IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 29, 2021, accepted February 19, 2021, date of publication February 23, 2021, date of current version March 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061370

Household-Level Energy Forecasting in Smart Buildings Using a Novel Hybrid Deep Learning Model

DABEERUDDIN SYED[1,2], (Graduate Student Member, IEEE), HAITHAM ABU-RUB[2], (Fellow, IEEE), ALI GHRAYEB[2], (Fellow, IEEE), AND SHADY S. REFAAT[2], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA
[2]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar

Corresponding author: Dabeeruddin Syed (dsyed@tamu.edu)

This publication was made possible by the National Priorities Research Program (NPRP) grant [NPRP10-0101-170082] from the Qatar National Research Fund (a member of Qatar Foundation), and co-funding by IBERDROLA QSTP LLC. The Open Access Funding was provided by Qatar National Library.

ABSTRACT Forecasting of energy consumption in Smart Buildings (SB) and using the extracted information to plan and operate power generation are crucial elements of the Smart Grid (SG) energy management. Prediction of electrical loads and scheduling the generation resources to match the demand enable the utility to mitigate the energy generation cost. Different methodologies have been employed to predict energy consumption at different levels of distribution and transmission systems. In this paper, a novel hybrid deep learning model is proposed to predict energy consumption in smart buildings. The proposed framework consists of two stages, namely, data cleaning, and model building. The data cleaning phase applies pre-processing techniques to the raw data and adds additional features of lag values. In the model-building phase, the hybrid model is trained on the processed data. The hybrid deep learning (DL) model is based on the stacking of fully connected layers, and unidirectional Long Short Term Memory (LSTMs) on bi-directional LSTMs. The proposed model is designed to capture the temporal dependencies of energy consumption on dependent features and to be effective in terms of computational complexity, training time, and forecasting accuracy. The proposed model is evaluated on two benchmark energy consumption datasets yielding superior performance in terms of accuracy when compared with widely used hybrid models such as Convolutional (Conv) Neural Network-LSTM, ConvLSTM, LSTM encoder-decoder model, stacking models, etc. A mean absolute percentage error (MAPE) of 2.00% for case study 1 and a MAPE of 3.71% for case study 2 is obtained for the proposed forecasting DL model in comparison with LSTM-based models that yielded 7.80% MAPE and 5.099% MAPE for two datasets respectively. The proposed model has also been applied for multi-step week-ahead daily forecasting with an improvement of 8.368% and 20.99% in MAPE against the LSTM-based model for the utilized energy consumption datasets respectively.

INDEX TERMS Appliances energy forecasting, bidirectional LSTM, deep learning, hybrid model, power forecasting.

I. INTRODUCTION

Energy consumption in buildings is one of the significant contributors to energy efficiency programs worldwide [1]. Additionally, a major component of the energy consumed in the buildings is wasted through over-utilization of energy appliances such as exhaust fans, and Heating, Ventilation, Air Conditioning (HVAC) systems, ineffective control over thermal comfort, and not optimizing the start-up time and sequencing of electrical equipment. Therefore, it is essential to properly manage energy consumption in buildings which could be done by creating smart buildings (SB) with installed sensors, measuring devices, and various control strategies [2]. One of the management stages to realize

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

demand-side response is to predict the household energy consumption with high accuracy. If the short-term load forecasting at the building level is obtained with high accuracy and precision, the power utilities can utilize the information to manage the operation and maintenance of power systems. This would help the utilities to match the generation of power with the load demand, thereby enabling them to plan and schedule the energy resources.

Energy consumption profile at the building level consists of three components [3]: 1) Regular consumption patterns that can be recognized from the inherent historical load patterns in the building, 2) Uncertain consumption patterns owing to the different daily weather conditions, and 3) Noise component that cannot be physically modeled.

