the noise of the input data. However, the LSTM networks although are designed to work with temporal correlations, they utilize only the attributes provided in the training set, while in contrast CNNs although are used to extract patterns of local trend as well as the same pattern which appears in different region of time-series data, they are not usually adapted for long temporal dependencies. Therefore, a hybrid model that exploits the benefits of both deep learning techniques could improve the forecasting accuracy.

In this paper, we propose a CNN-LSTM model that utilizes the ability of convolution layers to learn the internal representation of time-series data and obtain the important attributes as well as the usefulness of LSTM layers to identify short-term and long-term dependencies. The proposed model was developed and evaluated on real-world load consumption data of various individual households from the Smart Grid Smart City (SGSC) project funded by the Australian Government [13]. We compared the performance of the proposed model with existing state-of-the-art methods in individual household load forecasting. The effectiveness of the proposed model was further evaluated for various time horizons. The results show that the exploitation of convolutional layers along with LSTM layers could provide a significant improvement in the accuracy of individual household load forecasting.

The major contributions of this paper are: (1) developing a hybrid CNN-LSTM model, which can exploit the benefits of convolutional layers and LSTM layers; (2) Illustrating the effectiveness of the proposed model in individual household load forecasting in comparison with existing state-of-the-art methods; (3) validating the efficacy of the proposed model for various time horizons; (4) investigating the clustering behavior by grouping customers with structural similarity in their load profiles.

The remainder of the paper is organized in the following way. Section II provides a literature review of short-term load forecasting. Data analysis and problem formulation are provided in Section III. The proposed CNN-LSTM model is presented in Section IV. Finally, the results and discussions are elaborated in Section V.

II. RELATED WORK

A lot of research work has been conducted in the field of short-term power load forecasting. Previously, conventional statistical analysis techniques were used for such time-series analyses. Recently, with the enormous progress in the fields of artificial intelligence and machine/deep learning, researchers have developed various deep learning models for load forecasting problems.

Artificial neural networks (ANNs) have been effectively used for short term load forecasting at an industrial scale due to their nonlinear mapping attributes [14]. The main issue with ANN-based forecasting models is that these models can easily be stuck into local minima, which causes poor generalization. Moreover, the forecasting frameworks based on ANN models can be over-fitted, and their convergence rate is slow [15].

The forecasting problems can also be solved by applying other machine-learning models such as generalized regression neural networks (GRNNs) [16], support vector machines (SVMs) [17], extreme learning machine neural networks (ELMNNs) [18] and Kernel-based Support Vector Quantile Regression [19]. The prediction accuracy of the ELMNN is intensely reliant on the applied activation function. The unsystematically chosen activation function will cause poor generalization [20]. Moreover, it is not suitable for prediction problems that require deep extraction of features as it cannot encode the sequence of layers (it can encode one layer only). The GRNN model is computationally much more complex, which makes it inappropriate for such forecasting issues [21]. The attributes of all of these models, including large memory space requirements, high computational complexity, and optimal choice of a kernel of SVM based models, make them inappropriate for such forecasting problems [22].

Most of the developed models for short-term load forecasting focus on aggregated load forecasting [23]–[30]. For supporting future smart grid applications, short-term power load forecasting of individual energy customers is gaining increasing interest, a subject that has been targeted by very few researchers in the recent past.

The authors in [31] considered the functional time series approach to examine the individual household load forecasting. Their evaluation is based on the root mean square error (RMSE). In [32], the Kalman filter is used to estimate the load of the individual household for various time horizons and sampling periods. They argued that the chosen sampling rate provides a compromise between accuracy and computational complexity.

The authors in [33] applied SVM and ANN methods on high-resolution data collected over thirty days from three houses. They obtained considerable improvements in the mean absolute error of (4% – 33%). In [34], the authors proposed an approach based on the activity sequence and support vector regression. They concluded that the activity sequence variable is an impelling factor that could enhance the accuracy of individual household load forecasting for a time horizon of fifteen minutes ahead.

The authors in [35] explored several forecasting models, such as neural networks, ARIMA, and exponential smoothing for horizons ranging from 15 min to 24 h. They evaluated the developed models using two data sets. One dataset was from six households in the United States, while the other was from a single household in Germany. They obtained an average mean absolute percentage error (MAPE) of 85% and 30% for the data sets from the United States and Germany respectively.

In [36], the authors applied several different models, including SVM, classification and regression trees, and multilayer perceptron neural networks. They concluded that a combination of household behavioral data and historical electricity usage data from individual households could significantly improve the forecasting accuracy. They achieved a MAPE of 51% and 48 % for the neural networks and SVM, respectively. In our previous works in [37]–[40], we have developed several efficient models for power load forecasting.

In [12], the authors proposed a deep learning model based on LSTM for short-term residential load forecasting. They compared their model with state-of-the-art machine learning models as well as empirical models. Their proposed model outperformed all the rival techniques and achieved an average MAPE of 44.06% for short-term residential load forecasting of 69 customers.

Based on these recent explorations conducted by valuable researchers, there is a vibrant and increasingly understandable research tendency that looks at challenges related to behavioral and other factors that have an impact on the energy consumption of the individual household. The motivation is to observe and get feedback on power usage patterns of each particular household energy user and determine important fundamental relations between the contextual issues such as time of use, day of the week (weekday or weekend), season, etc. It is anticipated that the insights through such explorations may enhance the understanding and awareness of household energy consumption, which will lead us to better usage of the electricity.

Our proposed method not only outperforms [12] for the next time step forecasting but also achieved much better results for load forecasting of up to 3 h ahead. In [12], the authors forecasted the power load value after the next 30 minutes (look ahead). In our analysis, we included simulations for up to the next three hours. Additionally, we have performed clustering analysis based on the power consumption behavior of the energy users, in order to analyze whether the prediction accuracy could be improved by grouping users of similar energy profile.

III. DATA ANALYSIS AND PROBLEM FORMULATION

Individual household load forecasting is quite challenging because hourly consumption of electricity depends on several factors, such as the number of persons living in each household, the number of major appliances running at a particular time, weather conditions, economics, lifestyle, and daily routines, etc. An individual household load can lack a stable pattern and fluctuate even in consecutive hours. On the contrary, forecasting the aggregated power load at the community or utility level is comparatively easy [40]. The diversity in the aggregated power demand smooths daily power load shapes, which make relative forecasting errors quite low in terms of MAPE.

In this work, we have used the data gathered during the SGSC project initiated by the Australian Government [13]. The SGSC collected the power consumption data for about 10,000 customers in Australia. Short-term individual households electric load forecasting is one of the research areas that can utilize the data gathered during this project.

Our aim in this work is to forecast the power demand of a group of general individual customers. Therefore, it was unrealistic to consider all the customers available in the SGSC database. For the demonstration of the proposed method to make a one-on-one comparison with [12], we selected a subset of the SGSC dataset, the customers who owned a hot water system. Based on this selection criterion, we separated a reasonably-sized subset of data, which corresponded to 69 customers.

Below, we present a brief analysis of the power load profiles of some of the individual households. For evaluating the regularity in daily power consumption profiles, we applied the well-known density-based clustering technique called 'density-based spatial clustering of application with noise' (DBSCAN) [41]. The advantage of using the DBSCAN for regularity analysis in a power consumption profile is that it does not need cluster information in the dataset. Additionally, it includes outliers in the dataset. Generally, the power consumption behavior of the customers will be repeated during weekdays, which makes the DBSCAN an ideal clustering technique for identifying outliers in the dataset of daily power consumption. If the outcome of the DBSCAN shows a low number of outliers, it means that the regularity in power consumption behavior is high.

