IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 22, 2020, accepted December 25, 2020, date of publication January 20, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3053069

A Pyramid-CNN Based Deep Learning Model for Power Load Forecasting of Similar-Profile Energy Customers Based on Clustering

KHURSHEED AURANGZEB[1], (Senior Member, IEEE), MUSAED ALHUSSEIN[2], KUMAIL JAVAID[3], AND SYED IRTAZA HAIDER[1], (Student Member, IEEE)

[1]College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[3]Department of Computer Science, University of South Asia, Lahore 54000, Pakistan

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

This work was supported by the Deputyship for Research and Innovation, "Ministry of Education" in Saudi Arabia, through the project IFKSURG-1438-034.

ABSTRACT With rapid advancements in renewable energy sources, billing mechanism (AMI), and latest communication technologies, the traditional control networks are evolving towards wise, versatile and collaborative Smart Grids (SG). The short term power load forecasting of individual as well as group of similar energy customers is critical for effective operation and management of SG. Forecasting power load of individual as well as group of similar energy customers is challenging compared to aggregate load forecasting of a residential community. The main reason is the high volatility and uncertainty involved for the case of individual and group of similar energy customers. Several machine/deep learning models have been developed in the recent past for forecasting load of individual energy customers, but such explorations are ineffective due to the requirement of one trained model for every energy customer, which is practically not feasible. We plan to build a deep learning model using convolutional neural network (CNN) layers in pyramidal architecture for effective load forecasting for a group of similar energy-profile customers. Initially, we grouped a subset of energy customers from database of Smart Grid Smart City (SGSC) into clusters using DBSCAN approach. The CNN layers are used for extracting feature from historical load of each cluster. The extracted feature of similar energy-profile customers (grouped based on clustering) is combined to make training-databases for each cluster. We have used the power load data from SGSC project, which contain thousands of individual household energy customers data. The developed Pyramid-CNN model is trained based on these sets of databases. The trained model is evaluated on randomly selected customers from few clusters. We obtained significantly improved forecasting results for randomly selected user from different clusters. Our adapted strategy of clustering based model training resulted in upto 10 percent MAPE improvement for the energy customers. The essence of our work is that energy customers can be grouped into clusters and then representative model could be developed/trained, which can accurately forecast power load for individual energy-customer. This approach is highly feasible, as we do not need to train a model per energy customer and still achieve competitive forecasting results.

INDEX TERMS Smart homes, smart grids, power load forecasting, CNN, low energy consumers, high energy consumers, clusters.

I. INTRODUCTION

The forecasting of short-term electricity loads is an important part of the energy market. The maintenance, processes, and administration of the power system can be greatly helped by accurate load forecasting. Energy can not be stored in vast amounts, which necessitates that generation and demand must have a fair balance. [1].

For correct and effective scheduling of power activities, it is necessary to accurately predict the power generated and

The associate editor coordinating the review of this manuscript and approving it for publication was Shiping Wen.

the power load. The strategies for efficient use of available power are essential to design, planning and operation of the grid. In addition, predictive errors have a notable impact on safety checks, dynamic state estimates, and power grid load transmission [2], [3]. To help run and plan the system, power transmission company develop and apply accurate forecasting models for both production and demand of the power.

The advancement in information and communication technology (ICT), modern measurement infrastructure (MMI) helped in the transition from conventional grid (also known as typical power grid or TPG) to the smart grid (SG). The advantages of this transition is in the form of seamless connection between energy customers and utilities, which allow for two-way communication. Power production, power division, and power usage can be monitored and improved by integrating ICT into power grid. Thanks to intelligent technologies and ICT, SG empowers its clients with reliable, sustainable, economical, safe, and efficient energy. The demand side management (DSM) technologies used in SG allows efficient load usage by switching the loads to off-peak slots (hours) from on-peak slots (hours), helping to reduce costs and monitor the capacity of the power grid. The utilization of energy can be increased by creating an opportunity for a two-way contact stream between utilities and consumers, which helps to enhance power system management process and optimization [4]–[8].

Accurate prediction of power load of energy customers enable them to gauge their energy usage and, wherever possible (maintaining user comfort), shift their appliance usage across day/night. Accurate load forecasting based on deep learning models provides an opportunity for energy users to relate their current energy costs to their future usage patterns. As a result, these customers can take benefit from predictive algorithms by being aware of their energy usage and future estimates, and they can manage their energy consumption costs more efficiently.

Energy consumers play a key role in responding to the demand and nicking of the SGs. They can be categorized into; residential, business and industrial classes. A big portion of energy from the total produced is consumed by residential sector. In the residential sector, the MMI installed is very useful in projection of the short-term electricity load of energy-users [9].

Statistical and machine learning (ML) models were developed in the past to predict the output of renewable energy resources (RERs) as well as the prediction of individual and aggregated load. The performance of these models are heavily dependent on the load data and can be classified as data driven models [10], [11].

The authors in [12], [13] and [14] applied Memristive neural networks for various problems including periodic sampling, global exponential synchronization and cyber physical.

The authors in [15] and [16] evaluated the performance of hybrid artificial neural network model and various artificial intelligence model for time series forecasting problems, which are similar in nature to our adapted strategy. They compared the performance of several artificial intelligence models including ELM, ANFIS, ANN, SVM and GPR for daily stream-flow time series forecasting.

The authors of the work in [17], argued that the detailed technical procedures are supportive to replicate them. Keeping this in mind, we have added more details to both technical sections and the description of the used datasets.

In the literature for dealing with short-term electric load predictions, several methods have been published. However, very few of them have addressed families and individuals. For short-term individual power load prediction, an LSTM based model was developed and assessed by the authors in [18]. But the main problem with their approach is the development and training of model for each and every individual of the residential community, which practically is not feasible.

Therefore, the aim in current work is to group the energy customers in a residential community based on their load pattern using DBSCAN clustering. In this way, energy users with similar energy consumption patterns will be grouped together. The historical load data of individual customers can then be combined into separate databases for each cluster. Then, convolutional layers can be employed in the deep learning model for extracting features from the database of the target cluster. One model will be trained based on the data of each cluster, which can be used for forecasting the energy consumption for every individual member of the cluster with high accuracy. In this way, few representative models would be developed and trained for each cluster of the residential community, compared to the case of having a trained model for each and every individual user.