Short-term load forecasting at aggregated levels has a majority component corresponding to regular consumption patterns and therefore, it is easier to predict with a high degree of accuracy. Household-level energy consumption is volatile and uncertain owing to the unpredictable patterns due to weather conditions. Furthermore, customer consumption behavior may change based on various reasons, including the weather. Hence, the consumption depends on individual consumer behavior, and it is too stochastic to predict easily. Therefore, the challenge in short-term energy forecasting at the building level is to predict the uncertain patterns considering the weather conditions and stochastic nature of consumer consumption behavior.

Considering that the energy consumption in buildings is highly volatile and unpredictable, deep learning techniques are forerunners in developing highly accurate prediction models. Extensive research has been performed in the last decade to forecast load at aggregated levels [4]–[6]. In [4], density-based spatial clustering of applications with noise (DBSCAN) algorithm was utilized to aggregate different subzones into clusters based on the historical yearly energy consumption values and the forecasting models are developed at an aggregated level of cluster. Considering the variation of consumption behavior between households, the households were clustered, the aggregate forecasts for each cluster were determined separately, and finally, the forecasts were aggregated in [5]. The aggregation of the residential load was proposed and the percentage of users within a cluster that should be equipped with smart meters and submetering capability was determined in [6]. However, the work on short-term energy forecasting at the household consumer level is still limited. Energy forecasting and management methods for household-level consumption include time series analysis, metaheuristic algorithms such as binary backtracking search algorithm in home energy management systems (HEMS), machine learning approaches such as support vector machines, deep learning methods such as neural networks, and ensemble deep learning models [7]–[16]. Among deep learning models, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been extensively used in energy forecasting problems. CNN is a feed-forward neural network that bases its calculations on convolutions. RNN is a recurrent neural network in time with cells whose internal states depict the dynamic behavior of the dependent feature with time. Long Short Term Memory (LSTM) is a type of RNN and it has three gates to decide which information to move forward and which information to discard. It has been found that LSTM models are more accurate than RNN models. The CNNs have been employed in hybrid models in the first phases to get accurate performance without feature extraction phase [16]. However, the complexity of CNN models depends on multiple factors including convolution process, kernel number, and large memory access. Whereas LSTM networks are based on space and time and the input size does not exponentially grow network requirements. After considering the above factors, the bi-directional LSTM layer is used to extract information from features instead of complex CNN layer, and stacking of LSTM layers on dense layers is used to forecast the dependent variable i.e. energy consumption in our work. Also, the proposed model is evaluated in terms of accuracy and training time performance against the widely used hybrid models including the LSTM encoder-decoder model, CNN-LSTM, ConvLSTM, and another ensemble model.

The key contributions of this work include the following:

1) A novel hybrid deep learning model is built using stacked bi-directional and uni-directional LSTM models allowing the learning of exceedingly non-linear and convoluted patterns and correlations in data that are beyond the reach of classical uni-directional architectures.
2) Two real case studies are depicted to demonstrate the accuracy of the proposed model in forecasting the appliances' energy consumption at the household level in buildings with sensors.
3) Quantitative analyses are performed through score metrics and the effectiveness of the proposed model is demonstrated in comparison with the existing state-of-the-art approaches.
4) The proposed model architecture, effects of model parameters and hyperparameters, and impression of adding lag energy features are meticulously studied. Furthermore, the proposed model is compared with other deep learning models on two different datasets.

The remainder of the paper is structured as follows. Section II describes the literature review on bi-directional LSTMs and machine learning models utilized in household-level energy forecasting. Section III discusses the architecture of the proposed component deep learning layers. Section IV proposes a short-term load forecasting methodology utilizing the proposed Hybrid Stacked Bi-directional Uni-directional Fully connected (HSBUFC) model architecture. In Section V, the performance of the proposed methodology is evaluated against the benchmark and traditional deep learning models. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