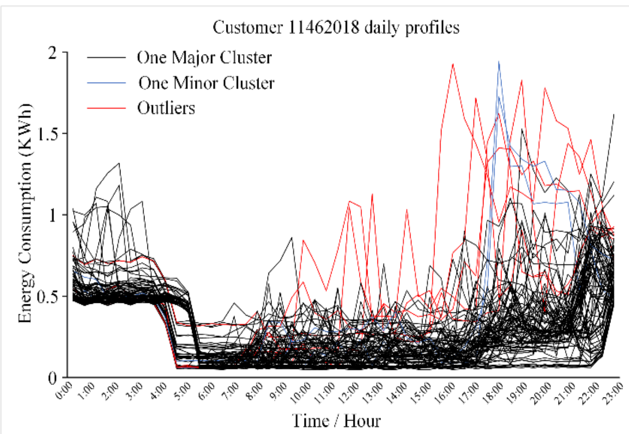

FIGURE 1. The 92 curves of Customer 11462018 grouped into one major cluster, one minor cluster, and five outliers.

Figure 1 shows the half-hourly power consumption profile of randomly selected customer (customer ID 11462018) for the considered period of 92 days. It can be inferred from the figure that the behavior of this customer varies over a span of three months and makes one major cluster, one minor cluster, and some outliers.

Figure 2 shows the number of major and minor clusters along with the number of outliers in the power consumption of the daily profiles of a few randomly selected customers. The independent axis shows customer identification (ID), while the dependent axis indicates the number of major/minor clusters and outliers. As seen in the figure, the number of major/minor clusters and outliers vary from customer to customer.

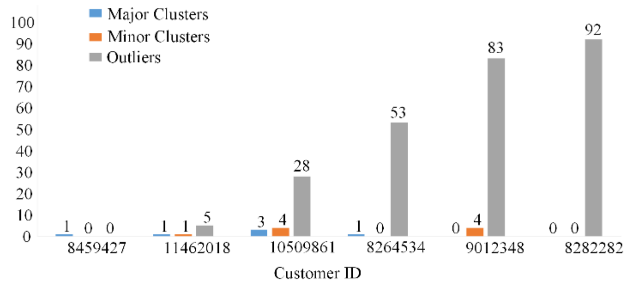

FIGURE 2. Number of major/minor clusters and outliers for a few randomly selected customers.

For instance, customer ID 8459427 has only one major cluster with no outliers, which signifies that the household has a regular power consumption pattern and could be easily predictable. On the contrary, customer ID 8282282 has all outliers with no clusters. This customer exhibits highly volatile power consumption behavior and is hence very difficult to predict accurately.

For some customers like (customer ID 10509861), there exist more than one prominent pattern in daily profiles. For such customers, it is difficult to apply the commonly used forecasting schemes that are based on features such as time of the day, any day of the week.

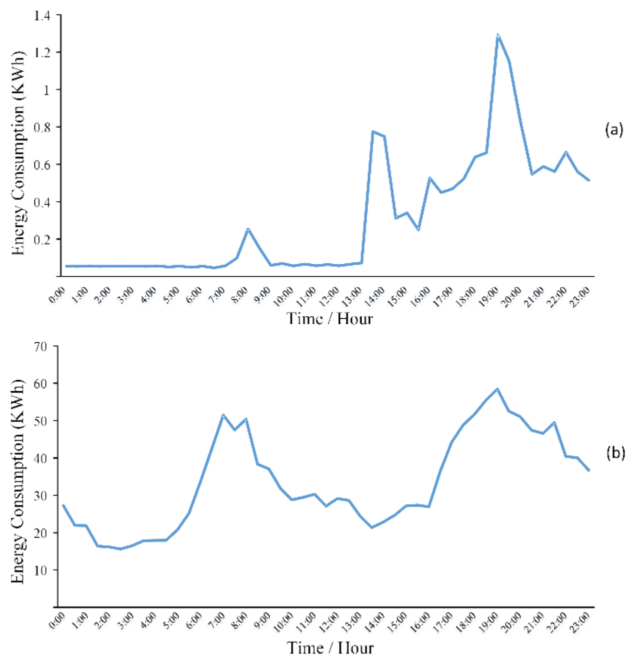

FIGURE 3. Electricity load profile across a 24 h period on 5th June 2013 based on data [13]; (A) for an individual customer (customer ID 8198267), (B) aggregated for 69 customers.

The load profile for a randomly selected day of 5th June 2013 for an individual customer (customer ID 8198267) as well as the aggregated load of 69 customers for the same day are shown in Figure 3. It is evident from this figure that compared to the power profile of an individual customer, the aggregated power profile of 69 customers have smooth variations. The individual customer profile shows a peak in the evening between 07:00 p.m. and 07:30 p.m. and a second peak that is smaller than the peak in the evening, between 01:30 p.m. and 02:00 p.m. On the contrary, the aggregated power profile shows a distinct peak in the morning and the second distinct peak in the evening.

The above analysis signifies that the load forecasting of an individual customer is challenging due to the abrupt variations in daily profiles. For every individual customer, a deep learning model should be trained and evaluated on his own set of data. This means that for 69 different customers, 69 different models should be trained and tested on their respective datasets.

The CNN and the LSTM are the most commonly used machine learning models. Our main purpose in designing the hybrid model of CNN and LSTM layers is to exploit their characteristics for developing an efficient model for load forecasting of the individual household. The individual household load is a time-series data for which we chose the LSTM layers because of their capability to extract the sequence pattern information as well as short-term and long-term dependencies. On the other hand, the CNN layers are employed due to their capability of extracting the valuable features embedded in the time series data. Additionally, the CNN layers are helpful in filtering out the noise of the input data. Consequently, a hybrid model that exploits the benefits of both CNN and LSTM is expected to enhance the load forecasting accuracy of the individual household.

IV. THE PROPOSED HYBRID DEEP LEARNING FRAMEWORK

In this section, we discuss the attributes considered for the developed model. Next, we describe the architecture of the proposed CNN-LSTM model. The convolutional layers are used to extract the valuable features from the input data while LSTM layers are used to exploit short-term and long-term dependencies.

A. FEATURE PREPARATION

In this work, we study the energy data from the SGSC project. The raw dataset acquired using smart-meters contains half-hourly load consumption of residential customers measured in kilowatt-hour (kWh). We used the most commonly exploited features in load forecasting literature to obtain the attributes from the data. The electricity load for residential customers can vary considerably across different hours of a day as well as various days of the week, therefore features such as an hour of the day, day of the week, and holiday indicator are considered. The seasonal impact is excluded from the analysis as the data only corresponded to the winter season in Australia.

The input feature vector shown in Figure 4 is composed of the following attributes:

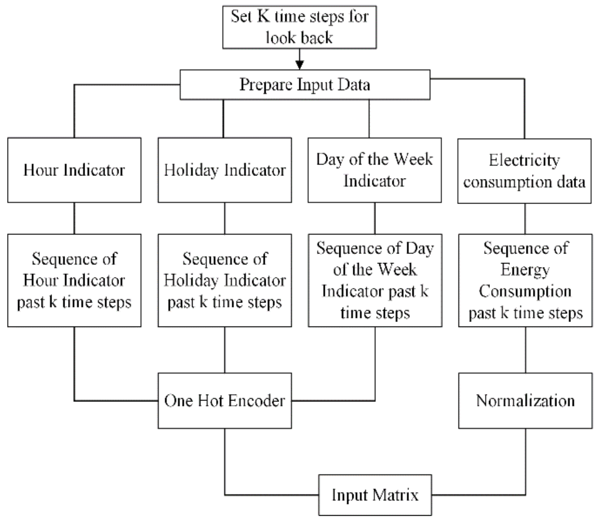

FIGURE 4. Data preparation steps.

1) The energy consumption sequence E_i for the past K time steps. The energy consumption data was first applied to the Min-Max normalization technique.
2) The one-hot encoded hour indicator T_i which indicates the time of the day for the past K time steps (ranges from 1 to 48)
3) The one-hot encoded day of week indicator D_i for the past K time steps (range of values 0 to 6)
4) The one-hot encoded holiday indicator H_i for the past K time steps (which can be 0 or 1).

B. PROPOSED MODEL ARCHITECTURE

The architecture of the proposed CNN-LSTM based deep learning framework is shown in Figure 5.