We have used real world power load data of individual customers collected during Australian Government project known as Smart Grid Smart City (SGSC) [19]. It must be noted that the convolution layers of CNN learn the internal representation of the given time series data and obtain the significant features. The performance of the developed pyramidal-CNN deep learning model remain unchanged even if the parameter combinations changes. So, the performance of the proposed deep learning model is independent of the parameter combination. Our achieved results indicate competitive forecasting values for the individual household load of various clusters. Instead of having trained model for every individual user, we adopted a clustering based strategy, where one model for a group of energy customer (cluster) is trained, which achieved competitive forecasting results for various members of clusters.

This article 's main contributions are:

1) Developed a deep learning model (Pyramid-CNN), which can benefit from convolutional layers for feature extraction and pyramidal architecture for complexity reduction.
2) Applied DBSCAN algorithm for determining major clusters, minor clusters and outliers in power consumption behavior of 69 customers over a period of 92 days.

3) Applied K-means clustering for grouping energy customers into categories of 'lec' and 'hec' coressponding to low and high energy customers respectively.
4) Developed and trained separate models for categories of 'lec' and 'hec' based on the real-world historical power load data.
5) The trained models are used for forecasting power load of few individual customers from each category. It helped in proving usefulness of the group based training strategy for effective power load forecasting of similar-profile energy customers.

The structure of the remaining paper is as follows. Next section, i.e., II describes the related work. Details of the proposed model are unfolded in Section III. Section IV uncovers the data analysis and clustering based on DBSCAN approach. The results and discussion are provided in Section V. Finally, Section VI conclude the research work.

II. LITERATURE REVIEW

The short term electricity forecasting is a hot research topic which attracted significant attentions from researchers. Previously, conventional data analysis methods have been applied load forecasting based on time-series historical data. Recently, researchers have developed several different deep neural network (DNN) models for load prediction problems. The attention of the researchers towards this area increased duw to various reasons including tremendous developments in the fields of artificial intelligence (AI) and machine/deep learning and availability large amount of data from smart meter.

Due to their non-linear mapping properties, the artificial neural networks (ANNs) have been employed for short term load forecasting [20]. But, these forecasting models could easily be trapped in local minima, which is their main limitation. In addition, the convergence rate of these models is slow [21].

The generic neural network regression (GRNN) [22], [23], support vector machines (SVM) [24], extreme learning machines neural network [24] and kernel-based quantile regression [25] can also be used for forecasting purposes. Poor generalization is caused by irregularly chosen activation function [26]. Additionally, forecasting problems that involve in-depth extraction of features is not suitable because the sequence of layers cannot be encoded. The GRNN model is unsuitable for predictive matters due to its computation complexity [27]. The features of all of these models make them an unsuitable choice for prediction problems. Their main limitation are wide specifications for memory, high computational complexity and a variety of kernels [28].

So far, most of the focus of researchers has been on aggregate power load forecasting [29]–[36]. These model achieved excellent performance for system-level power load forecasting. But for the effective planning and management of the SGs, individual household level power load forecasting is needed. Predicting the short-term power usage of individual energy users of smart grid is gaining growing attention, throughout the modern history. In [37], the authors have taken into account the expediency of a time series to check individual predictions of the load of households. Their estimation is based on the root mean square error, (RMSE).

To estimate total housing loads over multiple time horizons and sampling intervals, about the writers in [38] applied Kalman filter based approach. They claim that a balance between accuracy and computational complexity is given by the sampling rate of selected samples.

In [39], the authors adopted SVM and ANN models for data with high resolution obtained from three homes over a period of 30 days. These authors investigated the possible impact of automatic meter reading (AMR) for short-term power load forecasting of household level energy consumer. They modeled the real-time measured data from the energy user's smart meters as the sum of the Gaussian noise signal and the deterministic component. Their obtained results indicate that the availability of the vast amount of collected data significantly increases the accuracy of the power load forecast. They argued that better prediction accuracy is accomplished at the expense of high computational difficulty. The authors explored different methods for forecasting the peak electricity demand of individual households. They concluded that, at the household level, the occupancy and historic peak electricity load are significantly better features for peak load prediction than season and temperature.

The authors in [40], devised a technique based activity sequencing and used the support vector regression for forecasting. They concluded that the variable of the operation sequence is an important component which could increase the prediction accuracy of power load of individual household. The authors explored several predictive models in [41] including neural networks (NN), ARIMA, etc for time horizon of 15-minute to 24-hour. Using two data sets, they evaluated the developed model. One of these data sets belonged to Six families in the USA, while the second belonged to a single household in Germany. For data sets from United States, they achieved average mean absolute percentage error (MAPE) of 85 percent, while for data sets from Germany 30 percent MAPE is achieved.

Various models, including SVM, classification and regression trees, and neural networks with multilayer perceptrons (MPNN) were used by the researchers in [42]. They concluded that the accuracy of the forecast could be significantly improved by a blend of household activity and historical electricity usage data from individual households. For neural networks and SVM, they obtained MAPE of 51 percent and 48 percent, respectively. Several other researchers proposed and evaluated several different effective models for prediction of electric loads [43]- [47]. Specifically, in [47], we proposed a Hybrid deep learning model which, is composed of convolutional layers and LSTM layers, where the focus has been on power load forecasting of individual energy customer.

Among all these model exploration for power load forecasting of individual household, one of the innovative work was published by the authors of [18]. They implemented an

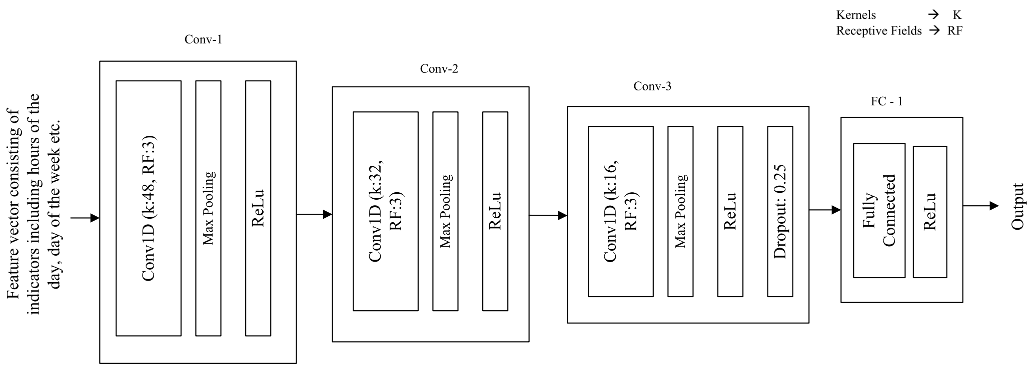

FIGURE 1. The proposed Pyramid-CNN based deep learning model.