Short-term load forecasting at the household level is currently of much research interest with the integration of renewable energy resources in smart grids at the distribution level [17]. The volatile load demand patterns at the consumer level, if predicted with high accuracy, will help in load balancing and renewable energy efficient utilization. Initial works on short-term load forecasting at the household level included the time series analyses and traditional statistical approaches. With the research interest shifting to artificial intelligence, various machine learning approaches and deep learning techniques have been utilized to forecast the energy consumption at the household level. In [18], Moradzadeh *et al.* proposed multilayer perceptron (MLP) and support vector regression (SVR) models trained on buildings' structural characteristics and technical parameters data for forecasting cooling and heating loads in residential buildings. They reported a very high correlation coefficient value of 0.9993 using their proposed models. In [15], Wang *et al.* proposed a two-stage forecasting methodology. In the first stage, the traditional time forecasting models were utilized to perform a day ahead load forecasts. To enhance the accuracy of the forecasts, the second stage utilized models such as support vector machines (SVM), linear regression, and quadratic models to generate predictions of deviations. These deviations were integrated with the forecasts from the first stage to yield the overall forecast values with an average Mean Absolute Percentage Error (MAPE) of 5.21 %. However, the SVM model is not suitable for big data as the training time for the SVM model scales super linearly with the increase in data records.

In [19], Rafiei *et al.* proposed wavelet pre-processing, improved wavelet neural networks, and generalized extreme learning machines (ELMs) on training data. The predictions of the load were provided as intervals keeping in mind the uncertainties of the forecasting models and data noise. ELMs are neural networks with a single hidden layer. The usual disadvantages of ELMs are that the forecasting accuracy is heavily dependent on the activation function, and the generalization is poor. These shortcomings were effectively tackled by the introduction of wavelets as the activation functions in their methodology. However, the ELM-based methods do not effectively perform deep extraction of inherent information and features associated with energy consumption data owing to their single layer-based modeling.

In [20], the authors established mathematical models of backpropagation neural networks and Elman neural network. These models were used with small learning rates, and layers to store internal states, and to deal with time-varying characteristics of energy consumption data. Their results concluded that Elman neural networks perform better in dynamic load forecasting than backpropagation neural networks. However, these neural network-based models are bound to converge to local minima rather than global minima. This leads to poor generalization and further, causes overfitting.

Recently, a lot of research attention has been focused on the development of deep learning models to recognize patterns in the energy consumption data and to perform the forecasts with high accuracy and efficiency. Typically, deep learning models suffer from the problem of exploding gradients (i.e. learning diverges) or vanishing gradients (i.e. the learning stops). This problem is taken care of by LSTM networks that introduce memory cells and computing gates. LSTMs are types of RNNs that have been utilized in the past for time-series analyses and load forecasting problems. In our recent works, multiple efficient and accurate energy consumption forecasting models were developed based on ensemble models, extreme learning machines, LSTMs, deep neural networks, and dimensionality reduction techniques [14], [21], [22].

In [23], the authors developed hybrid sequential learning based on the deep learning model. Their solution utilizes CNN in the first phase to extract the features from the energy consumption dataset and uses Gated Recurrent Unit (GRU) in the second phase to utilize its effective gated structure to make predictions. However, GRU-based models do not have as great volatility as LSTM-based models owing to their simplicity and a smaller number of gates for the gradient flow.

In [11], the authors proposed an advanced domain fusion methodology based on CNN, which derived the time-domain and frequency-domain features representing the changing energy consumption trends, LSTM layers, and Discrete Wavelet Transforms (DWT). The authors reported a MAPE of around 1% on two datasets, which comprise energy consumption (MW) information at aggregated levels. However, this methodology was not tested on the disaggregated level of household consumption or appliance energy use.

In [12], Kong *et al.* proposed an LSTM memory-based framework for short-term energy forecasting at the residential level. They incorporated the appliances energy data from a Canadian household to illustrate the efficacy of their deep learning framework. Although minutely data were available, an aggregation of thirty minutes has been utilized in their work. However, only six appliances' energy data were utilized in the study. Their results were compared against the benchmarking models of Feed Forward Neural Networks (FFNN) and k-Nearest Neighbors (k-NN). The superior performance of the LSTM-based model, with a MAPE of 21.99% was displayed. In the current work, the bi-directional LSTM-based hybrid model was utilized to improve the accuracy of forecasts.

