

This paper is a postprint of a paper published in Springer Nature link-Electrical Engineering (2022) 104:347–361
<https://doi.org/10.1007/s00202-021-01395> and is subject to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Weather classification-based load and solar insolation forecasting for residential applications with LSTM neural networks

Dipanshu Naware¹ · Arghya Mitra¹

Abstract

Forecasting has always been the backbone of planning studies in the power system. With the advent of the restructured power system and also the integration of renewable energy sources, it has become more challenging. Accurate load forecasting leads to a reduction in supply–demand mismatch, whereas solar insolation/wind speed forecasting improves generation side planning. With the rapid growth in computer-based intelligence, self-learning mechanisms together with continuously growing volume of the dataset and affordable cost-effective data storage options, deep learning performance keeps getting better and has made a huge impact to solve the complex problems associated with the power system network. In this paper, we propose a classical long short-term memory neural network model for a combined day-ahead load and solar insolation forecasting of residential consumers equipped with rooftop PV based on weather classification. The effectiveness of the proposed model based on mean absolute error, mean bias error and coefficient of correlation (R -Squared) is compared with that of the standard artificial neural network and convolutional neural network models.

Keywords Deep neural network · Load forecasting · Insolation forecasting · Machine learning · Microgrid · Weather classification

1 Introduction

Forecasting is one of the most important aspects of all types of planning studies. Load forecasting (LF) plays an important role when it comes to power system planning. It involves short-, medium- and long-term forecasting depending upon the time duration which may vary from hours to days and days to years. Prediction of load in advance requires consideration of many parameters like historical load data, weather-dependent data (like temperature, humidity, wind intensity, etc.), day of the week, the hour of the day, holidays, etc., in order to reduce supply–demand mismatch.

On the other hands, solar power generation is increasingly important. Indeed, solar and wind resources, along with the energy storage systems (ESS), are actively adopted within the existing power systems to reduce carbon footprint. The

adoption of renewable resources in distributed generation is a suitable option that is being supported in many countries. For example, the Ministry of New and Renewable Energy (MNRE), Govt. of India targets to achieve deployment of at least 10,000 renewable energy (RE)-based micro- and mini-grid projects across the country with a minimum installed RE capacity of 500 MW in next 5 years [1]. An ingenious techno-economic infrastructure, keeping in mind the intermittent nature of these alternate sources of energy, is thus the need of the day. Hence, forecasting of solar insolation enhances the generation side planning.

Recently, several advanced artificial intelligence (AI) technologies, viz. machine learning (ML), deep learning (DL), reinforcement learning (RL), deep reinforcement learning (DRL), natural language processing (NLP), and expert systems, have been developed, as documented in [2–4]. Computational efficiency improves as the set of training examples increases giving a promising outcome. With the continuously growing volume of the dataset and affordable cost-effective data storage options, DL performance keeps getting better. Various DL architecture includes restricted Boltzmann machine (RBM), recurrent neural network (RNN), convolutional neural network (CNN), long

✉ Dipanshu Naware
dipanshu.0104@gmail.com

Arghya Mitra
mitrarghya@gmail.com

¹ Department of Electrical Engineering, Visvesvaraya National Institute of Technology, Nagpur, M.S 440010, India

short-term memory (LSTM), and gated recurrent unit (GRU), as described in [5] with a focus on the use of deep neural networks (DNN) in image classification and retrieval.

Several applications of ML are proposed for the planning and operation of power systems; as huge amount of data are collected in modern-day grid. Combined load and insolation forecasting issue remains open for research as many studies have documented short-term load forecasting (STLF) and insolation forecasting separately and not as a combined operation. Also many studies are confined to parameters like historical load data, historical insolation data, temperature, humidity and wind speed. Illustrative examples are the following. An hourly forecasting model is designed and implemented in [6, 7] as a multiple regression that considers historical load data, weather-dependent variables, and some past errors. In [8], STLF model is demonstrated in a microgrid, based on the combination of self-organizing map (SOM), k-means clustering, and multi-layered perceptron (MLP), where SOM classifies patterns based on historical data and clustered using the k-means algorithm and each cluster is trained using MLP for load prediction. Similarly, a DNN with LSTM layer is proposed in [9] for STLF, considering the effect of weekdays and weekends in a day ahead Korean electricity market. A bagged neural network is presented for STLF in [10] to achieve improved performance. An autoregressive integrated moving average model is used in [11], for forecasting a residential electricity consumption, whereas a comparative analysis of four forecasting methodologies is carried out in [12] for long-term load prediction. In [13], an evolutionary-based optimization technique, called Bat algorithm, is proposed for STLF utilizing the effect of weather variables.

It is observed that neural network approaches have been used extensively over other methods due to their ability to deal with non-linearities. As LF requires historical data along with various weather variables, observing patterns from these events prove to be helpful in predicting future loads. In [14], STLF is conducted using ANN and bootstrap aggregation model. Here, hourly solar irradiance is taken as an additional input to the forecasting models. In one of the studies [15], a convolutional LSTM is developed for enhanced STLF at residential level by integrating autoregressive and exogenous feature selection. ANN for STLF is reviewed to predict the next hour's load, next day's peak load or next day's total load in [16]. Again, in [17], electricity price is considered as one of the inputs to ANN that influences the relation between load and other variables for STLF. A feed-forward neural network (FFNN)-based STLF model is developed in [18] that considers the holiday forecasting model to reduce the forecasting errors. The result shows that the inclusion of this model improves the accuracy of both normal days and holidays.

In a traditional FFNN, output at the current time step is independent of output at the previous time step. As load demand, electricity prices and PV generation vary on an hourly basis, time series forecasting is needed to overcome this drawback. Hence, RNN serves the purpose where the output from the previous step is fed as an input to the next step. But RNN's are not good at capturing long-term dependencies; also both FFNN and RNN suffer from vanishing or exploding gradient. To overcome this problem, the LSTM architecture is introduced in [19]. It includes a memory cell and is the most important element in an LSTM architecture. LSTM has received massive recognition in the area of smart grid. A meter-level residential LF is proposed in [20] and is based on appliance energy measurement using LSTM networks and compares with traditional FFNN and K-nearest neighbor. Performance comparison shows that forecasting accuracy improves under the proposed work as appliance-level energy consumption measurement is made available. To probe the efficacy of LSTM and GRU neural networks for LF, they are compared with the standard RNN model which forms the ground model [21]. The accuracy of the GRU model was found to be better than LSTM for energy consumption prediction.

In order to exploit several features of household energy consumption, CNN has emerged as one of the promising candidates in recent times for efficient and accurate forecasting. Few studies are documented below. CNN for building-level load forecasting is examined and compared with LSTM networks, factored restricted Boltzmann machines (FCRBM), shallow ANN and SVM [22]. The historical load data are passed through convolutional layers, and the output of CNN is fed to a dense layer along with other relevant informations. Similarly, a parallel CNN-RNN model with a three-layer DNN is used for feature extraction from the historical load and handling sequential dynamics, respectively, and compares the performance of the model with linear regression, SVR, DNN, and CNN_RNN [23]. A hybrid forecasting model combining the characteristics of CNN and LSTM for household energy forecasting is reported in [24]. This hybrid model outperforms other machine learning techniques for individual residence forecasting.

Load forecasting at residential level for individual consumers is challenging due to the inconsistency in their energy consumption behavior causing unfavorable forecasting accuracy. It is easy to forecast the usage behavior of household having regular routines or lifestyles. Also, at community level microgrid, inclusion of features such as holidays, weekends, varying electricity prices has improved the forecasting accuracy. The proposed approach includes historical data of hourly varying solar insolation as one of the features for prediction of day-ahead load demand.

For prediction of PV power, a deep learning forecasting model is presented in [25] with LSTM-based auto-encoder

for the purpose of energy management in the presence of an optimized electric vehicle charging station. As solar PV power generation highly depends on climatic conditions, cloud cover plays a vital role and can cause poor operation of PV modules. Partial shedding of PV cell may cause generation drop by 10–25%. Hence, it is an inevitable process experienced by the modules. A weather classification-based PV power forecasting is documented in [26]. Four different SVM models are used each for clear, cloudy, foggy and rainy day, whereas in [27], cloud amount is included along with weather type classification of sunny, cloudy and rainy days and radial basis function neural network (RBFNN) model is implemented for each classification for day ahead PV power prediction. However, relation between the impact of cloud amount and PV power generation for each category would have been established using some mathematical equation.