LSTM-based deep learning model to forecast a short-term residential load of individual energy customer. They compared their model to the newest model of machine learning, as well as with the experimental model. In order to forecast the short-term residential load of 69 consumers, their developed model achieved an average MAPE of 44.06 percent.

There is a diverse research trend by valuable researchers that looks at behavioral and other factors that influence the energy usage of a single household. The goal is to analyze and obtain feedback on each individual household energy user's electricity consumption patterns and to determine the important fundamental relationship between contexts such as time of use, day of the week (day of the week or weekends), season, etc. The insights can increase understanding and awareness of domestic energy usage through such studies, that will lead us to effective use of energy.

The main problem with the idea and exploration in [18] is the development/training of models for each and every individual household in a residential community, which is practically not feasible. A more practical approach would be to group the energy customers of a residential community into several clusters based on their energy consumption patterns/profiles. The main challenge is to achieve competitive load forecast for the individual user of each clusters, by adopting a strategy of training only one model for each cluster. This approach will be more practical, feasible as well as reliable if the developed model achieve better or competitive forecasting accuracy based on the model trained on the data of the specific cluster.

III. PROPOSED MODEL

Our proposed Pyramid-CNN model is unfolded in this section. The feature vector which is feed into the model contains several information about the power consumption of the electricity users. It include the energy consumption from previous instances of the time. Additionally, it also contain information about the hour of the day and the day of the week information, which is provided in the form of one-hot encoding. The output from the trained model represents the predicted energy consumption for the specific time instance in future. The developed model is significantly enhanced from the previous conference version both in terms of kernel of CNN for feature extraction and the receptive fields. Additionally, some dropout layers are added for better training of the developed model. The specific architecture of the developed Pyramid-CNN models is presented in Fig. 1. Three feature extraction blocks in the form of 1D convolutional layers with decreasing kernels size can be seen in Fig. 1. In the revised model, the MaxPooling layer is added before the Rectified Linear Unit (ReLU) layer, which decreases the spatial dimension (feature maps) with a factor of 2. This aids in reducing the computational complexity of the model. On the other hand, for reducing the impact of gradient vanishing problem, the ReLU layer are added in the developed model. The dropout layer is added for preventing the overfitting issue. The output of the dropout layer is connected to the fully connected layer for producing output.

Adopting the coarse-to-fine framework is a common approach developing CNN based deep learning models. The coarse-to-fine framework exhibits significantly high computational cost due to the involvement of large proportions of trainable parameters. For our implementation, we have selected the well-known pyramid architecture from the work in [48]. In this pyramid architecture, the number of kernels are large in the initial stage and are reduced with a constant factor as we go deeper in the network. For the developed Pyramid-CNN model, the parameter settings is unfolded in Table 1. We selected the 'Adam' optimizer and mean absolute error.

The training flow for our developed Pyramid-CNN model is shown in Fig. 2. The prepared input data in the form of feature vector is divided into three sets: training; 70%, validation; 20%, test ;10%. At the beginning, the validation and the training data are loaded for initializing the training process. The validation loss is calculated/evaluated to see whether it is decreasing or not. Our adopted strategy is to check validation loss for continuous decrements. If it is decremented in every next step, then the trained model till that time is stored with its weights. But, if it is not decremented and stays constant for 15 consecutive epochs, then we reduced the learning rate. The model is trained for 200 epochs. The last stored model

TABLE 1. Configuration of the developed model.

The Parameters	Setting
Optimization Function	Adam optimizer
Loss Function	MAE
Learning Rate	0.001
Changing learning Rate if stuck	The validation loss is monitored for 15 Epochs and changed with a factor of 0.80, where minimum allowable is 1e-5
Batch Size	64
Epoch	200

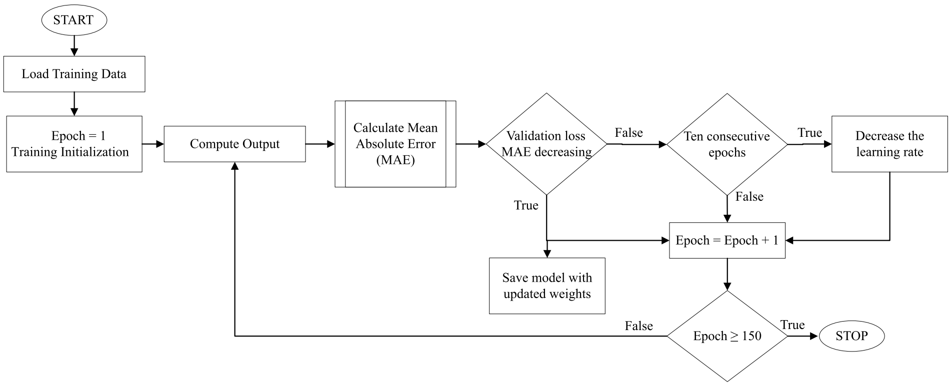

FIGURE 2. The flow chart for the training of the developed deep learning model.

is our best model and is loaded for forecasting the power consumption of the electricity users.

A. THE PYRAMID-CNN MODEL

The different layers employed in the developed deep learning model such as MaxPooling, 1D convolution, ReLU, dropout, and FC are elaborated briefly below.

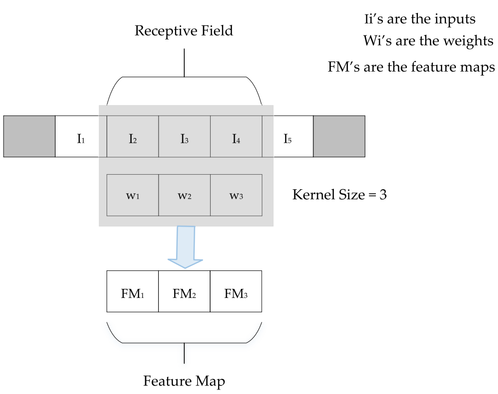

FIGURE 3. The 1D convolution process.