To combine the strengths of different models and their knowledge representations, many studies have been made on hybrid models to perform accurate energy forecasting. In [24], four learning algorithms i.e. k-NN regressor, SVR, XGBoost, and genetic algorithm (GA) were combined to propose an ensemble model for forecasting electricity consumption at the distribution transformer level. Forecasting day-ahead photovoltaic power has been performed in [25] using LSTM and auto-encoder persistence model while handling uncertainties and predictions for complex weather variables. Air conditioning energy prediction has been performed in [26] using a meta ensemble machine learning model based on stacked auto-encoders to optimize prosumer energy management. A hybrid ensemble strategy involving bagging, boosting, and random subspace with pruning on LSTM-based models has been suggested in [27] to extract the features from multi-feature data on industrial power load and a new loss function was proposed to balance the tradeoff between bias-variance and to reduce peak load prediction error. In [28], deterministic and probabilistic low-voltage load forecasting was performed using a hybrid ensemble deep learning model based on deep belief networks. In the same work, the regression ability of the networks was improved by utilizing bagging and boosting techniques and the weights of sub-models in the ensemble were optimally determined by using the k-nearest neighbor method. A comparative study of a novel hybrid artificial intelligence (AI) and deep learning (DL) model for power distribution networks was performed in [29], where AI techniques such as optimally pruned extreme learning machines (OP-ELM), adaptive neuro-fuzzy inference system (ANFIS) and deep learning techniques such as LSTM were examined.

Several recent works were based on LSTM-based models and these models were developed using past data. However, there are other invariants of LSTMs that consider not only the past inputs but also the future context values [30]. The output of distinct hidden layers in either direction is passed through connections to the same layer, and this is the concept of bidirectional LSTMs that are utilized in the proposed work. The advantages of bidirectional LSTMs are 1) to exploit the data features to extract the bidirectional temporal dependencies from available data, and 2) to preserve the information from past and future inputs in the hidden states of LSTM cells.

The proposed model in this paper outperforms [31] in energy consumption forecasting. The same dataset is utilized in the current paper and reference work [31]. In [31], the authors developed data-driven models based on gradient boosting machines, random forests, support vector machines with radial basis function, and multiple linear regression. The authors reported the lowest MAPE of 13.43% for their model predictions using random forests. Although various deep learning or hybrid models are developed to forecast energy consumption at the household level, the error rate is high. Hence, in this work, a hybrid model (HSBUFC) is developed based on the stacking of bi-directional and uni-directional LSTMs followed by fully connected dense layers to achieve the forecasts with high accuracy and low error percentages as compared to the mentioned models.

III. ARCHITECTURE OF DEEP LEARNING MODELS

In this section, the architectures of the uni-directional and bidirectional LSTMs are discussed. The forward and backward passes in bi-directional LSTMs are utilized to recognize inherent patterns in energy consumption data using past and future inputs.

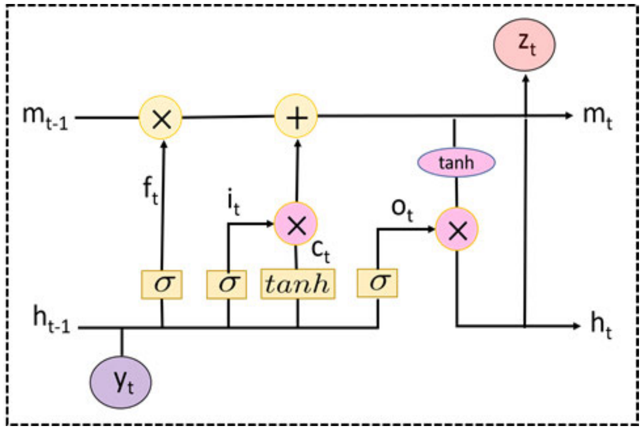


FIGURE 1. Overview of LSTM cell.