The CNN feature extraction block consists of three 1D convolutional layers. We incorporated the MaxPooling layer and Rectified Linear Unit (ReLU) layer in between the two consecutive convolution layers. The convolutional operation is highly effective and piling several convolutional layers in a deep learning framework enables the initial layers to learn low-level features in the applied input. The feature map, which is the output of the convolutional layers has a limitation that it keeps track of the precise location of the features in the input. It means that little movements in the location of the feature in the input will lead to a different feature map. A pooling layer is usually added after the convolutional layer for mitigating the limitation of the invariance of the produced feature map whereas the activation function is applied for enhancing the capability of the model for learning complex structures. In our developed model, we have added a MaxPooling layer which is a down-sampling scheme that reduces the spatial dimension of the feature maps by a factor of 2, hence reduces the overall computational load. The ReLU activation function is resilient against the gradient vanishing problem and has been widely implemented by various researchers to make the network more trainable.

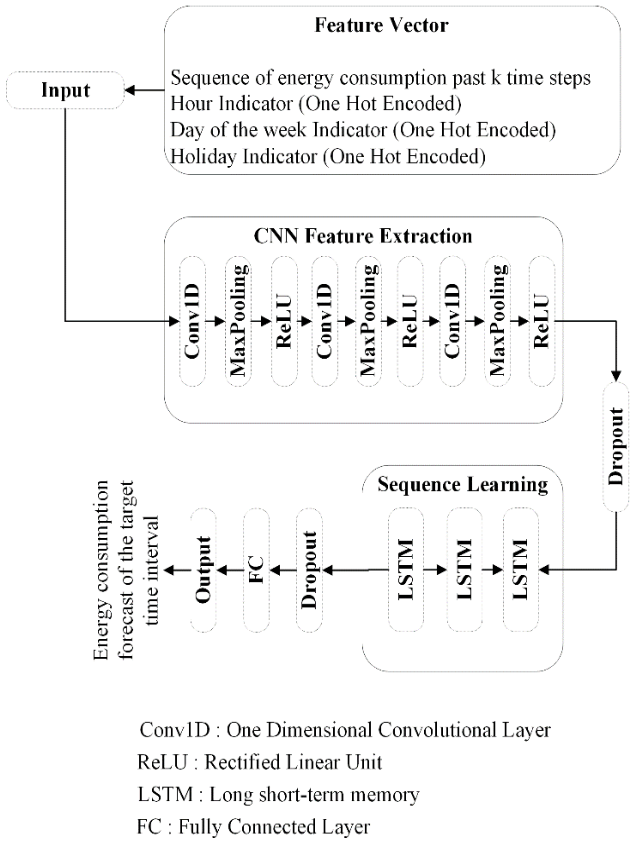

FIGURE 5. Proposed deep CNN-LSTM framework.

In the development of any deep learning model, the dropout layer offers a cool way to relieve the overfitting issue. This layer includes the random selection of neurons and deactivating some of them in the training process. In this work, we have incorporated a dropout layer between the CNN feature extraction block and the LSTM sequence learning to prevent overfitting. The output of the sequence learning block is connected to a dropout layer, followed by a fully connected layer to produce the final output.

It is a common practice to adopt a coarse-to-fine approach when developing a CNN model. This structure introduces higher computational complexity as it involves a large number of trainable parameters. We chose a pyramid architecture, as discussed in [43] when developing our CNN feature extraction block, where the number of kernels is large in the lower-level layers, which are gradually decreased by a constant as we go down the higher-level layers. The configuration of various layers of the proposed model are provided in Table 1. We select a kernel size of 48 for the first convolution layer, which is reduced to 32 and 16 for the second and third convolution layers. This type of structure avoids overfitting and reduces the number of trainable parameters.

In the sequence learning block, we used three LSTM layers with 20 neurons each. The return sequence is set to true for the first two LSTM layers so that the network will output the full sequence of hidden states whereas, in the final LSTM layer, the return sequence is set to false so that the network will output the hidden state at the final time step. We used the dropout layer before the fully connected layer to avoid overfitting. The fully connected layer has 20 neurons. The number of neurons in the output layers are varied from one to six for evaluating the different number of lookahead (up to 3 h ahead load forecasting).

TABLE 1. Configuration of the various layers of the proposed model.

Proposed Model		
Convolution1	Kernels	48
	Size of Receptive Field	3
MaxPooling	-	-
ReLU	-	-
Convolution2	Kernels	32
	Size of Receptive Field	3
MaxPooling	-	-
ReLU	-	-
Convolution3	Kernels	16
	Size of Receptive Field	3
MaxPooling	-	-
ReLU	-	-
Dropout1	-	0.25
LSTM1	Hidden Nodes	20
	Return Sequence	True
LSTM2	Hidden Nodes	20
	Return Sequence	True
LSTM3	Hidden Nodes	20
	Return Sequence	False
Dropout2	-	0.25
Fully Connected	Hidden Nodes	20
Output	Hidden Nodes	1 / 2 / 6

The parameter settings of the developed deep learning framework are presented in Table 2. In this work, we have used a well-known optimizer 'Adam' and mean absolute error as a loss function.

TABLE 2. Parameter settings of the developed model.

Parameter	Setting
Optimizer	Adam
Loss Function	Mean Absolute Error (MAE)
Learning Rate	{0.001}
Learning Rate Adjustment	Monitor = validation loss, patience = 10 Epochs, factor = 0.8, minimum learning rate = 1e-5
Batch Size	{128}
Epoch	{150}

Figure 6 shows the training flow of the proposed deep learning model. The input data is split into 70% training, 20% validation, and 10% test data. We used mean absolute error (MAE) as a loss function to monitor the validation loss. Initially, the training data and the validation data are loaded,

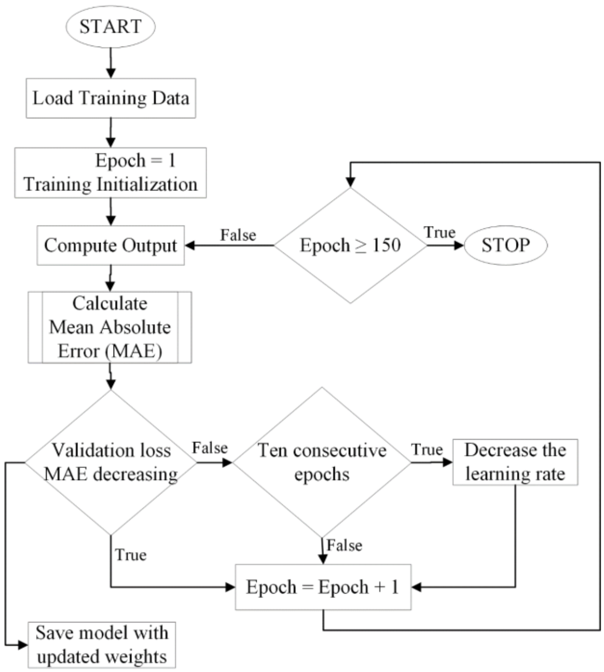

FIGURE 6. Training flow of the proposed model.

and the training process is initialized. After completing each epoch, the validation loss is determined and checked to see if it is decreasing. If the validation loss is decreasing, then the model is saved with the updated weights, and the epochs are incremented. However, if the validation loss is not decreasing for ten consecutive epochs, then the learning rate is decreased, and the epochs are incremented. The training stops when the epoch count reaches 150. We load the last saved best model for prediction and evaluation on the test data to avoid overfitting.

V. RESULTS AND DISCUSSIONS

In this study, we chose a pool of 69 customers out of thousands of customers' data. The data were retrieved from the SGSC project initiated by the Australian Government [13]. The 69 customers were chosen based on the criterion of energy users having a hot water system. Some customers had more than one year of data available, while some had only a few months of data available. We selected data starting from June (01 June 2013) until the end of August (31 August 2013), when all of the customers' data was available. We excluded the seasonal impact from our analysis as the data only corresponded to the winter season in Australia. For each customer, the data spanned 92 days. We partitioned the data into training, validation, and test data as 70%, 20%, and 10%, respectively.