1) 1D CONVOLUTION LAYER

This is the most important layer in our feature extraction block, for which the operational model is shown in Fig. 3. It consists of learnable filters, which are used to perform the convolution with the input signal. In the learnable filters of CNN layer, all the receptive field are similar. The kernel weights are w_1, w_2, w_3 which are convolved with inputs i_1, i_2, i_3, i_4, i_5 in a sliding fashion for attaining the feature maps FM_1, FM_2, FM_3. It is evident from Fig. 3 that the feature map FM2 is obtained in the following way using Equation 1.

$$\mathrm{FM}_2 = \mathrm{w}_1{}^*\mathrm{I}_2 + \mathrm{w}_2{}^*\mathrm{I}_3 + \mathrm{w}_3{}^*\mathrm{I}_4 \tag{1}$$

The learnable filters in the CNN layers of our developed model are applied to the inputs for obtaining the feature maps. The convolutional layer is very effective and stacking suitable number of convolutional layers in the developed deep learning model enable to learn the features at the early stage in the input data. The produced feature maps keep tracking the exact location of different features of the input signal. It is highly important to note that minor changes in location of various features of the input data will produce a totally different feature map.

2) MAXPOOLING LAYER

Normally, the convolutional layer is followed by a pooling layer in order to mitigate the invariance of feature map. It affects all the produced feature map differently for creating new pooled feature maps. Normally, the pooling filter size is significantly lower than size of the produced feature map. Incorporating the pooling layer in a deep model produces feature map (pooled) that shows summary of the features in the input signal. Pooling have different types in which we have selected the MaxPooling that is a down sampling scheme. According to the scheme of MaxPooling, a filter slide across the input signal and the maximum value in the overlapped region is chosen as the output. We added a MaxPooling layer after the convolution layer in the proposed Pyramid-CNN model. The operational example of 1D MaxPooling is shown in Fig. 4

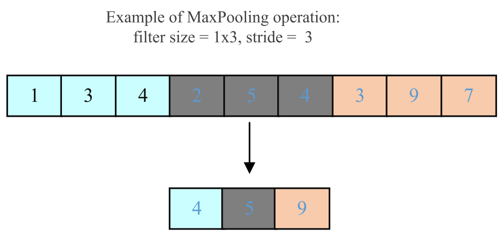

FIGURE 4. The 1D MaxPooling with kernel size 1 × 3 and stride of 3.

3) RELU

The Rectified Linear Unit (ReLU) is an activation function, which is added in the deep learning models for augmenting its capability to learn the complex structures. In our developed deep learning model, the ReLU is added after every MaxPooling layer and also the fully connected layer. The mathematical operation of ReLU is thresholding, which is given in Equation 2.

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

4) DROPOUT

The overfitting of a deep learning model is a major issue, which can be relieved with the help of adding a dropout layer. It randomly select some neurons and deactivate them during the training of the deep learning model. The result of deactivated neurons is zero.

5) FULLY CONNECTED LAYER

The fully connected (FC) layer is added in any deep learning model for the nonlinear mapping to output from input. Usually, the FC layer is added at end of deep learning model.

IV. DATA ANALYSIS

For the current research work, we selected smart meter data from the SGSC project [19] of the Australian Government. For power consumption analysis of individual households, the data gathered in SGSC project could be used as it have the load data for ten thousands of energy customers from Austalia.

Our focus in this research work is analysis of power consumption of a set of individual electricity customers from the whole database of 10,000 customer of SGSC project. It is not realistic approach to analyze all the customers, so for demonstration purpose, we have chosen the customers having installed the hot water system. This criterion is applied as it will enable us to compare our obtained results with best model previously published in IEEE Transaction on Smart Grid [18]. With the restriction of customers having hot water system, we divided the set of 10,000 customers and selected a pool of 69 customers for our analysis. The process for extraction of pool of 69 energy customers from the whole set of 10000 customers is shown graphically in Fig. 5.

For a residential customer in a community, the power load at a specific instance can be forecasted with reasonable accuracy [34]. Normally, the variety in the power consumption at community level stabilize the day-to-day load profiles which makes it relatively easier to forecast the substation power consumption. On the other hand, the power load of the individual household lacks the obvious patterns, which makes it difficult to achieve high forecasting accuracy. The power load of the individual residential customer varies with the variations in the weather condition. For individual energy customer different parameters including affordability, lifestyle, and

---

Algorithm 1 The Pseudocode of Our Proposed Approach

Input: The power load database from SGSC project
Procedure: Selecting customers having hot water system
while The whole database is not searched **do**
  for each customer **do**

- Check whether the hot water system is installed;
- Maintain a new database for the customers having hot water system;

The Pool of 69 Customers:The pool of 69 customers who have installed the hot water system;
New Database:For pool of 69 customers, extract the data for three months and impute the missing values;
Model Development:

- Design and Develop the pyramid-CNN deep learning model;

Customers-Grouping, Training and Testing:

- The customers having similar power consumption profile are grouped using KNN;
- The dataset for each group is separated;
- Train, validate and test separate model for each group
- Use the MAPE metric for evaluating the model

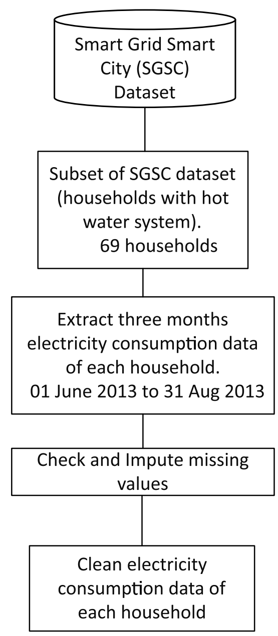

FIGURE 5. The data extracted for the 69 household from the whole database of 10000 customers of SRSG project.