A. UNI-DIRECTIONAL LSTMS (OR LSTMS)

LSTM is a type of Recurrent Neural Network that is classically built to process, analyze, and forecast sequence data [32]. The RNN model predicts based on the input of the current time step and the output from the previous time step. In addition to the ability to utilize information from the recurrent connections to the outputs of previous time steps, LSTMs also have memory cells to accumulate steps over prediction sequences enabling them to perform better with long-term dependency tasks such as energy forecasting. Also, the existence of gates and more complex recurrent units in LSTMs enable them to control the information that is passed through and overcome the problem of vanishing gradients [33].

The LSTM unit can be defined as a collection of vectors in R^d at every time step t. The overview of the LSTM cell is illustrated in Figure 1. LSTM vectors are presented as follows [34].

Memory cell (m_t): It is given by the following:

$$m_t = f_t \cdot m_{t-1} + i_t \cdot c_t \tag{1}$$

where c_t is given by

$$c_t = Tanh(W_m \cdot [h_{t-1}, y_t] + b_m) \tag{2}$$

where t symbolizes the current time step, $t-1$ symbolizes the previous time step, W_m represents the weight matrix for memory cell neurons, h_{t-1} represents the hidden state at step $t-1$, y_t represents input at step t, and b_m represents bias for memory cell units.

Input gate (i_t): It is given by the following:

$$i_t = \sigma(W_i \cdot [h_{t-1}, y_t] + b_i) \tag{3}$$

Forget gate (f_t): a gate that resets the old memory.

$$f_t = \sigma(W_f \cdot [h_{t-1}, y_t] + b_f) \tag{4}$$

Output (o_t): Output of LSTM unit and output valve.

$$o_t = \sigma(W_o \cdot [h_{t-1}, y_t] + b_o) \tag{5}$$

Hidden cell state (h_t): h_t is given as follows.

$$h_t = o_t * Tanh(m_t) \tag{6}$$

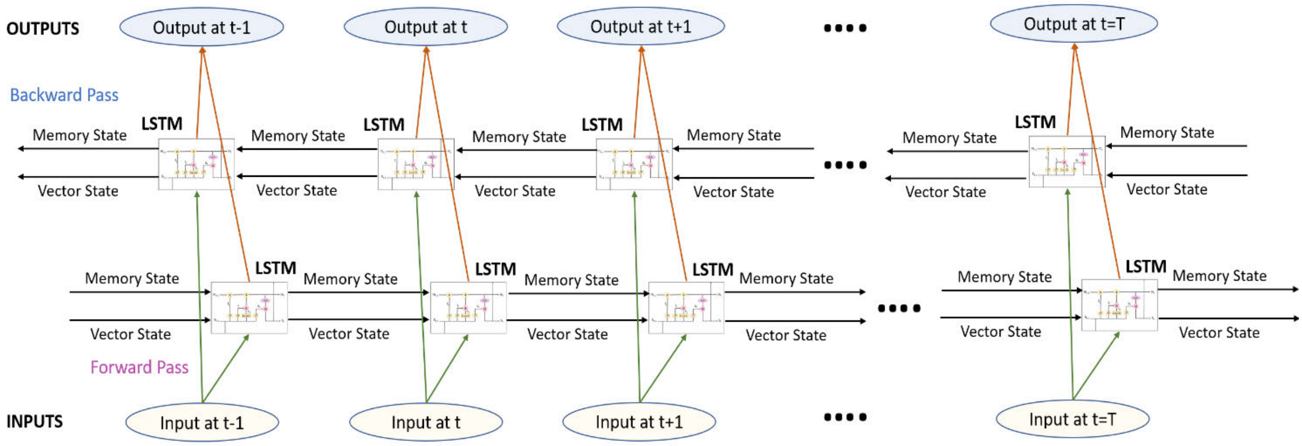


FIGURE 2. The architecture of unfolded bi-directional LSTM model.