The developed model was experimented with different configurations of various lookback time steps to include two, six, and twelve. In addition, the model was evaluated for a different look ahead/forward time steps such as half-hour, one hour, and 3 h corresponding to one, two, and six look forward time steps.

TABLE 3. Attributes of the data set used for training and testing of the developed model.

Attribute No.	Description	Formula
1	Energy Consumption sequence	E
2 – 49	Time of the day Indicator (dummy variable)	Ti, i = 1, 2, .., 48
50 – 56	Day of the week Indicator (dummy variable)	Di, i =1, 2, .., 7
57 – 58	Holiday Indicator (dummy variable)	Hi, i = 0 / 1

The various attributes of the data, including the energy consumption value, the time of the day indicator, the day of the week indicator, and the holiday indicator, are organized column wise. The numbering of the various attributes of column-wise data is shown in Table 3.

A. SINGLE-STEP FORECAST

In this section, we investigated the comparison of various machine learning and deep learning models with the proposed model for a single-step forecast, i.e., to predict the value at the next time step.

In [12], the authors evaluated various machine learning models including the ELM [27], KNN [28], [29], back-propagation neural network (BPNN), and the input selection combined with hybrid forecasting (IS-HF) [30] on the same dataset. The comparison of the MAPE results for various machine learning models, including the deep learning model based on LSTM [12] with the proposed hybrid model for different lookback time steps, is shown in Table 4.

We did not re-evaluate all of the machine learning models, and empirical methods mentioned in Table 4 solely for comparison. However, we did take the MAPE values of these methods from Table 2 of [12]. The LSTM based model achieved an average MAPE of 44.06% for a look back of 12-time steps. As concluded in [12], the average MAPE of the deep learning model based on the LSTM architecture is better than those of the other machine learning models as well as empirical methods for various lookback time steps. Based on this analysis, we selected the deep learning model based on the LSTM architecture for evaluating the performance of the proposed hybrid CNN-LSTM model. It can be inferred from Table 4 that the proposed hybrid CNN-LSTM model achieved an average MAPE of 40.38% with a look back of two-time steps. On the contrary, the model based on LSTM achieved the best result with a look back of twelve-time steps. The proposed method achieved better forecasting performance as compared to many state-of-the-art methods including LSTM based approach.

The comparison of the proposed hybrid CNN-LSTM model with the LSTM based deep learning model [12] is shown in Figure 7. The independent axis in this figure shows the varying number of outliers while the dependent axis shows the MAPE. It can be observed in this figure that the

TABLE 4. Comparison of the proposed deep learning framework with other models.

Method	Scenario (Lookback)	Average MAPE Individual Forecasts (%)
CNN-LSTM	2-time steps	**40.38**
	6-time steps	41.07
	12-time steps	42.85
LSTM	2 time steps	44.39
	6-time steps	44.31
	12-time steps	**44.06**
BPNN	2-time steps	49.62
	6-time steps	49.04
	12-time steps	49.49
KNN	2-time steps	74.83
	6-time steps	71.19
	12-time steps	81.13
ELM	2-time steps	122.90
	6-time steps	136.49
	12-time steps	123.45
MAPE Minimization	-	46.00
IS-HF	-	96.76
Empirical Mean	-	136.46

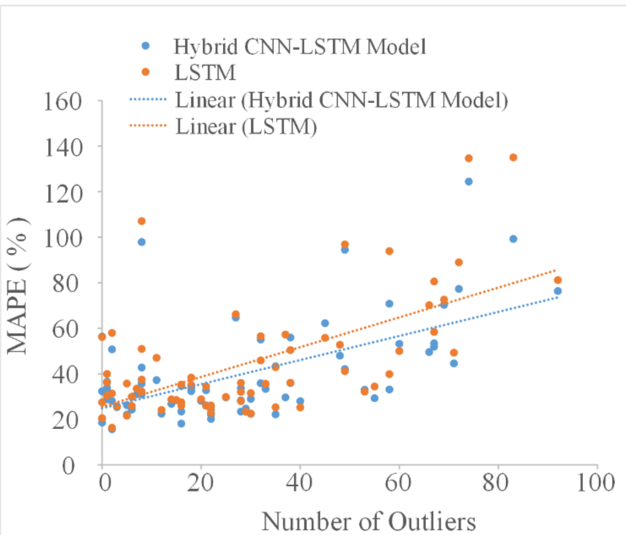

FIGURE 7. MAPE vs. no. of outliers for the proposed hybrid model and the LSTM model.

MAPE of the proposed hybrid model is lower than that of the LSTM model. For a look back of 2-time steps, the hybrid CNN-LSTM model performs the best (lower MAPE value) for 57 out of 69 households, whereas the LSTM is the best predictor for 12 households. It is pertinent to note that as the number of outliers increases, forecasting energy consumption became more challenging for both the LSTM based model and the proposed hybrid approach. However, the proposed model improves the overall average MAPE of individual household energy consumption forecasting. This improvement is more noticeable when the outliers are relatively large.

B. MULTI-STEP FORECAST

In this section, we evaluated the effectiveness of the proposed model for different time-horizons, i.e., multi-step forecasting. For each instance of test data, the proposed model was assessed such that if the time horizon is set to 3 h ahead, then the model will forecast the next six values. We compared the proposed hybrid CNN-LSTM model with the LSTM based model for multi-step forecasting. Table 5 shows the average MAPE for the individual household forecasts for different lookback and look-forward time steps.

TABLE 5. The comparison of proposed model with the LSTM model for various lookback and look-forward configurations.

Lookback (time steps)	Look forward (time steps)	Average MAPE Individual Forecasts (%)		Percentage Improvement (%)
		Proposed	LSTM	
2	1	40.38	44.39	4.01
	2	46.98	51.74	4.76
	6	59.26	65.24	5.98
6	1	41.07	44.31	3.24
	2	46.73	53.66	6.93
	6	59.86	67.24	7.38
12	1	42.85	44.06	1.21
	2	48.05	56.08	8.03
	6	61.91	69.55	7.64

It is evident from this table that for all the lookback and look-forward time steps, the average MAPE of the proposed model is lower than that of the LSTM. It is observed that the percentage improvement in terms of MAPE is slightly decreasing when the lookback is varied from 2-12 time steps. This trend is only observed for a look-forward of the next time step which may be due to the LSTM layers learning more from the lag features. Other than that the gap between the MAPE values of the proposed hybrid model and the LSTM model gets wider for increasing look-forward time steps. This signifies that the proposed hybrid model not only outperforms the LSTM based model for a single-step forecast but the multi-step forecast as well. Compared to the LSTM based deep learning model [12] for two lookback time steps, with our developed model, we obtained 4.01%, 4.76%, and 5.98% improvement for one, two, and six look-forward time steps respectively.

The proposed model and the deep learning model based on LSTM are applied to the test data, and the results are presented in Figure 8 for three different customers. The independent axis shows the time instances of the different days of the test data. By carefully observing different parts

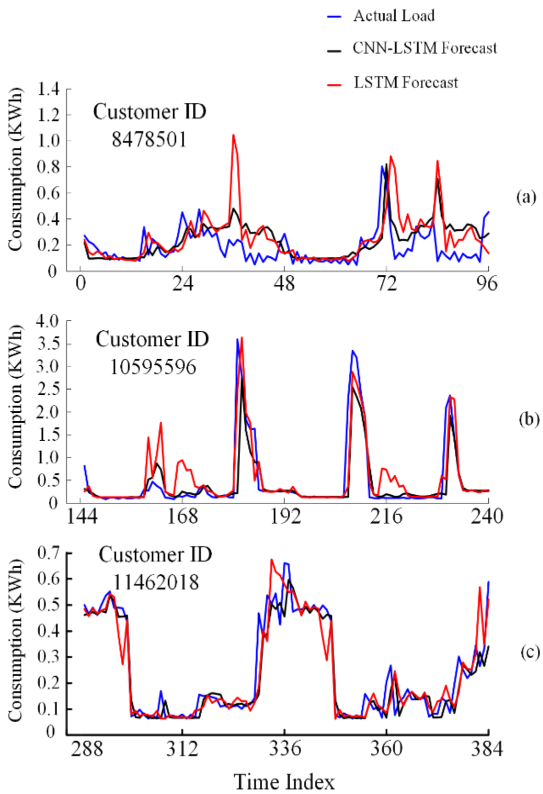

FIGURE 8. Forecast results of three customers for the proposed and the LSTM models.

of this figure, we can see that a deep learning approach based on LSTM only has unreasonably high peaks at various instances of time. On the contrary, our proposed model, which is based on a combination of CNN and LSTM quite closely follow the actual original load at almost all the time instances. Additionally, the LSTM based model also shows some peaks at time instances '168' and '216' in Figure 8 (b) where there are no peaks in the actual load. In part (c) of Figure 8, there are some peaks at time instances ahead of the real peaks. Based on all these observations, we can conclude that the proposed CNN-LSTM based deep learning framework is an efficient approach for forecasting individual household power consumption.