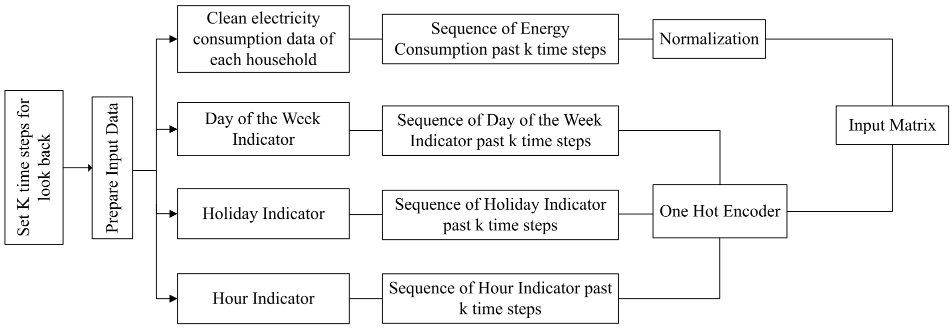

FIGURE 6. Data extracted for the 69 household from whole database of SCSG project.

work-home hours ratio affect the power consumption behavior. These different factors make it difficult to accurately forecast the power consumption behavior of the individual household. The different features which play role in the power consumption behavior of individual household and how we combined them to a make the feature vector for our model are shown in Fig. 6

We have selected the well known clustering method (density-based) called DBSCAN [34]. The plus point of applying the DBSCAN for clustering analysis of power demand of individual customers is that the cluster information is not needed in advance. Furthermore, DBSCAN also determines the outliers of the analyzed dataset. Normally, the power demand of end user is similar in the weekdays and in the weekend days. Due to this characteristics of the power demand of the individual customers, the DBSCAN is the most suitable approach for clustering in order to determine the outliers in the data set. The lower number of outliers in the results of DBSCAN will indicate that the consistency of individual customer is high.

TABLE 2. The proportion of outliers, minor and major clusters in the selected subset of 69 energy users.

S. No	Customers	Major	Minor	Outliers
1	8459427	1	0	0
2	8198319	1	0	1
4	8196671	1	0	6
3	8342852	2	2	1
5	8196669	1	1	12
6	8487285	1	4	35
7	8568209	2	2	37
8	8540084	1	1	69
9	9012348	0	4	83
10	8282282	0	0	92

Table 2 presents the number of minor and major clusters as well as outliers for ten randomly chosen electricity users from the whole set of 69 user over a period of 92 days. It can be easily observed in this table that there is significant variations in the number clusters (minor and major) and the number of outliers. It is evident from this table that customer with ID 8342852 has two major clusters, two minor clusters and one outlier. It is also evident that customer with ID 8198319 has one major cluster, no minor clusters and one outlier, which indicate that the power consumption profile of this user can be predicted easily. On the other hand, customer with IDs 8282282 and 9012348 have no major cluster. Furthermore, customer 9012348 has four minor clusters along with 83 outliers. Also, customer 8282282 has no major and minor cluster and has 92 outliers. This indicate that power consumption behavior of the customer with ID 8282282 and 9012348 is significantly volatile and hence is difficult to forecast precisely.

The power consumption (half-hourly) of one randomly chosen energy customer from the dataset (customer ID 8523058) over a period of three months is presented in Fig. 7. It can be observed in this figure that power consumption profile of this user has minor variation due to which it makes a single major cluster for all the 92 days without having any outlier.

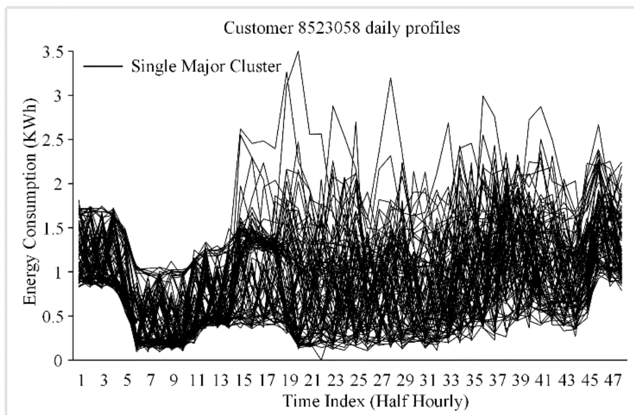

FIGURE 7. Case (Customer 8523058) with no outliers and only one major cluster.

The histogram representing number of outliers in the whole pool of selected 69 customers is shown in Fig. 8. It can be observed in this figure that 29 customers (first bar) have outliers in the range of 0 to 20. Similarly, 22 customers have outliers in the range of 20 to 40, which means that 22 out of 69 customers have outliers in range of 20 to 40 (huge range). This is a critical results and it (higher number of outliers) implies that a single trained model will not be able

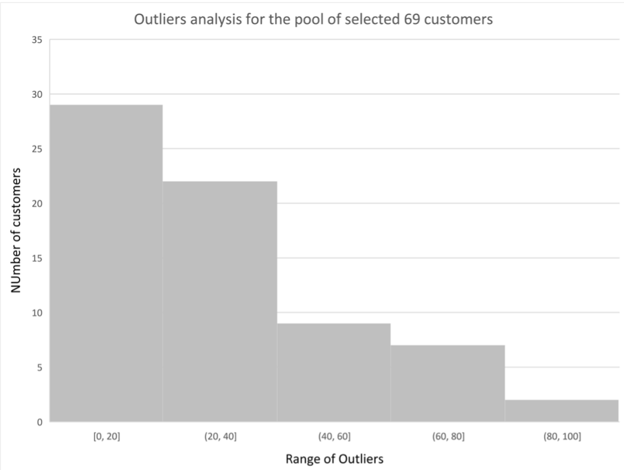

FIGURE 8. The histogram representing the number of outliers for the pool of 69 customers.

to accurately predict the power consumption for the whole community. At the other extreme developing 69 different models (the strategy adapted in [18]) and their training and testing for each individual customer is also impractical and not feasible.

The outlier's analysis in the consumption behavior of the electricity users is shown in Fig. 9.

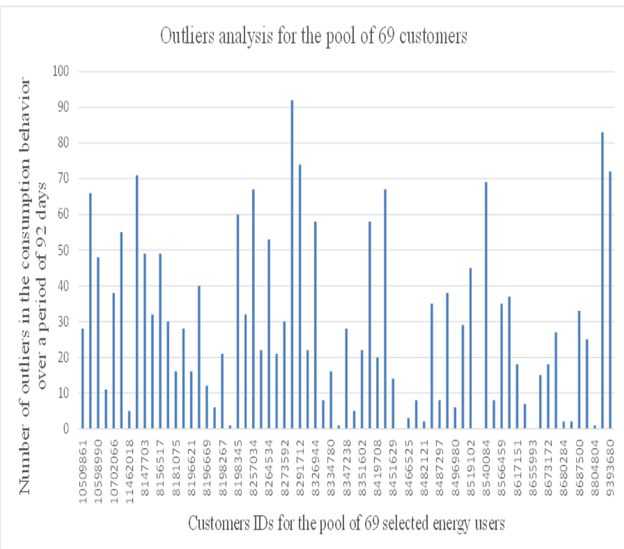

FIGURE 9. The outlier's analysis in the consumption behavior of 69 customers over a period of 92 days.