The LSTM cell using its structure (as depicted in Figure 1) processes information in four steps as follows. The cell forgets the irrelevant parts of the previous states, stores the relevant parts of new information into its cell state, updates the relevant information from prior and current input into its internal cell states, and yields outputs as inputs for the next time step.

B. BI-DIRECTIONAL LSTMS

Bi-directional LSTM is a development over uni-directional LSTM models. The bi-directional LSTMs process the inputs in two directions - in the forward pass, from past inputs to future inputs, and in the backward pass, from future inputs to past inputs. The combination of hidden states from the forwarding pass and backward pass preserves the information from both past inputs and future inputs through two different hidden layers. The output from these hidden layers is passed to the single identical output layer. This allows the bidirectional LSTMs to preserve the context and data patterns better from both past and future inputs without delay. It has been proven that bi-directional LSTMs perform better predictions and classifications than uni-directional LSTMs in diverse fields such as speech recognition [35]. However, the advantages of bi-directional LSTMs have not been much explored in the field of energy consumption forecasting in smart grids.

The architecture of the unfolded bi-directional LSTM model, comprising of forwarding LSTM units and backward LSTM units, is depicted in Figure 2. The forward pass output ($\vec{h}$) is successfully determined using inputs in the positive sequence of time from $T-k$ to $T-1$. Whereas the backward pass output ($\overleftarrow{h}$) is successfully determined using inputs in the negative sequence of time from $T+k$ to $T+1$. There are no hidden-to-hidden layer connections between the forward LSTM units and the backward LSTM units. The calculations of outputs of the forward pass and backward pass utilize the traditional LSTM functions, i.e. equations (1) to (6). The final output vector of the bi-directional LSTM layer is represented as $Z_T = [z_{T-k}, z_{T-k+1}, \ldots, z_{T-1}]$. Each element in the final output vector is given by the following:

$$z_t = \sigma(\vec{h}_t, \overleftarrow{h}_t) \tag{7}$$

where σ represents a function used to integrate the outputs from forwarding pass and backward pass. The σ function can be a summation, averaging, concatenating, or multiplication function. And the forward pass output ($\vec{h}$) and backward pass output ($\overleftarrow{h}$) are given as follows.

$$\vec{h} = \mathrm{H}\left(W_{y\vec{h}}\mathrm{y}_t + W_{\vec{h}\vec{h}}\vec{h}_{t+1} + b_{\vec{h}}\right) \tag{8}$$

$$\overleftarrow{h} = \mathrm{H}\left(W_{y\overleftarrow{h}}\mathrm{y}_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t+1} + b_{\overleftarrow{h}}\right) \tag{9}$$

where H represents the hidden layer function, y_t represents the input sequence.

IV. PROPOSED HYBRID METHODOLOGY

In this section, the proposed methodology that performs high accuracy predictions of energy consumption at the household level is presented.

A. DATA ACQUISITION

1) UCI APPLIANCES ENERGY DATASET

To build and evaluate the proposed hybrid model, a real-world dataset of appliances energy consumption is used. The dataset is available online and was collected in a smart building fitted with different temperature and humidity sensors [36]. The different features in data used in the current analyses with a proposed hybrid deep learning model are described in Table 1. Apart from data on temperature and humidity in different rooms, appliances energy use, light fixtures energy use, weather data from the nearby weather station, outdoor temperature, outdoor relative humidity, atmospheric pressure, wind speed, visibility, and dewpoint temperature data are available.

The data were collected for a period of five months ranging from 11 January 2016 to 27 May 2016. The frequency of

TABLE 1. Description of features in the UCI appliances energy dataset.