C. CLUSTERING ANALYSIS

The research on household clustering has shifted its focus from attributes-oriented factors to consumption-pattern-oriented factors. This shift is mainly due to the use of smart meters that provides high-resolution power consumption time-series data.

In this section, we have used the k-mean clustering technique to group similar households based on their consumption pattern, i.e. load profiling. For clustering analysis, we have considered the training data of the same pool of 69 customer that were used in the single and multi-scale forecast in Section V. For our analysis, we have divided a day into four periods that are described below:

1) Breakfast period: 6.30 AM – 9.00 AM
2) Daytime period: 9.00 AM – 3.30 PM
3) Evening period: 3.30 PM–10.30 PM
4) Overnight period: 10.30 PM – 6.30 AM

The different attributes of the considered data are explained below:

Attributes 1 – 4: For each time period, the relative mean power using $P_{ti} = \sum P/P_d$ where numerator is the sum of power for each time interval t_i and P_d is the daily mean power over the entire training data.

Attribute 5: The mean relative standard deviation using $\sigma = 1/4 \sum_{i=1}^{4} \sigma_i/P_i$ over the entire training data, where σ_i and P_i corresponds to standard deviation and average power in each time period respectively.

Attribute 6 – 9: For each time interval, weekend versus weekday difference score using $W_{ti} = \left|P_i^{Weekday} - P_i^{Weekend}\right|$ where $P_i^{Weekday}$ and $P_i^{Weekend}$ are the average power during the weekdays and weekend respectively over the entire training data.

For each household, the above mentioned nine attributes summarize their consumption behavior for the training data. In order to find the optimal number of clusters, different alternatives have been explored by the researchers, and the silhouette score analysis is one of the appropriate technique. The silhouette index ranges from -1 to +1 and is an indicator of how closely similar an object is to the rest of the members of its cluster. A higher value of the silhouette index (close to +1) shows that an object is highly similar to other members of its own cluster and is highly dissimilar to the objects of other clusters.

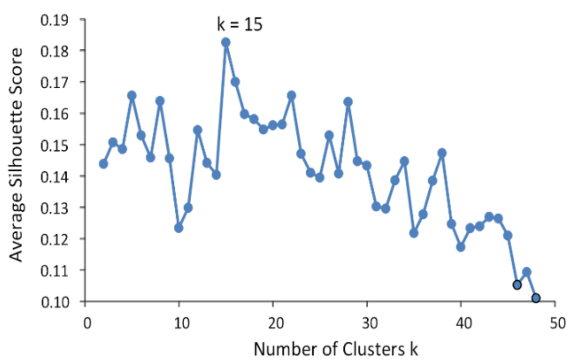

FIGURE 9. Average silhouette score based on nine attributes of 69 customers.

The k-means clustering is iteratively performed on the above mentioned nine attributes of the 69 households from 2 to 50 clusters and the result is shown in Figure 9. As shown in the figure, silhouette score is maximum at k equal to 15 and decreases afterwards. The high-resolution time-series load profiles of households often carry comprehensive attributes. It is common to find larger number of clusters, i.e. more than 10 clusters (e.g. [45]–[47]).

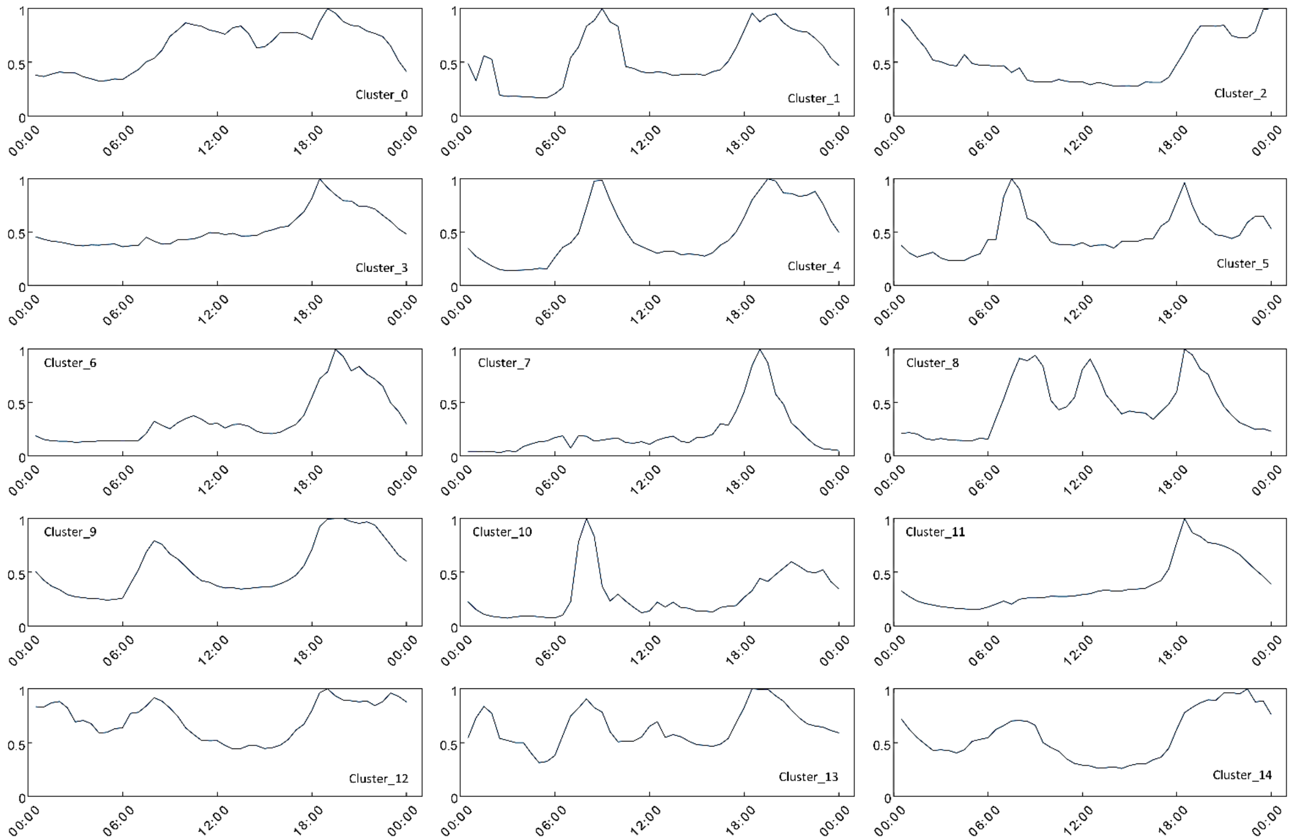

FIGURE 10. The average power consumption patterns of the fifteen clusters.

The average power consumption patterns of the fifteen clusters over the entire day is shown in Figure 10. For each cluster, only the training and validation data (83 days out of 92 days) of their representative households are considered. The average power consumption is normalized with the maximum value in each cluster. The resulting clusters show some differences from each other. For example, Cluster_1, Cluster_8, Cluster_12, and Cluster_13 has three distinct peaks, the main difference between them is the time. Cluster_1 has a smaller overnight peak whereas Cluster_8 doesn't have an overnight peak. Cluster_12 and Cluster_13 don't have a distinct peak in the daytime. Cluster_3, Cluster_6, Cluster_7, and Cluster_11 have only one distinct peak in the evening time.