Nevertheless, the literature discussed above focus on individual load and insolation forecasting at service provider level and ignores the prime importance of residential users at community level. Future smart grid requires intelligent forecasting techniques for residential consumers. Proper planning of on-site generation and consumption ensures smooth local energy management between PV, energy storage and utility grid. Furthermore, prior idea of forecasted variables will be useful in studying the degradation performance of energy storage device under various weather conditions. In order to validate the above study, we propose a combined load and insolation forecasting at local community level considering the impact of cloud cover on forecasting accuracy. The incorporation of cloud cover has a reasonable impression on energy consumption as well as PV generation. High amount of cloud cover causes humid weather with high energy consumption while drop in PV power generation.

Weather classification-based approach has been reported in the past; however, it is restricted to insolation forecasting. The proposed approach uses cloud cover as an additional feature for both load and insolation forecasting. In our work, we propose three LSTM models (each for cloudy, sunny and rainy) based on cloud cover for LF and one LSTM model for insolation forecasting. We also propose the modified clearness index (MCI) for calculating the PV power generation. The performance of proposed LSTM model is compared with one statistical method—autoregression (AR), and two machine learning methods—ANN and CNN, in order to show its effectiveness. Our LSTM model is also compared with a hybrid CNN-LSTM model used in [24] for forecasting individual household energy consumption to prove the superiority of the proposed approach.

The main contributions of this paper are:

1. Day-ahead load and solar insolation forecasting of residential consumers in a grid-tied rooftop PV microgrid using a proposed LSTM model.
2. Establishment of a novel framework based on weather classification by incorporating the concept of cloud amount along with clearness index in PV power generation.

The paper is structured as follows: Sect. 2 overviews various forecasting methodologies; Sect. 3 gives the methodology proposed for load and insolation forecasting; Sect. 4 gives the model selection criterion through comparison; in Sect. 5, the experimental results are discussed; and Sect. 6 gives the conclusive remarks and future scope.

2 Forecasting methodologies

This section explains the overview of various forecasting methodologies adapted in this work. Both statistical and machine learning methods are discussed along with their basic structure followed by the detailed description of the related methods.

2.1 Autoregression

Autoregressive model is one of the simplest types of statistical time series forecasting methodology. As the name suggests, it takes into account the previous time steps information as input to predict the value at the next time step, i.e., regression with itself as visible from (1).

$$y_t = C_1 y_{t-1} + C_2 y_{t-2} \dots C_p y_{t-p} \quad (1)$$

where $t-1, t-2, \dots t-p$ are the lagged values of forecasting variables, $C_1, C_2 \dots C_p$ are coefficients, and p is the order of autoregressive model.

2.2 Artificial neural network

Neural networks are an interconnection of neurons to perform computation efficiently. It adapts to the changing environment and mimics the human brain. It is capable of handling noisy data and can perform a nonlinear mathematical operation with ease. It consists of an input layer, a hidden layer, and an output layer. The number of neurons/nodes in each layer is decided based on the requirement. FFNN is the most widely used topology.

Figure 1 shows the basic structure of a neural network. A neural network with two or more hidden layers is often referred to as DNN. Mathematically the output of a neural network can be expressed as,

$$O_j = f \sum_{k=1}^n (w_{jk} x_k) + b_{jk} \quad (2)$$

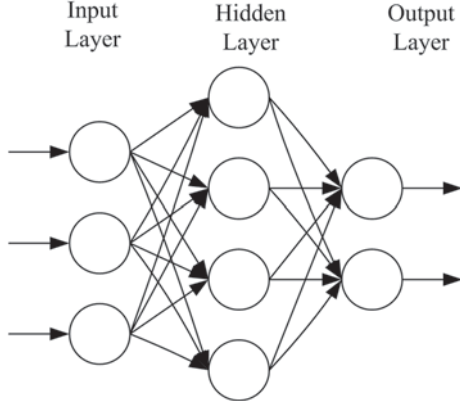


Fig. 1 Basic schematic of a neural network

where O_j is the output of neural network, f is the activation function, $w_{jk}x_k$ is the weighted inputs to the neural network, and b_{jk} is the biases.

Each input node is connected to each hidden node through weights and bias. A variety of activation functions are available to determine the output of a neural network. The most widely used activation functions are differentiable and non-linear, such as sigmoid function, hyperbolic tangent function (tanh), rectified linear unit (ReLU) function. Authors in [28] conducted a study to analyze the performance of these functions.

2.3 Convolutional neural network

Convolutional neural network (CNN) belongs to the family of ANN and is specially designed for image processing. They can be used with one-, two- and three-dimensional data. Convolutional layer forms the heart of CNN, it performs linear operation by assigning weights and biases and successfully captures the spatial and temporal dependencies. Figure 2 shows the generalized architecture of CNN.

The input image or text is fed to the convolutional layer with an appropriate input shape and size of kernels followed by a MaxPooling layer with ReLU as an activation function. More number of such layers can be connected in cascade manner for better feature extraction. Finally, a fully connected dense layer is connected to give the output. The ability to deal with noisy data and to extract valuable characteristics from the time series data, CNN has shown effectiveness over other methods.

2.4 Long short-term memory

To overcome the drawback of vanishing or exploding gradient, RNN was introduced and has shown impressive performance while dealing with sequential data. The presence of memory unit in RNN ensures backward connection

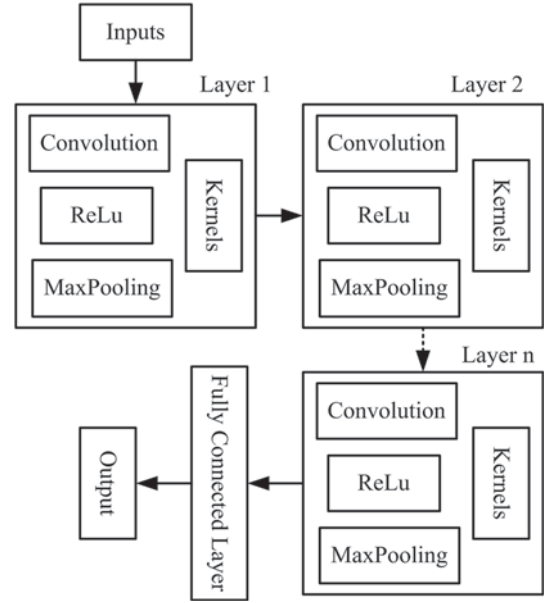


Fig. 2 CNN architecture

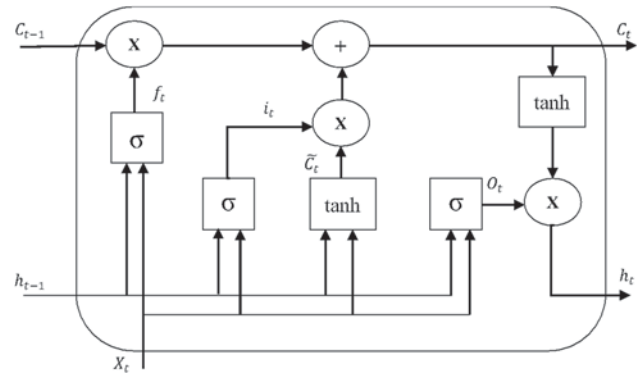


Fig. 3 Schematic diagram of an LSTM cell

for storage of information. Due to its high computational time and fading of inputs, LSTM, a modified version of RNN, was introduced for capturing long-term dependencies. The primitive structure of LSTM is shown in Fig. 3. As the name suggests, it has two states, C_t and h_t for long-term and short-term, respectively. The operations are handled by three gates, mainly input gate, forget gate, and output gate, and are described as follows.

The input X_t and h_{t-1} are fed to all the layers, i.e., f_t , i_t , \tilde{C}_t and O_t where \tilde{C}_t uses tanh as an activation unit, while others are using sigmoid. The function of forget gate layer is to identify the relevant information required to be retained for further operation, while irrelevant information will be thrown away from the cell state as given by (3),

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (3)$$

Input gate layer decides new information to be stored in the cell state expressed as,

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (4)$$

$$\tilde{C}_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (5)$$

After this process, old cell state gets updated to new cell state by performing element-wise multiplication and addition given by (6),

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (6)$$

Finally, part of the cell state is processed by the output gate layer as,

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (7)$$

where h_{t-1}, C_{t-1} are the outputs of previous cell, x_t is the new input vector, \tilde{C}_t is the intermediate output, $w_f, b_f, w_i, b_i, w_c, b_c, w_o, b_o$ are the weights and biases for each layer, respectively, σ and \tanh are the activation units, and $+$, \times indicate pointwise operation.