S.No.	Features	Units	Feature description
1	Date	mm-dd-yy hh:mm	Date & Time of energy use
2	Appliances Energy Use	Wh	Energy consumed by all the appliances in the house
3	Light Fixtures Energy Use	Wh	Energy consumed by the light fixtures in the house
4-12	Indoor Room Temperature	Celsius	Room temperature is recorded for nine rooms in the house including but not limited to the laundry room, bedroom, kitchen room, etc.
13-21	Indoor Relative Humidity	%	The relative humidity is recorded for nine rooms in the house including but not limited to the laundry room, bedroom, kitchen room, etc.
22	Outdoor Room Temperature	Celsius	The outdoor external temperature is collected from the nearest weather station.
23	Outdoor Relative Humidity	%	Outdoor external relative humidity recorded.
24	Pressure	mm Hg	Atmospheric pressure is collected from the nearest weather station.
25	Windspeed	m/s	Speed of the wind as collected from the nearest weather station.
26	Visibility	km	Visibility as collected from the nearest weather station.
27	Dewpoint	Â°C	Dewpoint temperature as collected from the nearest weather station.
28-48	Lag values	Wh	Past values of appliances energy used as features in the dataset.

TABLE 2. Descriptive statistics of data features in UCI appliances energy dataset.

Features	Min (Wh)	Max (Wh)	Mean (Wh)	Std. deviation (Wh)
Appliances	10.00	1080.00	97.6949	102.5248
Lights	0.00	70.00	3.8018	7.9359
Indoor Temperature Features (4-12)	-6.0650	29.8567	19.3817	4.9952
Indoor Relative Humidity (13-21)	1.00	99.90	42.7094	12.8473
Outdoor Temperature	-5.00	26.10	7.4125	5.3184
Outdoor Relative Humidity	24.00	100.00	79.7504	14.9010
Atmospheric Pressure	729.30	772.30	755.5226	7.3994
Wind speed	0.00	14.00	4.0397	2.4512
Visibility	1.00	66.00	38.3308	11.7947
Dewpoint temperature	-6.60	15.50	3.7609	4.1952
Rv1	0.0053	49.9965	24.9880	14.4966
Rv2	0.0053	49.9965	24.9880	14.4966

data was 10 minutes. For the forecasting, the next time step temperature and humidity values are utilized in regression models with an assumption that these values are obtainable. Also, weather forecasting is exceptionally accurate in current times. The descriptive statistics of data features are mentioned in Table 2. The standard deviation, mean, maximum, and minimum values of different attributes are tabulated. The table indicates that the appliances' energy consumption has a high standard deviation corroborating the fact that the energy consumption at the household/appliances level is highly volatile. Figure 3 depicts the data distribution of the target feature that is appliance energy use. It can be observed that the distribution has a long tail indicating the high variance of data.

There is a strong correlation between the appliance energy use and weather features such as outdoor relative humidity, outdoor temperature, visibility, wind speed, etc. Also, these weather features have irregular distribution. This is depicted in the distribution plot of the various weather features, as shown in Figure 4.

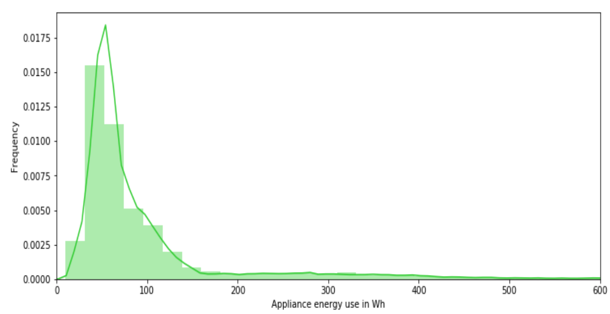


FIGURE 3. Distribution plot of appliance energy use.

2) UCI HOUSEHOLD ENERGY DATASET

An individual household electric energy consumption dataset is utilized to evaluate the performance of the proposed model. The dataset is available in the dataset archive of the University of California, Irvine (UCI) Machine Learning repository [37]. The data contain 2075259 records for 9 attributes. The measurements are collected over a period of 4 years with a frequency of one minute from December

TABLE 3. Data Attributes in UCI household energy dataset.