However, the difference between them is the average power consumption during overnight, breakfast, and daytime as well as the width of the peak span. Cluster_2 exhibits a decreasing trend during the day-time whereas average power consumption drastically increases in the late evening. Cluster_4, Cluster_5, Cluster_9, and Cluster_14 have two distinct peaks in the breakfast and in the evening. However, Cluster_4 has a wider peak span as compared to Cluster_5. Moreover, Cluster_9 has a low average power consumption during the night time whereas a sharp rise in the early breakfast.

In the previous section, we compare the individual household forecasting performance of the proposed model

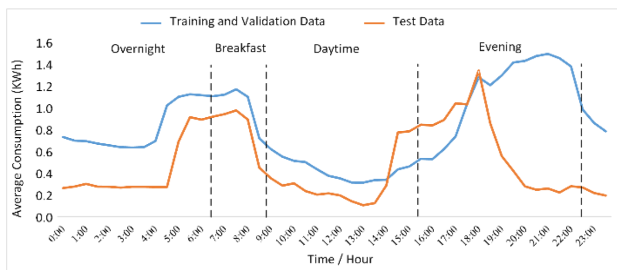

FIGURE 11. The average hourly power consumption of customer id 8680284.

with the LSTM model, where the average MAPE of the LSTM model was obtained from [12] as mentioned in Table 4. In order to make a one-to-one MAPE comparison between the proposed model and the LSTM model for all the households, we have implemented the LSTM model presented in [12]. The obtained average MAPE of 44.68% for the LSTM model is very close to the average MAPE of 44.39% presented in [12]. We present the comparison of the individual forecast using LSTM (third column), and individual forecast using the proposed method (fourth column) in terms of MAPE for each of the 69 households in Table 6.

In this section, we perform the clustering analysis using the proposed method to examine if by grouping the similar households will further improve the forecasting accuracy.

TABLE 6. Compared MAPEs for Sixty Nine Households in Fifteen Clusters with or without clustering.

Cluster No.	Customer ID / No. of Outliers in Training & Validation Data	LSTM without Cluster MAPE (%)	Proposed Method without Cluster MAPE (%)	Proposed Method with Cluster MAPE (%)	Cluster No.	Customer ID / No. of Outliers in Training & Validation Data	LSTM without Cluster MAPE (%)	Proposed Method without Cluster MAPE (%)	Proposed Method with Cluster MAPE (%)
Cluster_0	8804804 / 01	36.15	36.50	41.41	Cluster_9	8451629 / 14	28.76	26.79	31.37
	8350006 / 05	21.58	22.09	21.88		8181075 / 16	27.50	18.15	24.35
	8496980 / 06	29.94	29.85	30.61		8673172 / 16	38.34	33.96	35.55
	8618165 / 07	33.44	31.04	30.04		8617151 / 17	35.09	32.41	32.52
	8328122 / 08	32.25	30.99	28.73		8679346 / 26	66.19	64.71	62.37
	8308588 / 21	26.09	24.33	24.77		8184653 / 28	32.02	28.30	25.65
	8176593 / 30	22.52	22.36	23.69		8211599 / 31	45.86	35.83	46.42
	8566459 / 35	42.87	43.35	42.69		10598990 / 47	52.74	47.94	41.90
	8282282 / 83	81.15	76.39	99.15		11081920 / 52	34.43	29.28	30.49
Cluster_1	8342852 / 01	39.90	33.76	36.28		8376656 / 56	39.87	33.10	43.76
	8482121 / 02	31.36	28.02	39.60		8540084 / 63	72.62	70.30	84.39
	8196671 / 06	25.83	24.27	25.52	Cluster_10	9393680 / 65	88.97	77.29	77.29
	8487461 / 34	50.44	55.97	61.65	Cluster_11	8459427 / 00	20.49	18.55	21.11
Cluster_2	11462018 / 05	35.72	26.26	34.00		8466525 / 03	25.72	25.37	27.66
	8196669 / 11	24.16	22.43	22.86		8478501 / 08	50.94	42.71	37.84
	10692972 / 11	47.06	37.23	37.41		8334780 / 16	26.04	23.41	23.66
	9012348 / 75	135.07	99.25	129.30		8196621 / 16	35.32	34.66	34.87
Cluster_3	8685932 / 02	16.34	15.65	14.75		8419708 / 19	28.84	28.08	24.55
	8257054 / 20	25.62	20.13	21.64		8198267 / 20	34.52	32.76	34.91
Cluster_4	8557605 / 07	37.46	35.74	34.41		8733828 / 23	29.85	29.65	31.72
	8661542 / 15	28.48	28.38	30.04		8347238 / 23	35.98	33.72	36.00
	8196659 / 39	25.21	27.98	38.56		8687500 / 30	35.56	33.32	40.58
	8156517 / 46	96.90	94.51	144.65	Cluster_12	8523058 / 00	56.13	56.39	43.42
	8326944 / 54	93.94	70.81	74.63		8655993 / 00	27.43	32.33	27.67
Cluster_5	10509861 / 25	27.99	23.38	23.54		8198319 / 01	30.44	28.74	36.80
	8273592 / 30	31.51	28.97	28.24		8487297 / 07	107.12	97.88	109.71
	8147703 / 47	41.12	42.08	48.68	Cluster_13	8273230 / 19	25.96	26.16	48.82
Cluster_6	8149711 / 31	56.51	55.09	52.78		8351602 / 22	22.84	22.05	25.17
	8568209 / 37	57.23	29.61	68.91		8504552 / 28	23.34	24.69	27.70
	8145135 / 65	49.28	44.57	48.10		8487285 / 34	25.25	22.16	27.49
Cluster_7	8264534 / 49	32.16	32.98	39.10		10702066 / 37	36.10	35.91	47.67
	8198345 / 54	49.99	53.18	55.37	Cluster_14	8680284 / 02	57.96	50.70	51.39
	8291712 / 68	134.73	124.47	131.34		8432046 / 63	58.42	53.40	64.44
Cluster_8	8519102 / 44	55.82	62.21	53.61		8257034 / 66	80.61	51.91	48.18
	10595596 / 62	70.17	49.55	42.66	Average		**44.68**	**40.38**	**44.76**

We identify patterns in energy usage profiles over four periods, i.e. overnight, breakfast, daytime and evening, and group the households with similar profiles. In this experiment, the training, validation, and test data for each cluster are prepared by concatenating their representative household attributes. For example, to train a single model for Cluster_0, we vertically concatenate the training data attributes mentioned in Table 3 of each representative household. The result of the clustering analysis is presented in column five of Table 6. The green and the red color represents the best and the worst model for each household in Table 6. We observed that some clusters contain very less number of households since the number of households for the analysis is not large enough (only 69 customers).

Column two of Table 6 shows the customer ID along with the number of outliers in daily profiles of training and validation data. For each household, the total number of days in the training (67 days) and validation data (16 days) is 83. The households in each cluster are arranged in increasing order of the number of outliers in their training and validation data. For instance, for Cluster_0, customer id 8804804 has only one outlier whereas customer id 8282282 has all the outliers with no distinct daily profile in the training and validation data.