The main aim of this study is to forecast the energy generation and consumption behavior of a residential consumer equipped with solar rooftop PV. While power system planning studies usually take place at service provider end, our work focuses on load and solar insolation forecasting at community level. Advanced metering infrastructure (AMI) allows measurement of consumer's energy consumption based on minutely or hourly resolution via smart meters installed at consumer's premises, while pyranometer is used for measuring global solar radiation. As forecasting techniques need huge amount of historical data for predicting the future data to understand the behavior of consumers, preserving crucial information for long time is necessary.

Hence, to deal with sequential data available through measurement devices, to understand the energy generation and consumption pattern, and to capture long-term dependency, forecasting is carried out using LSTM with series to supervised learning approach.

3 Proposed methodology

3.1 Weather-based classification

Weather is a mixture of events happening in our day-to-day life in different parts of the world. Air pressure, temperature, humidity, wind speed, wind direction, cloud cover (cloud amount), etc., are the various parameters of measurement defining a climate.

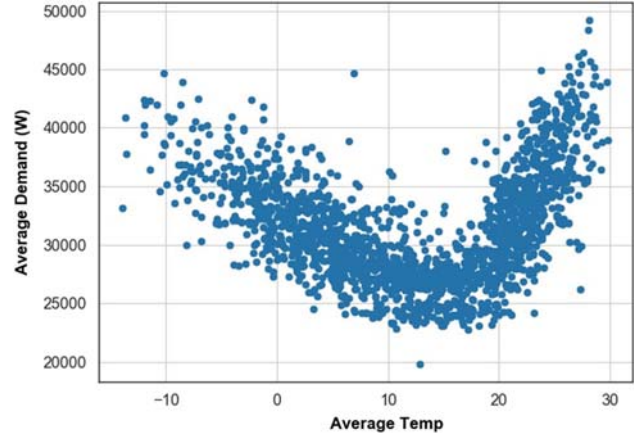


Fig. 4 Relationship between daily average temperature and load demand

Table 1 Cloud cover coding by WMO in Oktas and Tenths

Code	Oktas	Tenths	Category
0	0, no cloud	0, no cloud	Fine
1	≤ 1 , but not 0	≤ 1 , but not 0	Fine
2	2	2–3	Fine
3	3	4	Partly cloudy
4	4	5	Partly cloudy
5	5	6	Partly cloudy
6	6	7–8	Cloudy
7	≥ 7 , but not 8	≥ 9 , but not 10	Cloudy
8	8	10	Overcast

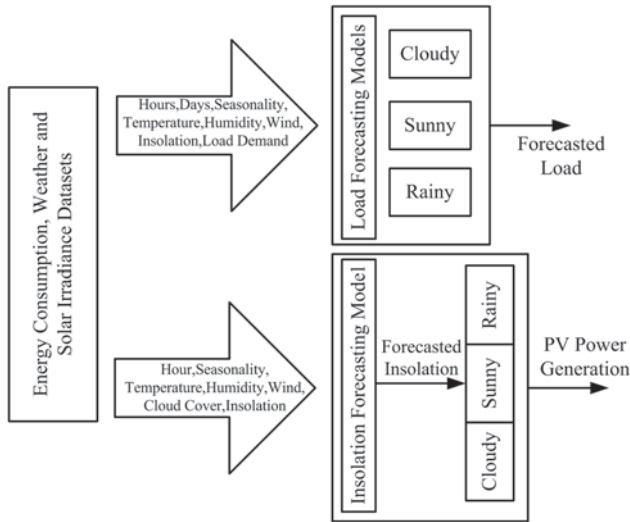
Weather variables (such as temperature, humidity, wind, solar radiation) are responsible for energy consumption behavior of customers as well as planning of solar PV generation. Extreme high and low temperatures may result in higher energy consumption as visible from Fig. 4. It shows the relation between daily average load demand with respect to daily average temperature, pre-processed from the datasets. Similarly, the amount of solar radiation reaching earth's surface depends on latitude of the location, sunrise and sunset hour angle, daily maximum possible sunshine duration and amount of cloud cover.

In meteorology, an Okta is a unit of measurement used to describe the amount of cloud cover at any given location [29]. Cloud cover (CC) in accordance with the recommendations by World Meteorological Organization (WMO) [30, 31] can be categorized as shown in Table 1. It is assessed by estimating the total sky area covered with clouds, i.e., how many eighths/tenths of the sky are covered in cloud, ranging from 0 Oktas/0 Tenths to 8 Oktas/10 Tenths.

As we group our original dataset into sunny, cloudy and rainy, the correlation between input variables belonging to the same group and the output variable becomes strong as

Table 2 Correlation among variables

	Combined (%)	Sunny (%)	Cloudy (%)	Rainy (%)
Load demand	13.5	12.83	1.22	29.26
Insolation	- 1.5	- 0.4	2.88	0.38

**Fig. 5** Workflow of the proposed problem

evident from Table 2. Based on this, we can categorize cloud cover as sunny (0–3), cloudy (4–6) and rainy (7–10) where 0 corresponds to 0% cloud amount and 10 corresponds to 100% cloud amount.

Figure 5 shows the workflow of the proposed problem. The original dataset is grouped into three individual datasets, namely sunny, cloudy and rainy, each having their own LSTM networks for LF and IF. The dataset contains weather-dependent variables, time and day information with hourly resolution. After grouping, each individual dataset is reframed as supervised learning problem before actually passing through the neural network. Each model has different numbers of hidden layer neurons calculated as per the equations shown in upcoming section. Each model will predict load demand and solar insolation for the next 24 h, and the corresponding PV power generation can be calculated. The positive difference between forecasted load and PV power generation gives the deficit power required to satisfy the demand through ESS and utility grid.

Although weather-based methodologies for forecasting solar PV generation are documented in the literature, our approach presents a novel concept by integrating load forecasting and insolation forecasting along with the concept of cloud cover. Due to the rising trend of AMI, it is easy for the residential consumers, with rooftop PV installed, to have weather-based forecasting model in their premises.

This arrangement empowers prosumers to have prior knowledge of energy generation and consumption pattern using personal area network (PAN), thereby maximizing the self-consumption through PV and ESS while minimizing the cost of electricity. As our approach is not restricted to single household, but a group of prosumers, the common load shared by the prosumers, are limited in numbers such as charging of electric vehicles, water pumping for overhead tank. Hence, to improve the local energy management and to ensure smooth operation of PV-ESS, these common loads can utilize the local generation as desired.

It also assures privacy and security aspects of smart grid for the individual consumers which remains an obstacle in the success of smart grid technologies. As forecasting is done at the prosumer's end, it does not jeopardize the privacy of the customer unlike forecasting at service provider end, where the sensitive personal data are collected and accessed by the unknown people. Similarly, arrangement at prosumer's end is less exposed to false data injection and ensures more secured network in local energy community, whereas it is more prone to malicious cyber-attacks in a centralized environment.

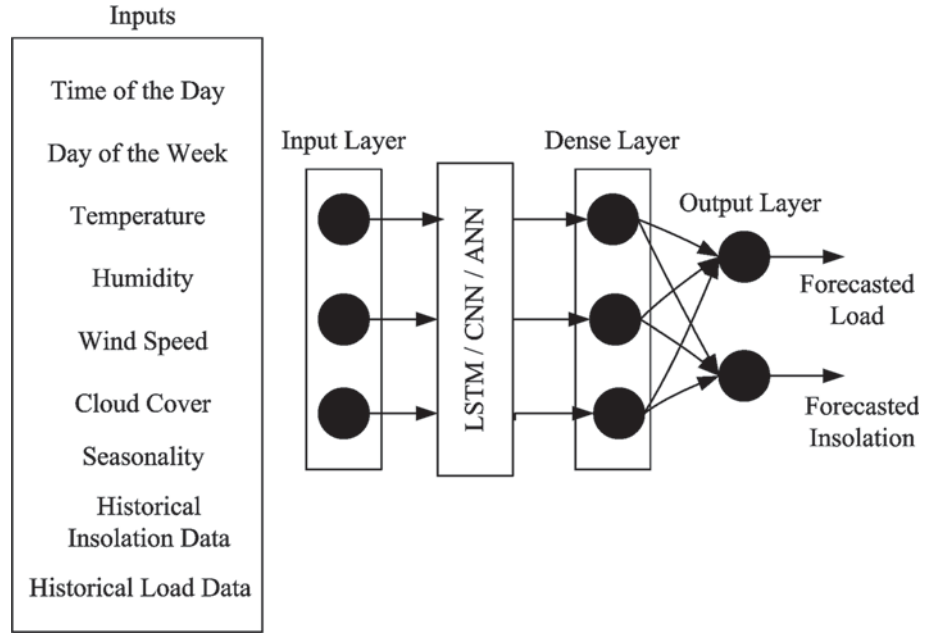
3.2 Architecture

The proposed architecture consists of one input layer, two hidden layers and an output layer as shown in Fig. 6. The first hidden layer is an LSTM layer. Hourly varying weather variables along with historical energy consumption data are used as inputs. The size of the input and output layers is same as the number of inputs and output, while selection of number of neurons in hidden layers is discussed in subsequent subsection.