So, a good solution is the approach based on clustering: To develop and train one model for each cluster and use it for forecasting power load of various customers within the selected cluster only. In this way, several representative models for clusters in a residential community will be needed, which is feasible. The promising solution for this problem is the clustering. By grouping the customers into various clusters based on some criterion, few representative models could be developed and training, which will be able to predict reasonably better compared to the single model solution. Furthermore, such an approach is feasible and practical.

V. RESULTS AND DISCUSSIONS

We have used the power load data from SGSC project, which contain thousands of individual household energy customers data. At every time instance, the data consists of 57 different features representing useful information such as time of the day, day of the week, week day or week end day etc. For one time instance, the data contains 57 different values, which is recorded every half hour. In total, for one day we have 24*2 = 48 samples, where each sample have 57 different values representing valuable information. In total, the training data is huge and require an efficient deep learning model for effective and accurate forecasting.

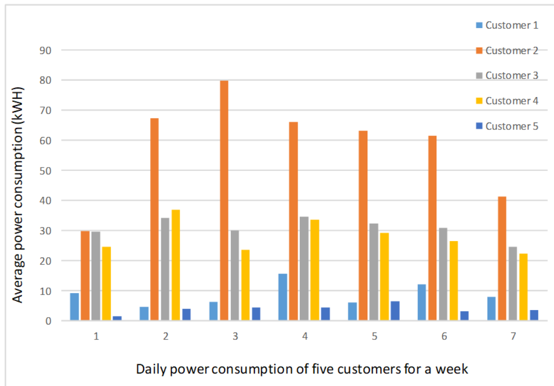

FIGURE 10. Average daily power consumption of five customers for a week.

The daily accumulated power consumption of five individual customer (randomly selected from the pool of 69 customers) for a week is presented in Fig. 10. It is evident from this figure that accumulated power consumption of the individual customers even in a period of a week is highly dissimilar. It is also evident that the power spans over a range from approximately 3 kwh to 79 kwh. Significantly dissimilar power consumption of individual energy customer consistently over a period of a week advocate that developing and training one deep learning model for 69 different customers cannot precisely predict the power consumption even in a single residential community. The residential customers could be grouped into various clusters based on various clustering strategies and then representative deep learning models could be developed and trained for relatively better forecasting.

We can observe in Fig. 10 that the day power consumption of customer 5 remains lower than 9kWH with an average of 3 kWH. There are many other customers with similar power consumption behavior in a residential community. These energy consumers are low energy consumers and are represented with "lec". It should also be observed in the same figure that the power consumption of customer 2 is significantly higher and reaches up to 80kWH in some days.

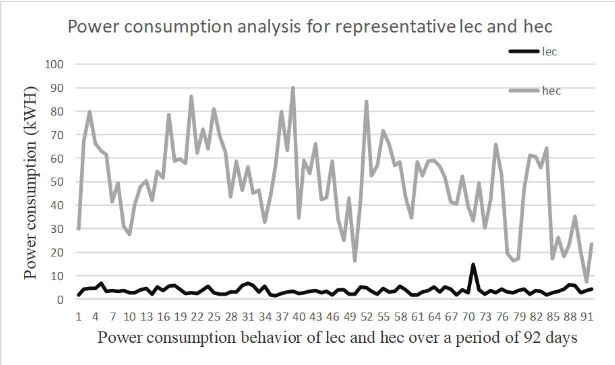

FIGURE 11. **The power consumption analysis of representative lec and hec over a period of 92 days.**

Many other customers in the residential community also have power consumption in similar range and we term them higher energy consumers represented by "hec". For demonstration purpose, the power consumption behavior of lec and hec over a period of 92 days shown in Fig. 11.

Using the same database, the performance of different notable machine learning methods such as ELM [33], KNN [49], back propagation (BP) and Artificial Neural Network (ANN) is compared with the performance of our proposed deep pyramid-CNN based on the evaluation metrics of MAPE and the results are provided in Table 3.

TABLE 3. **Comparison of proposed model with different machine learning models.**

Machine learning model	Achieved MAPE (%)
ELM	122
KNN	71
BPNN	49
ANN	47
LSTM	44
Pyramid-CNN (Proposed)	39

The MAPE values in Table 3 show that the general machine learning models including BP, ANN, ELM, KNN and LSTM could be used for simple time series forecasting with very few attributes, but when the time series involves multiple features such as in our case, then high forecasting accuracy cannot be ensured with simple models, necessitating the use of deep learning approach.

The CNNs are perhaps the most commonly used efficient deep learning approaches. Our main motivation for applying CNN model on the time series data is that it is highly useful for extracting the useful hidden features from the input data as well as it will help in removing the noise from the data. So, the main aim of using CNN layers in our proposed pyramid-CNN deep learning model is to extract patterns of local tendency and same local-pattern found in other region of the given time series data. Hence, we proposed to develop pyramidal-architecture based deep learning model which, uses convolution layers of CNN for learning the internal representation of the given time series data and obtain the significant features. The pyramidal architecture is selected for reducing the computational complexity of the overall deep learning model.

As mentioned earlier, we have used the publicly available power load data from Australian project named Smart Grid Smart City (SGSC). In this database, at every time instance, the data consists of 57 different features representing useful information such as time of the day, day of the week, week day or week end day etc. For one time instance, the data contains 57 different values, which is recorded every half hour. In total, for one day we have $24*2 = 48$ samples, where each sample have 57 different values representing valuable information. In total, the training data is huge and require an efficient deep learning model for effective and accurate forecasting.

We applied the k-means clustering for grouping the energy users having identical power consumption patterns. The pool of 69 customers are divided into 15 groups of varying number of energy users in each group which is shown in Fig. 12.