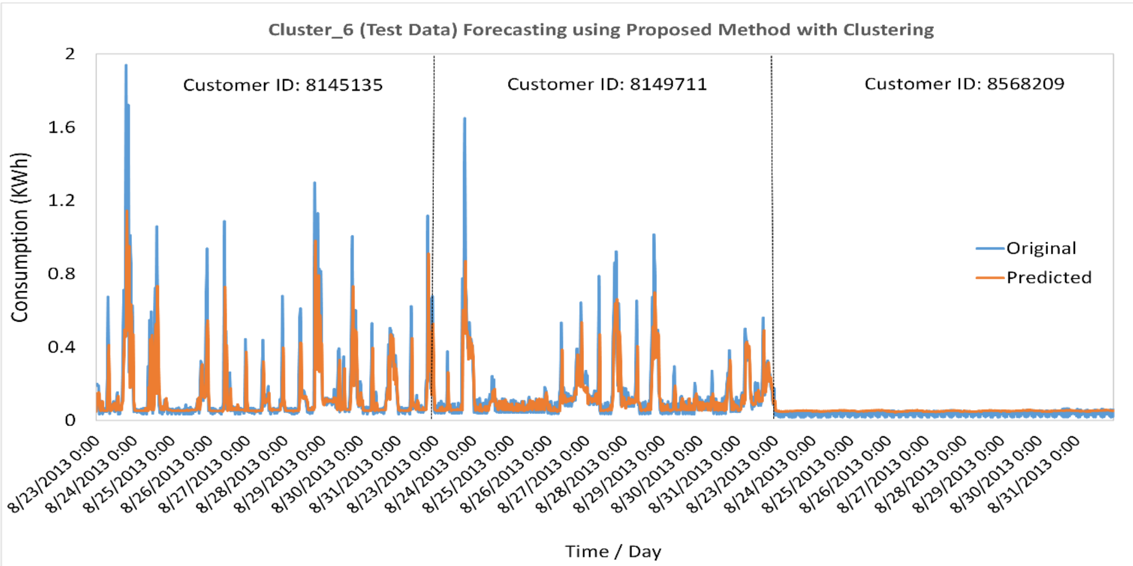

FIGURE 12. Cluster_6 forecasting using proposed method with clustering.

It is evident from the below table that for households with mostly outliers in the training and validation data will result in larger MAPE, e.g. customer id 8282282, 9012348, 8291712, 8540084, and 9393680 in Cluster_0, Cluster_2, Cluster_7, Cluster_9 and Cluster_10 respectively. As shown in the table as we move down the cluster, MAPE is generally increasing. Such a large error is because of the fact that there are no distinct daily profiles of such customers i.e. most of the data are outliers. For such households, the proposed method without clustering achieved better results compared to the individual LSTM based approach as well as the clustering-based approach. One of the reasons why the clustering-based approach may not be performing generally well as compared to the proposed method without clustering could be due to a large number of outliers in various households in different clusters. It results in more variation in the training data in terms of load characteristics that degrade the optimal learning of the model.

Another observation can be made from the table that for some customers the average MAPE is relatively large with very few or no outliers in the training data. For instance, customer id 8680284 in Cluster_14 has a larger value of MAPE even though it has only two outliers in the training data.

Figure 12 shows how the proposed method with the clustering approach forecasts the energy consumption of each household in Cluster_6. As shown in the figure, the prediction curve follows the fluctuation of the original consumption on an hourly basis for all three customers, except for a few peak usage hours for customer id 8145135 and 8149711. This occurs due to the fact that there are random peaks for an individual household.

We observed from the hourly load profile of customer id 8680284 that the average hourly power consumption in the test data was lower than that of training data as shown in **FIGURE 11**Figure 11. It is pertinent to note that for the training and validation data, the average power consumption is gradually rising from around 06:00 pm whereas the peak average consumption occurs in the late evening. However, for the test data, there is a sharp decrease in average consumption after 06:00 pm. The model may predict a peak in the late evening however the actual power consumption is quite low which results in larger MAPE. The proposed method without clustering achieved a MAPE of 50.70 whereas the clustering-based approach achieved second-best MAPE of 51.39 for customer id 8680284.

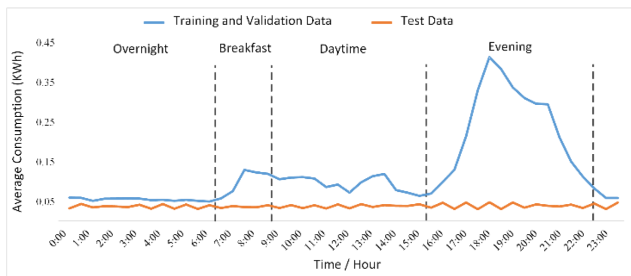

FIGURE 13. The average hourly power consumption of customer id 8568209.

The hourly profile of customer id 8568209 shows that the peak average consumption occurs in the evening for the training and validation data whereas there is no peak in the test data as shown in Figure 13. The prediction curve follows the original consumption for customer id 8568209 as shown in Figure 12, however results in larger MAPE. This occurs due to the fact that the actual value of power consumption for the entire test data is close to zero which results in larger MAPE.

The results presented in Table 6 indicates that the clustering-based approach performs the best (lower MAPE value) for 17 out of 69 households, whereas the LSTM without clustering is the best predictor for only 10 households. The proposed method with clustering approach obtained an average MAPE of 44.76% which is very close to the average MAPE of 44.68% obtained using the LSTM model. The proposed method without clustering outperform the LSTM model as well as the clustering-based approach for most of the customers.

The cluster analysis revealed information on the electricity use pattern of various households in SGSC database. The customers are classified in their energy profiles, depending on their structural similarity. The obtained findings are constrained by the small sample size, i.e., only 69 households. Future study may have greater sample size and more diverse samples.

VI. CONCLUSION

This paper seeks to explore the short-term energy consumption prediction problem of the individual household customers in the residential sector. Load forecasting at an individual household level is quite challenging because it lacks a stable pattern and fluctuates even in consecutive hours.

First, a clustering technique is applied to identify the number of outliers and to discover the regularity in daily power consumption profiles of individual household data. Next, a hybrid model is proposed, which is based on a combination of CNN and LSTM.

The developed framework is tested on a publicly available residential smart meter data from the SGSC project. The performance of the developed framework is comprehensively compared to other state of the art systems in short-term electric load forecasting. The results indicate that the proposed hybrid CNN-LSTM based deep learning framework outperforms the other rival techniques in forecasting individual household energy consumption having both regular and irregular usage behavior.

The prediction problem becomes more challenging for both the LSTM based model and the proposed hybrid approach as the number of outliers increases. However, the proposed model improves the overall average MAPE of individual household energy consumption prediction for both single-step and multi-step forecasting. This improvement is more noticeable when the outliers are relatively large.

The load forecasting model would reveal a better prediction result if parameters such as appliance ownership, sociodemographics data and household occupancy can be detected and added as feature. Future research should focus on further exploring the behavioral characteristics of customers and using that data to the load forecasting model.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

All three authors equally contributed to the manuscript.