The characteristics of a multivariate time series are extracted by performing data pre-processing such as seasonal decompose, lagged loads, weekdays vs. weekends. In this paper, seasonal decomposition and lagged load are derived from the datasets. After data pre-processing and data cleaning, day-ahead forecasted load and solar insolation are generated by a fully connected dense output layer, and the model is analyzed by evaluating performance metrics. Various parameters such as hidden neurons, activation function can be adjusted to improve the performance of the model. Design of architecture depends on size of the datasets and may vary based on various applications. The number of hidden neurons is altered to check the performance of the model as described in Sect. 3.3. Also, first hidden layer is replaced by a conventional dense layer, a CNN layer and a hybrid CNN-LSTM layer, one at a time to compare with the proposed method.

The input to LSTM is a time series signal containing hourly energy consumption information as well as solar insolation. It is a well-known fact that weather-dependent variables play a key role while forecasting load and solar

Fig. 6 Overall architecture of the forecasting model



insolation and hence inclusion of these exogenous variables is important to enhance the performance of our model. Therefore, the multivariate time series problem is reframed as a supervised learning problem, i.e., from sequence to pairs of input and output sequences. This is done to ensure utilization of various linear and nonlinear ML algorithms based on problem.

As our problem is reframed as supervised learning, selection of input and target variables from the given data frame for a neural network plays a vital role. Values of each variable at different time steps of ' t ' are taken as input to the model. The input to LSTM for LF is 17×1 for each weather-based model and 16×1 for solar IF with 1-h resolution. It is passed through a dense layer and an output layer. Table 3 summarizes the LSTM models for LF and IF.

3.3 Selection of hidden layer neurons

The selection of the number of hidden neurons/nodes is a tricky challenge possessed by neural networks. The performance of any neural network is evaluated based on error metric and poor choice of hidden neurons can influence the error by making the network unstable. Random selection may cause an under-fitting or over-fitting problem. Authors in [32] review several methods for fixing the number of hidden neurons while proposing a method for wind speed prediction.

In this work, number of hidden neurons is altered to check the performance of our model. A Keras Tuner library is used for selecting the optimal number of hidden neurons along with hidden layer neurons calculated using Eqs. (8)–(10), as in [33], with a variation of ± 5 neurons.

Table 3 Summary of the proposed LSTM models

Model	Layer (type)	Output shape	Parameters #
<i>Load forecasting</i>			
Cloudy	LSTM	34	7072
	Dense	23	805
	Dense	1	24
Total parameters			7901
Sunny	LSTM	18	2592
	Dense	30	570
	Dense	1	31
Total parameters			3193
Rainy	LSTM	34	7072
	Dense	30	1050
	Dense	1	31
Total parameters			8153
<i>Insolation forecasting</i>			
	LSTM	16	2112
	Dense	24	408
	Dense	1	24
Total parameters			2545

$$N_h = \frac{N_i + N_o}{2} + \sqrt{N_s} \quad (8)$$

$$N_h = \frac{2 * N_i}{3} + N_o \quad (9)$$

$$N_h \leq 2 * N_i \quad (10)$$

where N_h is the number of hidden neurons, N_i and N_o are the number of inputs and outputs, and N_s is the total number of samples in the dataset.

Table 4 Dataset features

Features	Description
Hour of the day	hh:mm:ss
Day of the week	Starting from Monday
Seasonality	Captured from seasonal decomposition
Temperature	Hourly varying temperature, °C
Humidity	Hourly varying humidity, %
Wind	Hourly varying wind, m/s
Cloud cover	Sunny, cloudy, rainy
Insolation	Hourly varying solar irradiance, W/m ²
Clearness index	Clearness of sky
Load demand	Hourly varying load, W

Table 5 Variables to forecasting models

Load forecasting	Insolation forecasting
<i>Inputs</i>	
Hour of the day	Hour of the day
Day of the week	Seasonality
Seasonality	Temperature
Temperature	Humidity
Humidity	Wind
Wind	Cloud cover
Insolation	Clearness Index
Load demand	Insolation
<i>Outputs</i>	
Load demand at $t + 1$	Insolation at $t + 1$

4 Model selection through comparative experimental results

4.1 Datasets

The proposed method is implemented on hourly energy consumption dataset with *Python 3.7.4*. It contains load consumption measurements between October 2012 and November 2017 with hourly resolution [34], whereas a weather dataset contains record of temperature, humidity, wind, and cloud cover description [35]. In addition to these, hourly varying solar irradiance is also used [36]. These datasets are combined as a multivariate time series, and a ratio of 70:30 is allotted for training and testing of LSTM network.

The total load demand from the original dataset is scaled down to achieve the magnitude, equivalent to residential sector, which is further divided among two type of customers with a ratio of 60:40 where 60% are the one with rooftop PV generation, known as ‘Prosumers,’ while 40% are traditional consumers. Datasets features are depicted in Table 4. In Table 5, details of various inputs to model are given.

4.2 Performance metrics

The difference between the actual value (y) and the model’s prediction \hat{y} , known as residual, plays a compelling role in model assessment. Mean absolute error (MAE), the simplest of error metric, describes positive magnitude of residuals averaged over total samples (n), whereas mean bias error (MBE) is similar to MAE without the absolute operation and estimates the average bias in the prediction as shown by (11) and (12), respectively. MBE indicates under-performance or over-performance of model due to the absence of absolute operation which is not the case with MAE. Also, MAE is more robust to outliers as it gives less weightage to them. This study includes both MAE and MBE for the comparison of the effectiveness of the proposed model.

Moreover, a coefficient of determination (R -Squared) is used to assess the regression accuracy of the proposed approach given by (13). It indicates the correlation between the actual and forecasted values where values close to 1 is better.

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (11)$$

$$\text{MBE} = \frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j) \quad (12)$$

$$R\text{-Squared} = 1 - \frac{\sum (y_j - \hat{y}_j)^2}{\sum (y_j - \bar{y})^2} \quad (13)$$

4.3 Variation in hidden layer neurons

Number of neurons are varied in hidden layer using Keras tuner. It gives an optimal number of hidden layers/neurons by tuning the hyperparameters specified while designing the neural network sequential model. The performance of forecasting models is evaluated and compared with Eqs. (8)–(10).

AR, ANN, CNN and LSTM models are simulated and compared using the concept of weather-based classification proposed in this paper. Tables 6, 7, 8 and 9 show the comparison of various models for both LF and IF methodologies in different weather conditions.

Overall comparison shows that LSTM models suit best for the proposed methodology where the original dataset is grouped as per weather classification and hidden layer neurons are adjusted as per Eqs. (8)–(10) and Keras tuner library. Performance metrics in bold digits indicate the superiority of LSTM model over others in Tables 6, 7, 8 and 9.