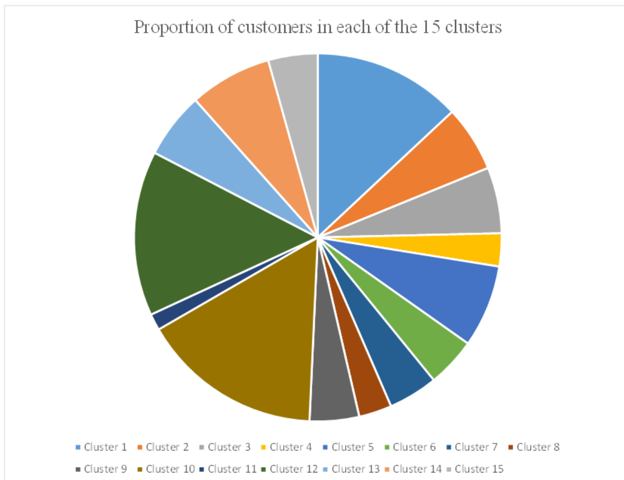

FIGURE 12. **The grouping of 69 customers in to 15 groups using k-means clustering.**

The developed Pyramid-CNN model is trained based on the historical consumption data of users of Cluster 2. This trained model is used for predicting the power consumption of the four energy users of Cluster 2 and the result is shown in Fig. 13. It is evident from Fig. 13 that even in the same cluster, the power consumption behavior is somewhat different. But still, the developed and trained model quite accurately predicted the power consumption for the four different customer of one cluster, i.e., Cluster 2.

Another model is trained based on the historical power consumption data of the three energy customers of Cluster 14. This newly trained model is used for predicting the power consumption of the three energy users of Cluster 14 and the prediction analysis is shown in Fig. 14. It can be observed in this figure that in this cluster, the power consumption behavior is quite similar due to which the newly trained model efficiently predicted the power consumption of the

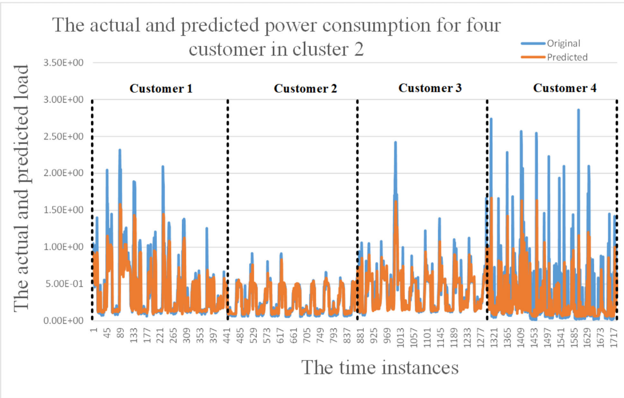

FIGURE 13. The prediction of power consumption behavior of four energy users of Cluster 2.

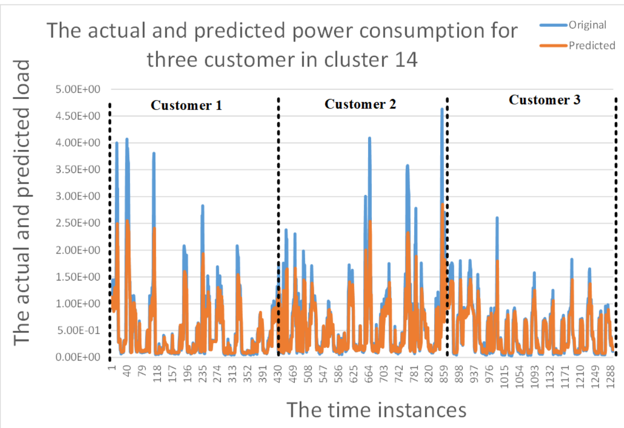

FIGURE 14. The prediction of power consumption behavior of three energy users of Cluster 14.

three different customer of this cluster. Based on the accurate prediction results of the two differently trained models proved both the efficiency of the developed deep learning model as well as the efficiency of our adapted strategy of "grouping similar energy customers into clusters".

TABLE 4. Comparison of MAPE for four customers of Cluster 2.

Customer ID	MAPE achieved in [18]	Our achieved MAPE
Customer 1	35	34
Customer 2	24	22
Customer 3	47	37
Customer 4	135	129

Customer IDs in the database of SGSC project for four members of Cluster 2, i.e., Customer 1, 2, 3 and 4 are 11462018, 8196669, 10692972 and 9012348 respectively. Similarly, customer IDs in the database of SGSC project for three members of Cluster 6, i.e., Customer 1, 2 and 3 are 8149711, 8568209 and 8145135 respectively. The comparison of forecasting result for customers of Cluster 2 and Cluster 6 are shown in Table 4 and Table 5 respectively. It can

TABLE 5. Comparison of MAPE for three customers of Cluster 6.

Customer ID	MAPE achieved in [18]	Our achieved MAPE
Customer 1	56	52
Customer 2	57	55
Customer 3	49	48

be observed in these two tables that the trained model for each cluster produced significantly better MAPE score.

So, based on the obtained results and observation in Fig. 13, Fig. 14, Table 4 and Table 5, we can say that our adapted strategy of clustering based model training is highly useful both in terms of number of models to be trained and obtaining significantly better forecasting accuracy. The number of models to be trained are reduced because we proposed to train a single model for each cluster compared to per customer model training approach adopted previously [18]. The forecasting results are improved due to improved performance of our developed model, which we adopted from the work in [48] (they used it for healthcare application). Based on the observation for various energy customers of Cluster 2 and Cluster 6 in Table 4 and Table 5, we can say that our adapted strategy of clustering based model training resulted in upto 10 percent MAPE improvement.

VI. CONCLUSION

The energy management and planning of Smart Grids is a challenging issue due to unpredictable energy consumption behavior of energy customers. Accurate load forecasting of individual and group of similar-profile energy user is highly important for successful operation of SG. Power load forecasting of individual energy customers has been studied by some researchers but it is unfeasible practically due to requirement of trained-model per energy user. We have developed Pyramid-CNN based deep learning model for forecasting load of a group of similar-profile energy user. Initially, DBSCAN based clustering is performed for grouping the energy customers into low energy customers (LEC) and high energy customers (HEC). Then, the CNN layers are used for extracting feature from historical load date of the cluster of energy customer. The extracted features are used for training the developed Pyramid-CNN model for the two clusters of customers. We obtained comparative forecasting results even for randomly selected customers from the group of both LECs and HECs. This proved the efficiency of our proposed approach of grouping similar-profile energy customers and then developing/training a representative model for each cluster. It eliminate the need for developing and training models for individual household, which is practically not feasible. Compared to the result of the basic LSTM model, we have obtained improved MAPE of 34, 22, 37, 129 from 35, 24, 47, 135 for low energy customers and of 52, 55, 48 from 56, 57, 49 for high energy customers. Based on the observation for various energy customers of two clusters representing low and high energy customers, we achieved an average MAPE improvement of 10 percent.