REFERENCES

[1] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts," *Energy Buildings*, vol. 84, pp. 214–223, Dec. 2014.
[2] R. Ballini and R. R. Yager, "OWA filters and forecasting models applied to electric power load time series," *Evolving Syst.*, vol. 5, no. 3, pp. 159–173, Sep. 2014.
[3] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using Long-Short-Term-Memory based recurrent neural network," in *Proc. 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2017, pp. 1–6.
[4] S. Aslam, Z. Iqbal, N. Javaid, Z. Khan, K. Aurangzeb, and S. Haider, "Towards efficient energy management of smart buildings exploiting heuristic optimization with real time and critical peak pricing schemes," *Energies*, vol. 10, no. 12, p. 2065, Dec. 2017.
[5] A.-H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.
[6] A. Khalid, S. Aslam, K. Aurangzeb, S. Haider, M. Ashraf, and N. Javaid, "An efficient energy management approach using Fog-as-a-Service for sharing economy in a smart grid," *Energies*, vol. 11, no. 12, p. 3500, Dec. 2018.
[7] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 53–59, Apr. 2011.
[8] N. W. A. Lidula and A. D. Rajapakse, "Microgrids research: A review of experimental microgrids and test systems," *Renew. Sustain. Energy Rev.*, vol. 15, no. 1, pp. 186–202, Jan. 2011.
[9] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Short-term load forecasting in a non-residential building contrasting models and attributes," *Energy Buildings*, vol. 92., pp. 322–330, Oct. 2015, doi: 10.1016/j.enbuild.2015.02.007.
[10] C. Paoli, C. Voyant, M. Muselli, and M.-L. Nivet, "Solar radiation forecasting using ad-hoc time series preprocessing and neural networks," in *Emerging Intelligent Computing Technology and Applications*. Berlin, Germany: Springer, 2009, pp. 898–907.
[11] L. Martín, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning," *Sol. Energy*, vol. 84, no. 10, pp. 1772–1781, Oct. 2010.
[12] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
[13] (2014). *Smart Grid, Smart City, Australian Govern., Australia, Canberra, ACT, Australia*. [Online]. Available: http://www.industry.gov.au/ENERGY/PROGRAMMES/SMARTGRIDSMARTCITY/Pages/default
[14] S. M. Elgarhy, M. M. Othman, A. Taha, and H. M. Hasanien, "Short term load forecasting using ANN technique," in *Proc. 19th Int. Middle East Power Syst. Conf. (MEPCON)*, Dec. 2017, pp. 1385–1394.
[15] S. Bouktif, A. Fiaz, A. Ouni, and M. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, p. 1636, Jun. 2018.
[16] H. Bendu, B. B. V. L. Deepak, and S. Murugan, "Multi-objective optimization of ethanol fuelled HCCI engine performance using hybrid GRNN–PSO," *Appl. Energy*, vol. 187, pp. 601–611, Feb. 2017.
[17] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Appl. Energy*, vol. 168, pp. 568–593, Apr. 2016.
[18] M. Hossain, S. Mekhilef, M. Danesh, L. Olatomiwa, and S. Shamshirband, "Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems," *J. Cleaner Prod.*, vol. 167, pp. 395–405, Nov. 2017.
[19] M. Khan, N. Javaid, S. Javaid, and K. Aurangzeb, "Short term power load probability forecasting by kernel based support vector quantile regression for real-time data analysis," in *Proc. Int. Conf. Innov. Intell. Inform., Comput. Technol.*, Nov. 2018, pp. 1–5.
[20] S. Lin, X. Liu, J. Fang, and Z. Xu, "Is extreme learning machine feasible? A theoretical assessment (Part II)," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 1, pp. 21–34, Jan. 2015.

[21] J. Mareáek, "Usage of generalized regression neural networks in determination of the enterprise's future sales plan," *Littera Scr*, vol. 3, pp. 32–41, Oct. 2016.

[22] H. Bhavsar and M. H. Panchal, "A review on support vector machine for data classification," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 1, no. 10, pp. 1–5, 2012.

[23] X. Cao, S. Dong, Z. Wu, and Y. Jing, "A data-driven hybrid optimization model for short-term residential load forecasting," in *Proc. IEEE Int. Conf. Comput. Inf. Technol.*, Oct. 2015, pp. 283–287.

[24] Z. Yun, Z. Quan, S. Caixin, L. Shaolan, L. Yuming, and S. Yang, "RBF neural network and ANFIS-based short-term load forecasting approach in real-time price environment," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 853–858, Aug. 2008.

[25] H. Li, Z. Zhang, Y. Zhao, and X. Hu, "Short-term load forecasting based on the grid method and the time series fuzzy load forecasting method," in *Proc. Int. Conf. Renew. Power Gener. (RPG)*, 2015, pp. 1–6.

[26] P. Qingle and Z. Min, "Very short-term load forecasting based on neural network and rough set," in *Proc. Int. Conf. Intell. Comput. Technol. Automat.*, vol. 3, May 2010, pp. 1132–1135.

[27] R. Zhang, Y. Xu, Z. Y. Dong, K. Meng, and K. P. Wong, "Short-term load forecasting of australian national electricity market by an ensemble model of extreme learning machine," *IET Gener., Transmiss. Distrib.*, vol. 7, no. 4, pp. 391–397, Apr. 2013, doi: 10.1049/iet-gtd.2012.0541.

[28] R. Zhang, Y. Xu, Z. Y. Dong, W. Kong, and K. P. Wong, "A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2016, pp. 1–5, doi: 10.1109/pesgm.2016.7741097.

[29] F. H. Al-Qahtani and S. F. Crone, "Multivariate k-nearest neighbour regression for time series data—A novel algorithm for forecasting UK electricity demand," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–8, doi: 10.1109/ijcnn.2013.6706742.

[30] M. Ghofrani, M. Ghayekhloo, A. Arabali, and A. Ghayekhloo, "A hybrid short-term load forecasting with a new input selection framework," *Energy*, vol. 81, pp. 777–786, Mar. 2015, doi: 10.1016/j.energy.2015.01.028.

[31] M. Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014, doi: 10.1109/tsg.2013.2277171.

[32] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, and M. S. Fadali, "Smart meter based short-term load forecasting for residential customers," in *Proc. North Amer. Power Symp.*, Aug. 2011, pp. 1–5, doi: 10.1109/naps.2011.6025124.

[33] H. Ziekow, C. Goebel, J. Struker, and H.-A. Jacobsen, "The potential of smart home sensors in forecasting household electricity demand," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Oct. 2013, pp. 229–234, doi: 10.1109/smartgridcomm.2013.6687962.

[34] R. P. Singh, P. X. Gao, and D. J. Lizotte, "On hourly home peak load prediction," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun.*, Nov. 2012, pp. 163–168, doi: 10.1109/smartgridcomm.2012.6485977.

[35] Y. Ding, J. Borges, M. A. Neumann, and M. Beigl, "Sequential pattern mining — A study to understand daily activity patterns for load forecasting enhancement," in *Proc. IEEE 1st Int. Smart Cities Conf.*, Oct. 2015, pp. 1–6, doi: 10.1109/isc2.2015.7366169.

[36] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H.-A. Jacobsen, "Household electricity demand forecasting: Benchmarking state-of-the-art methods," in *Proc. 5th Int. Conf. Future Energy Syst.*, 2014, pp. 233–234.

[37] A. Naz, N. Javaid, M. B. Rasheed, A. Haseeb, M. Alhussein, and K. Aurangzeb, "Game theoretical energy management with storage capacity optimization and photo-voltaic cell generated power forecasting in micro grid," *Sustainability*, vol. 11, no. 10, p. 2763, May 2019, doi: 10.3390/su11102763.

[38] K. Aurangzeb, "Short term power load forecasting using machine learning models for energy management in a smart community," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Apr. 2019, pp. 1–6, doi: 10.1109/ICCISci.2019.8716475.

[39] R. Khalid, N. Javaid, F. A. Al-zahrani, K. Aurangzeb, E.-U.-H. Qazi, and T. Ashfaq, "Electricity load and price forecasting using jaya-long short term memory (JLSTM) in smart grids," *Entropy*, vol. 22, no. 1, p. 10, Dec. 2019, doi: 10.3390/e22010010.

[40] M. Alhussein, S. I. Haider, and K. Aurangzeb, "Microgrid-level energy management approach based on short-term forecasting of wind speed and solar irradiance," *Energies*, vol. 12, no. 8, p. 1487, Apr. 2019, doi: 10.3390/en12081487.

[41] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1591–1598, Jul. 2017, doi: 10.1109/tsg.2015.2493205.

[42] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, Aug. 1996, pp. 226–231.

[43] I. Ullah, M. Hussain, E.-U.-H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Syst. Appl.*, vol. 107, pp. 61–71, Oct. 2018, doi: 10.1016/j.eswa.2018.04.021.

[44] D. Hsu, "Time series forecasting based on augmented long short-term memory," 2017, *arXiv:1707.00666*. [Online]. Available: https://arxiv.org/abs/1707.00666

[45] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," *Appl. Energy*, vol. 87, no. 11, pp. 3538–3545, Nov. 2010.

[46] J. J. López, J. A. Aguado, F. Martín, F. Muñoz, A. Rodríguez, and J. E. Ruiz, "Hopfield–K-Means clustering algorithm: A proposal for the segmentation of electricity customers," *Electr. Power Syst. Res.*, vol. 81, no. 2, pp. 716–724, Feb. 2011.

[47] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 420–430, Jan. 2014.