Also, the proposed method is compared with one of the recent studies where a hybrid CNN-LSTM is used as a forecasting model for individual household energy consumption. The model parameters are adopted from [24]. As noticeable from Table 9, the overall comparison of MAE and MBE

Table 6 Comparison between various forecasting models evaluated as per Eq. (8)

Weather	Model	MAE	MBE
<i>LF</i>			
Cloudy	AR	0.0563	3.0403
	ANN	0.0699	- 0.0303
	CNN	0.0678	- 0.0312
	LSTM	0.0675	- 0.0307
Sunny	AR	0.0299	- 0.4855
	ANN	0.0289	0.0052
	CNN	0.0318	0.0132
	LSTM	0.0258	0.0081
Rainy	AR	0.0610	8.3803
	ANN	0.0622	- 0.0240
	CNN	0.0852	- 0.0692
	LSTM	0.0582	- 0.0964
<i>IF</i>			
NA	AR	0.1792	0.0230
	ANN	0.0534	- 0.0037
	CNN	0.1059	- 0.0133
	LSTM	0.0513	- 0.0090

Bold values of MAE and MBE indicate the superiority of the proposed model

Table 7 Comparison between various forecasting models evaluated as per Eq. (9)

Weather	Model	MAE	MBE
<i>LF</i>			
Cloudy	AR	0.0563	3.0403
	ANN	0.0643	- 0.0068
	CNN	0.0664	- 0.0303
	LSTM	0.0526	- 0.0134
Sunny	AR	0.0299	- 0.4855
	ANN	0.0383	0.0170
	CNN	0.0305	0.0051
	LSTM	0.0269	0.0047
Rainy	AR	0.0610	8.3803
	ANN	0.0808	- 0.0510
	CNN	0.0714	- 0.0318
	LSTM	0.0640	- 0.0297
<i>IF</i>			
NA	AR	0.1792	0.0230
	ANN	0.0608	0.0031
	CNN	0.1013	0.0017
	LSTM	0.0560	- 0.0186

Bold values of MAE and MBE indicate the superiority of the proposed model

Table 8 Comparison between various forecasting models evaluated as per Eq. (10)

Weather	Model	MAE	MBE
<i>LF</i>			
Cloudy	AR	0.0563	3.0403
	ANN	0.0554	- 0.0031
	CNN	0.0660	- 0.0289
	LSTM	0.0542	- 0.0062
Sunny	AR	0.0299	- 0.4855
	ANN	0.0298	0.0039
	CNN	0.0300	0.0130
	LSTM	0.0297	0.0115
Rainy	AR	0.0610	8.3803
	ANN	0.0665	- 0.0199
	CNN	0.0692	- 0.0251
	LSTM	0.0675	- 0.0323
<i>IF</i>			
NA	AR	0.1792	0.0230
	ANN	0.0511	- 0.0193
	CNN	0.0503	- 0.0113
	LSTM	0.0505	- 0.0163

Bold values of MAE and MBE indicate the superiority of the proposed model

Table 9 Comparison between various forecasting models evaluated by Keras Tuner

Weather	Model	MAE	MBE	R-squared
<i>LF</i>				
Cloudy	AR	0.0563	3.0403	0.6957
	ANN	0.0530	- 0.0104	0.7087
	CNN	0.0583	- 0.0168	0.6840
	CNN-LSTM	0.0719	- 0.0185	0.6006
Sunny	LSTM	0.0519	- 0.0053	0.7160
	AR	0.0299	- 0.4855	0.9030
	ANN	0.0269	- 0.0018	0.9063
	CNN	0.0316	0.0098	0.9027
Rainy	CNN-LSTM	0.0593	0.0252	0.7535
	LSTM	0.0257	0.0128	0.9033
	AR	0.0610	8.3803	0.6972
	ANN	0.0631	- 0.0216	0.6625
<i>IF</i>	CNN	0.0863	- 0.0625	0.6738
	CNN-LSTM	0.0713	0.0019	0.6509
	LSTM	0.0591	- 0.0169	0.6862
	AR	0.1792	0.0230	0.8930
NA	ANN	0.0617	- 0.0072	0.9824
	CNN	0.0541	- 0.0028	0.9910
	CNN-LSTM	0.0486	- 0.0205	0.9882
	LSTM	0.0426	- 0.0103	0.9823

Bold values of MAE, MBE, and R-squared indicate the superiority of the proposed model

Table 10 Overall comparison of weather-based LSTM models as per Eqs. (8)–(10) and Keras Tuner

N_h calculation using	Weather	MAE	MBE
<i>LF</i>			
Equation (8)	Cloudy	0.0675	− 0.0307
	Sunny	0.0258	0.0081
	Rainy	0.0582	− 0.0964
Equation (9)	Cloudy	0.0526	− 0.0134
	Sunny	0.0269	0.0047
	Rainy	0.0640	− 0.0297
Equation (10)	Cloudy	0.0542	− 0.0062
	Sunny	0.0297	0.0115
	Rainy	0.0675	− 0.0323
Keras Tuner	Cloudy	0.0519	− 0.0053
	Sunny	0.0257	0.0128
	Rainy	0.0591	− 0.0169
<i>IF</i>			
Equation (8)	NA	0.0513	− 0.0090
Equation (9)		0.0560	− 0.0186
Equation (10)		0.0505	− 0.0163
Keras Tuner		0.0426	− 0.0103

Bold values of MAE and MBE indicate the superiority of the Keras Tuner

shows the effectiveness of the proposed forecasting model in all the cases of LF and IF, whereas the accuracy of our model is superior in most of the cases.

As mentioned earlier, there is no thumb rule for fixing number of hidden neurons, and it is totally problem specific. Table 10 depicts the overall comparison of each weather-based LSTM models as per Eqs. (8)–(10) and Keras tuner library. The superiority of Keras tuner can be observed in Table 10. In Eqs. (8)–(10), N_h largely depends on the number of inputs, outputs and samples in the datasets which may be problem specific, whereas in case of Keras tuner, irrespective of these factors it tries to give the set of hidden neurons/layers that optimizes the model performance and outperforms the other hidden neurons selection criterion.

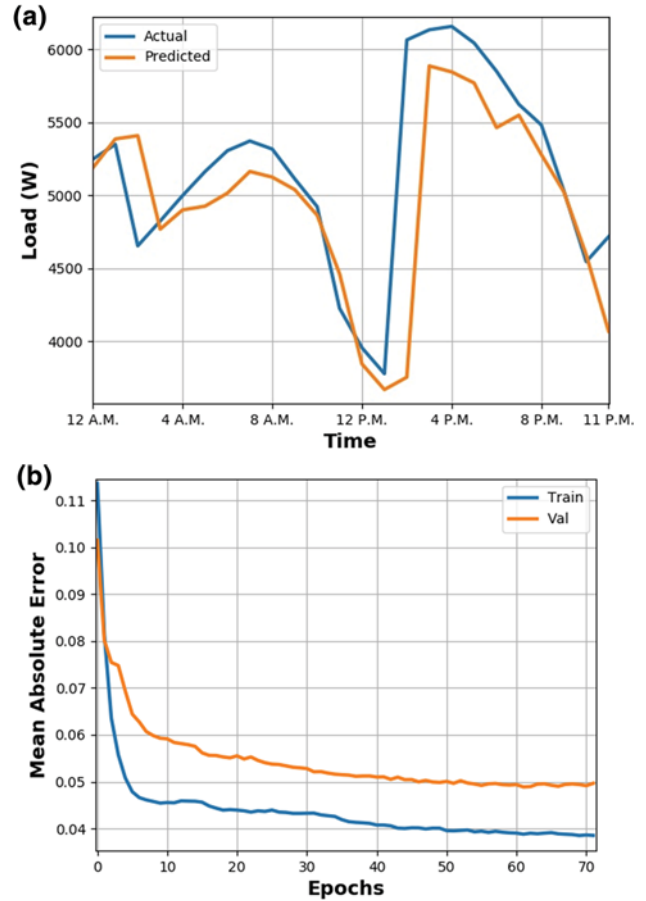


Fig. 7 a Forecasted load for cloudy weather b error plot for cloudy weather forecasted load

5 Forecasting of load and solar power generation based on proposed model

5.1 Load and solar insolation forecasting

LSTM neural network is used to predict the day-ahead load and solar insolation in a deregulated residential community. LF is accomplished by using individual forecasting models for cloudy, sunny and rainy weather as shown in Figs. 7a, 8a and 9a, respectively, and the corresponding error plot is given by Figs. 7b, 8b and 9b, respectively. Hourly energy consumption dataset is used to assess the model performance. Weather classification is totally based on the concept of cloud cover as depicted in Table 1.