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, Apr. 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. ICIC* (Lecture Notes in Computer Science), vol. 5754, D. S. Huang, K. H. Jo, H. H. Lee, H. J. Kang, and V. Bevilacqua, Eds. Berlin, Germany: Springer, 2009, doi: 10.1007/978-3-642-04070-2_95.

[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] Y. Cao, Y. Cao, Z. Guo, T. Huang, and S. Wen, "Global exponential synchronization of delayed memristive neural networks with reaction–diffusion terms," *Neural Netw.*, vol. 123, pp. 70–81, Mar. 2020, doi: 10.1016/j.neunet.2019.11.008.

[13] Y. Wang, Y. Cao, Z. Guo, T. Huang, and S. Wen, "Event-based sliding-mode synchronization of delayed memristive neural networks via continuous/periodic sampling algorithm," *Appl. Math. Comput.*, vol. 383, Oct. 2020, Art. no. 125379, doi: 10.1016/j.amc.2020.125379.

[14] S. Wang, Y. Cao, T. Huang, Y. Chen, and S. Wen, "Event-triggered distributed control for synchronization of multiple memristive neural networks under cyber-physical attacks," *Inf. Sci.*, vol. 518, pp. 361–375, May 2020, doi: 10.1016/j.ins.2020.01.022.

[15] Z.-K. Feng and W.-J. Niu, "Hybrid artificial neural network and cooperation search algorithm for nonlinear river flow time series forecasting in humid and semi-humid regions," *Knowl.-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106580, doi: 10.1016/j.knosys.2020.106580.

[16] W.-J. Niu and Z.-K. Feng, "Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management," *Sustain. Cities Soc.*, vol. 64, Jan. 2021, Art. no. 102562, doi: 10.1016/j.scs.2020.102562.

[17] G. Ren, Y. Cao, S. Wen, T. Huang, and Z. Zeng, "A modified elman neural network with a new learning rate scheme," *Neurocomputing*, vol. 286, pp. 11–18, Apr. 2018, doi: 10.1016/j.neucom.2018.01.046.

[18] 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.

[19] (2014). *Smart Grid, Smart City, Australian Govern., Australia, Canberra, ACT, Australia*. [Online]. Available: http://www.industry.gov.au/ENERGY/PROGRAMMES/SMARTGRIDSMARTCITY/Pages/default

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

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

[28] 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. 185–189, 2012.

[29] 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.; Ubiquitous Comput. Commun.; Dependable, Autonomic Secure Comput.; Pervas. Intell. Comput.*, Oct. 2015, pp. 283–287.

[30] 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.

[31] 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)*, Oct. 2015, p. 6.

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

[33] R. Zhang, Z. Y. Dong, Y. Xu, 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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. (SmartGridComm)*, Oct. 2013, pp. 229–234, doi: 10.1109/smartgridcomm.2013.6687962.

[40] 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. (SmartGridComm)*, Nov. 2012, pp. 163–168, doi: 10.1109/smartgrid-comm.2012.6485977.

[41] 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. (ISC)*, Oct. 2015, pp. 1–6, doi: 10.1109/isc2.2015.7366169.

[42] 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.*, Jun. 2014, pp. 233–234.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting," *IEEE Access*, vol. 8, pp. 180544–180557, 2020, doi: 10.1109/ACCESS.2020.3028281.

[48] 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.

[49] 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.

KHURSHEED AURANGZEB (Senior Member, IEEE) received the B.S. degree in computer engineering from the COMSATS Institute of Information Technology Abbottabad, Pakistan, in 2006, the M.S. degree in electrical engineering (system on chip design) from Linkoping University, Sweden, in 2009, and the Ph.D. degree in electronics design from Mid Sweden University, Sweden, in June 2013. He is currently an Assistant Professor with the College of Computer and Information Sciences, King Saud University (KSU), Riyadh, Saudi Arabia. He has authored and coauthored more than 68 publications, including IEEE/ACM/Springer/Hindawi/MDPI journals, and flagship conference papers. He has obtained more than ten years of excellent experience as an instructor and researcher in data analytics, machine/deep learning, signal processing, electronics circuits/systems, and embedded systems. He has been involved in many research projects as a principal investigator and a co-principal investigator. His research interests are in the diverse fields of embedded systems, computer architecture, VLSI, signal processing, wireless sensor networks, camera-based sensor networks, and smart grids, with an emphasis on big data, precision agriculture, machine/deep learning, embedded and pervasive computing, mobile cloud computing, and healthcare.

MUSAED ALHUSSEIN received the B.S. degree in computer engineering from King Saud University (KSU), Riyadh, Saudi Arabia, in 1988, and the M.S. and Ph.D. degrees in computer science and engineering from the University of South Florida, Tampa, FL, USA, in 1992 and 1997, respectively. Since 1997, he has been on the Faculty of the Computer Engineering Department, College of Computer and Information Science, King Saud University. Recently, he has been successful in winning a research project in the area of AI for healthcare, which is funded by the Ministry of Education in Saudi Arabia. He is the Founder and Director of the Embedded Computing and Signal Processing Research (ECASP) Laboratory. He is currently an Associate Professor with the Department of Computer Engineering, College of Computer and Information Sciences, KSU. His research activity is focused on typical topics of computer architecture and signal processing with an emphasis on big data, machine/deep learning, VLSI testing and verification, embedded and pervasive computing, cyber-physical systems, mobile cloud computing, big data, eHealthcare, and body area networks.

KUMAIL JAVAID received the bachelor's degree from the University of Engineering and Technology, Lahore, and the master's degree in intelligent systems from the University of Essex, U.K. He is currently a Faculty Member of the Computer Science Department, University of South Asia.

SYED IRTAZA HAIDER (Student Member, IEEE) received the B.E. degree in electronics engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2010, and the M.S. degree in electronics engineering from King Saud University (KSU), Saudi Arabia, in 2015. He is currently a Researcher with the Embedded Computing and Signal Processing Lab (ECASP), King Saud University. His research interests include signal processing, mixed-signal design, and image processing.