Figure 7a shows the day-ahead actual and predicted load under cloudy weather with an accuracy of 0.7160, whereas Figs. 8a and 9a show the day-ahead actual and predicted load under sunny and rainy weather with an accuracy of 0.9033 and 0.6862, respectively. Similarly, Figs. 7b, 8b and 9b indicate the corresponding error plots for training and validation dataset. Here, the mean absolute error is plotted with respect

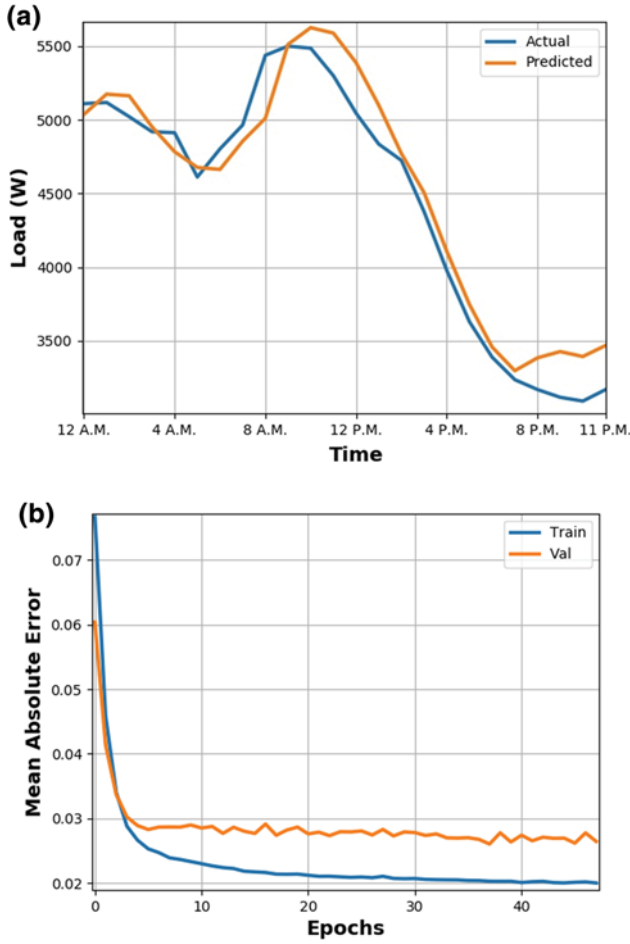


Fig. 8 a Forecasted load for sunny weather b error plot for sunny weather forecasted load

to number of epochs (max.100) and an early stopping criterion is used to avoid over-fitting.

It can be seen that the actual and predicted values of the load demand under various weather conditions are very close to each other and the same is visible from the regression accuracy metric (R -Squared). Hence, the proposed model has better forecasting accuracy indicating reduced supply–demand mismatch. Table 11 gives the basic details of load demand (minimum, maximum and average values) under different weather conditions for both actual and predicted cases. From the values in Table 11, the effectiveness of the proposed model for LF may be established.

For solar insolation, an LSTM model is used for predicting day ahead insolation with a maximum irradiance of 1057.5 W/m^2 . The efficacy of the proposed methodology is evident from Fig. 10a, and the same has been validated using R -squared metric with an accuracy of 0.9823. Also, Fig. 10b represents the error plot with respect to number of epochs (max.100) obtained via simulation for both training and validation data. As mentioned before, early stopping criterion is used to avoid over-fitting. The data used for this method

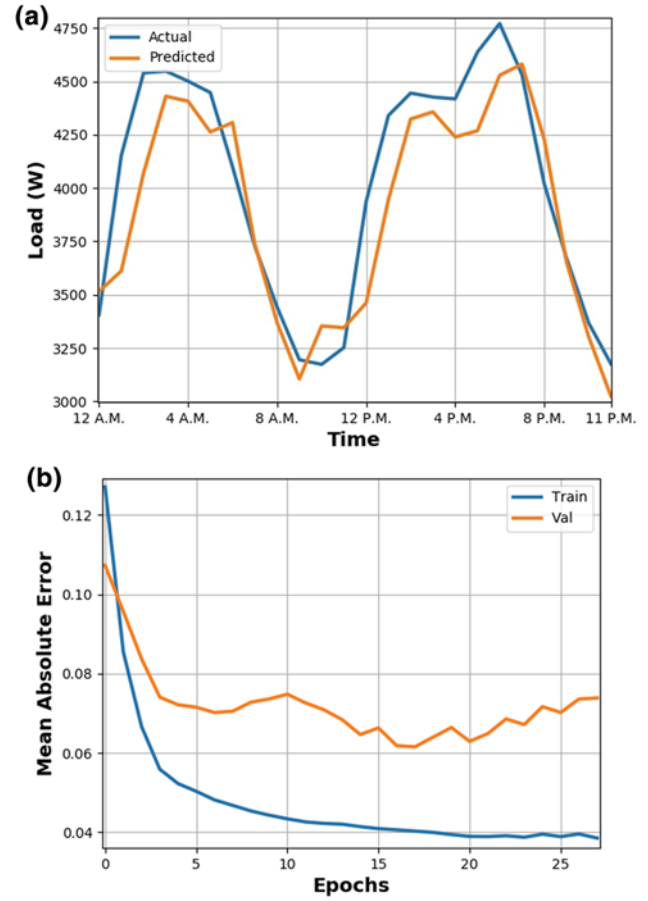


Fig. 9 a Forecasted load for rainy weather b error plot for rainy weather forecasted load

Table 11 Minimum, maximum and average load demand under different weather conditions

	Actual load (W)			Predicted load (W)		
	Min.	Max.	Ave.	Min.	Max.	Ave.
Cloudy	3779.7	6156.2	5161.5	3575.1	5924.3	4910.5
Sunny	3091.2	5499.7	4455.8	3297.4	5624.9	4545.1
Rainy	3173.4	4771.4	4009.6	3024.1	4581.2	3893.0

correspond to PJM region with an average day length varying between 9 to 15 h throughout the year. Apart from weather-dependent variables, cloud cover and clearness index are included as input to forecasting models because of their good correlation with the solar insolation.

5.2 Calculation of solar PV generation

After IF, the PV power generation (P_{pv}) can be calculated as in (14),

$$P_{pv} = \eta_{pv} * \eta_{inv} * \eta_{conv} * A * G / 1000 \quad (14)$$

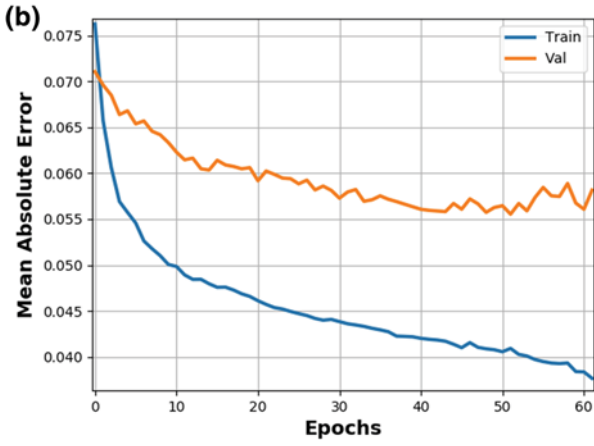
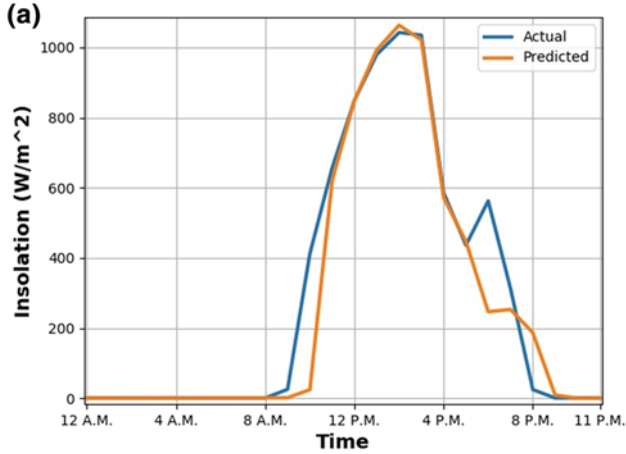


Fig. 10 a Forecasted insolation b error plot for forecasted insolation

where η_{pv} , η_{inv} , and η_{conv} are the efficiencies of PV panel, DC-AC inverter and DC-DC converter respectively, 'A' is the area required for the PV panels in m^2 , 'G' is the hourly solar irradiance in kW/m^2 .

Due to the presence of clouds and various atmospheric layers, a criterion, known as *Clearness Index* (K_t), is used to determine the clearness of sky using solar irradiance data expressed by (15),

$$K_t = I/I_0 \quad (15)$$

It is the ratio of hourly global solar irradiance I to the extra-terrestrial solar irradiance I_0 . Different authors have adopted different criteria to use clearness index. In [37], $K_t > 0.6$ and < 0.2 are considered as clear and cloudy, respectively, whereas [38] proposes $K_t > 0.65$ as clear and $0.12 \leq K_t \leq 0.35$ as cloudy. Similarly, authors in [39, 40] suggest $0 \leq K_t \leq 0.15$, $0.15 < K_t < 0.7$ and $K_t > 0.7$ as overcast, cloudy and clear skies, respectively.

Low value of K_t suggests high amount of cloud cover and vice versa. Also, as mentioned in Table 1, amount of cloud cover is coded in Oktas and Tenths for fine, partial cloudy and

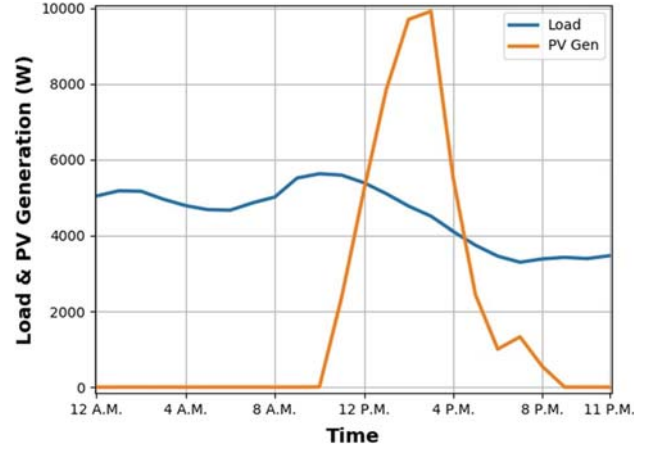


Fig. 11 PV power generation and forecasted load under sunny weather

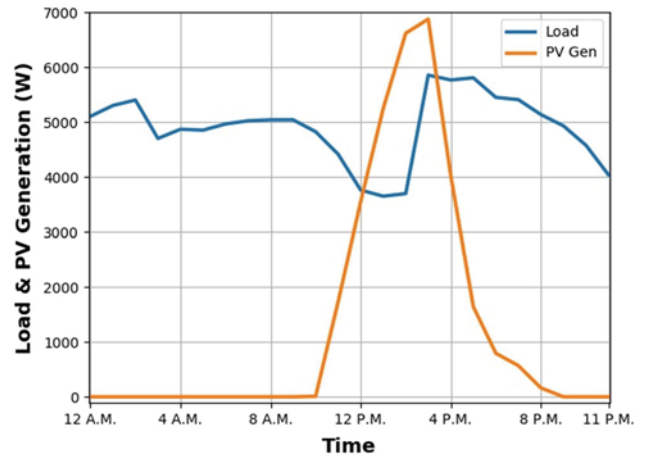


Fig. 12 PV power generation and forecasted load under cloudy weather

overcast conditions. In this paper, modified clearness index (MCI) is proposed and is expressed as shown by (16):

$$\tilde{K}_t = K_t * (1 - CA) \quad (16)$$

where \tilde{K}_t is the MCI, and 'CA' is the cloud amount ranging between 0 and 1 based on categorization. So, now PV power generation will be calculated as,

$$P_{pv} = \eta_{pv} * \eta_{inv} * \eta_{conv} * A * \tilde{K}_t * G/1000 \quad (17)$$

Based on the forecasted output from LSTM model, PV power generation is calculated as per Eq. (17). Figures 11, 12 and 13 show the PV power generation along with forecasted load corresponding to different weather conditions. Table 12 shows the solar PV specifications used here for the study [41].

It can be seen clearly that amount of cloud covered in the sky has a reasonable impact on PV power generation, i.e., during sunny weather (Fig. 11), the prosumers have sufficient PV

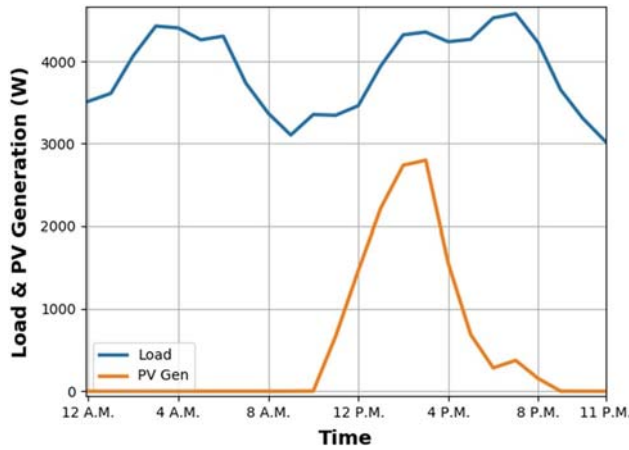


Fig. 13 PV power generation and forecasted load under rainy weather

Table 12 Solar PV specifications

Parameters	Value
Module efficiency (η_{pv})	16%
Inverter efficiency (η_{inv})	90%
Converter efficiency (η_{conv})	90%
Rooftop area (A)	100.30 m ²
Cloud index (MCI)	0–0.30 for Sunny 0.31–0.60 for Cloudy 0.61–1.0 for Rainy
Rated PV output (P_{pv})	13 kW _p
Rating of PV module	325 W _p
Module dimensions ($L \times W \times T$)	(1.96 m \times 1.01 m \times 0.035 m)
Modules in series (string)	10
Strings in parallel (array)	4

generation to satisfy the load demand, while surplus power generated can be stored in the ESS, whereas during cloudy (Fig. 12) or rainy weather (Fig. 13), considerable amount of PV power is lost due to the presence of clouds.

This concept of incorporating cloud amount as well as proposed modified clearness index in PV power generation will be useful while choosing the optimal size of the ESS in order to satisfy the deficit demand during low (or absence of) PV generation.

6 Conclusion

In this work, day-ahead LF and solar IF of residential consumers with rooftop PV units using LSTM are proposed. It establishes a novel framework based on weather classification along with the concept of cloud amount in PV power generation. LSTM has a tendency to capture long-term dependencies and hence produces promising result as

compared to traditional ANN and CNN. The results are also compared with conventional AR technique and a hybrid CNN-LSTM model, to establish the superiority of the proposed algorithm.

Due to the nonlinear correlation of PV power generation with various factors, it is difficult to model a single system that yields accurate forecasting; hence, weather-based classification serves the purpose. Cloud cover is a vital weather factor responsible for planning both generation and consumption. Based on the code given by WMO, cloudy, sunny, and rainy forecasting models are selected. For each weather-based models, the multivariate time series problem is reframed as supervised learning problem, and hidden layer neurons are varied to evaluate the model performance. The selection of the number of hidden layer neurons is also discussed in the study. Overall comparison shows LSTM outperforms ANN, CNN and hybrid CNN-LSTM methods for each of the weather condition. A modified clearness index is proposed here in particular, to show the impact of cloud cover on solar PV power generation. The effectiveness of the proposed methodology is validated using MAE, MBE and R -Squared metrics. The proposed approach may be validated for corrupted measurements by applying filters for outlier detection and the behavior of the forecasting models can be assessed. Another potential limitation of this study is removal of noise from the measured data, which can be assessed using LSTM Auto-encoders having ability to learn compressed representation of the inputs.

The share of green energy in present day is increasing, motivating residential consumers to maximize the self-consumption and storage of the local generation. With the prior knowledge of forecasted variables, the deficit energy demand can be calculated to study the charging/discharging behavior of ESS; hence, the optimal size and cost of ESS can be determined using battery aging model, keeping in mind various technical and economic factors.

Acknowledgements The authors express their sincere gratitude to Professor Alberto Borghetti, Department of Electrical, Electronic, and Information Engineering, University of Bologna, Italy, for his valuable suggestions and guidance for this study. Authors would like to acknowledge Department of Science and Technology, Government of India for the financial support under EU-India RE-EMPOWERED Project (File No.: DST/TMD/INDIA/EU/ILES/2020/50).

References

- Grids M (2003) National Policy for Renewable Energy based Micro and Mini Grids. <http://www.indiaenvironmentportal.org.in/content/430079/draft-national-policy-forrenewable-energy-based-micro-and-mini-grids/>. Accessed 6 Mar 2020
- Kong X, Xu X, Yan Z et al (2018) Deep learning hybrid method for islanding detection in distributed generation. *Appl Energy* 210:776–785. <https://doi.org/10.1016/j.apenergy.2017.08.014>

3. Andrew AM (1998) Reinforcement learning: an introduction. *Kybernetes* 27:1093–1096. <https://doi.org/10.1108/k.1998.27.9.1093.3>
4. Cheng L, Yu T (2019) A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int J Energy Res* 43:1928–1973. <https://doi.org/10.1002/er.4333>
5. Guo Y, Liu Y, Oerlemans A et al (2016) Deep learning for visual understanding: a review. *Neurocomputing* 187:27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>
6. Ramanathan R, Engle R, Granger CWJ et al (1997) Short-run forecasts of electricity loads and peaks. *Int J Forecast* 13:161–174. [https://doi.org/10.1016/S0169-2070\(97\)00015-0](https://doi.org/10.1016/S0169-2070(97)00015-0)
7. Papalexopoulos AD, Hesterberg TC (1990) A regression-based approach to short-term system load forecasting. *IEEE Trans Power Syst* 5:1535–1547. <https://doi.org/10.1109/59.99410>
8. Hernández L, Baladrón C, Aguiar JM et al (2014) Artificial neural networks for short-term load forecasting in microgrids environment. *Energy* 75:252–264. <https://doi.org/10.1016/j.energy.2014.07.065>
9. Kwon B-S, Park R-J, Song K-B (2020) Short-term load forecasting based on deep neural networks using LSTM layer. *J Electr Eng Technol* 15:1501–1509. <https://doi.org/10.1007/s42835-020-00424-7>
10. Khwaja AS, Naeem M, Anpalagan A et al (2015) Improved short-term load forecasting using bagged neural networks. *Electr Power Syst Res* 125:109–115. <https://doi.org/10.1016/j.epsr.2015.03.027>
11. Harris JL, Liu LM (1993) Dynamic structural analysis and forecasting of residential electricity consumption. *Int J Forecast* 9:437–455. [https://doi.org/10.1016/0169-2070\(93\)90072-U](https://doi.org/10.1016/0169-2070(93)90072-U)
12. Essallah S, Khedher A (2019) A comparative study of long-term load forecasting techniques applied to Tunisian grid case. *Electr Eng* 101:1235–1247. <https://doi.org/10.1007/s00202-019-00859-w>
13. Reddy SS (2018) Bat algorithm-based back propagation approach for short-term load forecasting considering weather factors. *Electr Eng* 100:1297–1303. <https://doi.org/10.1007/s00202-017-0587-2>
14. Vaish J, Datta SS, Seethalekshmi K (2020) Short term load forecasting using ANN and ensemble models considering solar irradiance. In: 2020 International conference on electrical and electronics engineering (ICE3). IEEE, pp 44–48
15. Li L, Meinrenken CJ, Modi V, Culligan PJ (2021) Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features. *Appl Energy* 287:116509. <https://doi.org/10.1016/j.apenergy.2021.116509>
16. Hippert HS, Pedreira CE, Souza RC (2001) Neural networks for short-term load forecasting: a review and evaluation. *IEEE Trans Power Syst* 16:44–55. <https://doi.org/10.1109/59.910780>
17. Chen H, Cañizares CA, Singh A (2001) ANN-based short-term load forecasting in electricity markets. In: Proceedings of the IEEE power engineering society transmission and distribution conference. IEEE, pp 411–415
18. Bakirtzis AG, Petridis V, Klartzis SJ et al (1996) A neural network short term load forecasting model for the greek power system. *IEEE Trans Power Syst* 11:858–863. <https://doi.org/10.1109/59.496166>
19. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
20. Kong W, Dong ZY, Hill DJ et al (2018) Short-term residential load forecasting based on resident behaviour learning. *IEEE Trans Power Syst* 33:1087–1088. <https://doi.org/10.1109/TPWRS.2017.2688178>
21. Kumar S, Hussain L, Banarjee S, Reza M (2018) Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster. In: Proceedings of 5th international conference on emerging applications of information technology, EAIT 2018. IEEE, pp 1–4
22. Amarasinghe K, Marino DL, Manic M (2017) Deep neural networks for energy load forecasting. In: IEEE international symposium on industrial electronics. IEEE, pp 1483–1488
23. He W (2017) Load forecasting via deep neural networks. *Procedia Comput Sci* 122:308–314. <https://doi.org/10.1016/j.procs.2017.11.374>
24. Alhussein M, Aurangzeb K, Haider SI (2020) Hybrid CNN-LSTM model for short-term individual household load forecasting. *IEEE Access* 8:180544–180557. <https://doi.org/10.1109/ACCESS.2020.3028281>
25. Suresh V, Janik P, Guerrero JM et al (2020) Microgrid energy management system with embedded deep learning forecaster and combined optimizer. *IEEE Access* 8:202225–202239. <https://doi.org/10.1109/ACCESS.2020.3036131>
26. Shi J, Lee WJ, Liu Y et al (2012) Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans Ind Appl* 48:1064–1069. <https://doi.org/10.1109/TIA.2012.2190816>
27. Chen C, Duan S, Cai T, Liu B (2011) Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol Energy* 85:2856–2870. <https://doi.org/10.1016/j.solener.2011.08.027>
28. Feng J, Lu S (2019) Performance analysis of various activation functions in artificial neural networks. *J Phys Conf Ser* 1237:022030. <https://doi.org/10.1088/1742-6596/1237/2/022030>
29. World Meteorological Organization (2014) Chapter 15. Observation of clouds. WMO Guide to Meteorological Instruments and Methods of Observations. https://library.wmo.int/?lvl=notice_display&id=19671#YUrQnFUzbDc. Accessed 21 June 2020
30. World Meteorological Organization (2008) Guide to instruments and methods of observation. https://library.wmo.int/index.php?id=12407&lvl=notice_display#YUrS7VUzbDc. Accessed 21 June 2020
31. Silva AA, Souza-Echer MP (2016) Ground-based observations of clouds through both an automatic imager and human observation. *Meteorol Appl* 23:150–157. <https://doi.org/10.1002/met.1542>
32. Sheela KG, Deepa SN (2013) Review on methods to fix number of hidden neurons in neural networks. *Math Probl Eng* 2013:1–11. <https://doi.org/10.1155/2013/425740>
33. Frederick MD (1996) *Neuroshell 2 user's manual version 4.2*. Ward Syst Gr Inc, Frederick
34. Mulla R (2017) Hourly energy consumption: over 10 years of hourly energy consumption data from PJM in Megawatts. In: Kaggle. <https://www.kaggle.com/robikscube/hourly-energy-consumption>. Accessed 27 May 2020
35. Beniaguev D (2017) Historical hourly weather data 2012–2017. In: Kaggle. <https://www.kaggle.com/selfishgene/historical-hourly-weather-data>. Accessed 27 May 2020
36. Andrey (2016) Solar radiation prediction: task from NASA hackathon. In: Kaggle. <https://www.kaggle.com/dronio/SolarEnergy>. Accessed 27 May 2020
37. McCormick PG, Suehrcke H (1991) Diffuse fraction correlations. *Sol Energy* 47:311–312. [https://doi.org/10.1016/0038-092X\(91\)90123-E](https://doi.org/10.1016/0038-092X(91)90123-E)
38. Kuye A, Jagtap SS (1992) Analysis of solar radiation data for Port Harcourt, Nigeria. *Sol Energy* 49:139–145. [https://doi.org/10.1016/0038-092X\(92\)90148-4](https://doi.org/10.1016/0038-092X(92)90148-4)

39. Li DHW, Lau CCS, Lam JC (2004) Overcast sky conditions and luminance distribution in Hong Kong. *Build Environ* 39:101–108. <https://doi.org/10.1016/j.buildenv.2003.06.001>
40. Li DHW, Lam JC (2001) An analysis of climatic parameters and sky condition classification. *Build Environ* 36:435–445. [https://doi.org/10.1016/S0360-1323\(00\)00027-5](https://doi.org/10.1016/S0360-1323(00)00027-5)
41. Adani Solar (2018) Solar panel specifications datasheet. “5BB multi-crystalline solar PV modules -1500V series”. [https://www.](https://www.adanisolar.com/-/media/Project/AdaniSolar/Media/Downloads/Downloads/Datasheets/Multi-Encore/Multi-Crystalline-Solar-PV-Modules.pdf)

[adanisolar.com/-/media/Project/AdaniSolar/Media/Downloads/Downloads/Datasheets/Multi-Encore/Multi-Crystalline-Solar-PV-Modules.pdf](https://www.adanisolar.com/-/media/Project/AdaniSolar/Media/Downloads/Downloads/Datasheets/Multi-Encore/Multi-Crystalline-Solar-PV-Modules.pdf). Accessed 24 Sept 2020