S.No.	Attributes	Units	Attribute details
1	Date	dd/mm/yyyy	Date of energy consumption
2	Time	hh:mm:ss	Time of energy consumption
3	Global Active Power	kW	household global minute-averaged active power
4	Global Reactive Power	kW	household global minute-averaged reactive power
5	Voltage	V	minute-averaged voltage
6	Global Current Intensity	Amps.	household global minute-averaged current intensity.
7	Sub Metering 1	Wh of active energy	energy sub-metering corresponding to the kitchen, containing mainly a dishwasher, an oven and a microwave.
8	Sub Metering 2	Wh of active energy	energy sub-metering corresponding to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light.
9	Sub Metering 3	Wh of active energy	energy sub-metering corresponding to an electric water-heater and an air-conditioner.
10-74	Lag values	kW	Past values of global active power used as features in the dataset.

TABLE 4. Descriptive statistics of data features in UCI household energy dataset.

Features	Min (kW)	Max (kW)	Mean (kW)	Std. deviation (kW)
Global Active Power	0.076000	11.122000	1.091615	1.057294
Global Reactive Power	0.000000	1.390000	0.123714	0.112722
Voltage	223.200000	254.150000	240.839858	3.239987
Global Current Intensity	0.200000	48.400000	4.627759	4.444396
Sub Metering 1	0.000000	88.000000	1.121923	6.153031
Sub Metering 2	0.000000	80.000000	1.298520	5.822026
Sub Metering 3	0.000000	31.000000	6.458447	8.437154

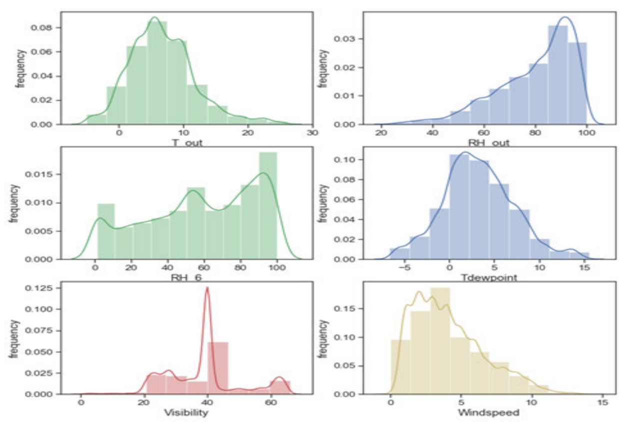


FIGURE 4. Distribution plot of features with irregular distribution.

2006 to November 2010. The different attributes in the energy consumption data are described in Table 3. Electrical quantities such as minute-average voltage values, minute-average current values, active power, reactive power, sub-metering values are available. There are missing values for 1.25% of the measurement records. Imputation methods have been utilized to deal with the missing values [38]. The descriptive statistics of data features are mentioned in Table 4. The standard deviation, mean, maximum, and minimum values of different attributes are tabulated.

The collected data were scaled using a minimum-maximum scaler, as per Equation 10 [39], before passing the data to the deep learning model. The range of the scaled values was set to [0, 1] inclusive of both the extreme values.

$$\hat{x}_m^j = \frac{x_m^j - x_{m,min}}{x_{m,max} - x_{m,min}} \tag{10}$$

where $\hat{x}_m^j$ and x_m^j correspond to the scaled value and actual value for feature 'm' at time step 'j', respectively. And, $x_{m,max}$ and $x_{m,min}$ correspond to the maximum and minimum of actual value for feature 'm' respectively.

The data in case study 1 has 19,736 records and data from case study 2 has 2075259 records. We have selected case study 2 with a larger dataset to prove the scalability of the proposed model and that the proposed model is computationally competent to the commonly employed deep learning hybrid models. Furthermore, additional features that represent the past values of energy consumption/power were added to the dataset for modeling. These past values are referred to as lag values.

In this work, the short-term energy forecasting problem is formulated (with a frequency of 1 to 10 minutes) as follows. Consider that the recorded energy consumption or power values are given by the time series E up to time steps 'j'. E is represented as:

$$E = e_1, e_2, e_3, \ldots.., e_t, \ldots\ldots, e_j, \quad 1 \leq t \leq j \tag{11}$$

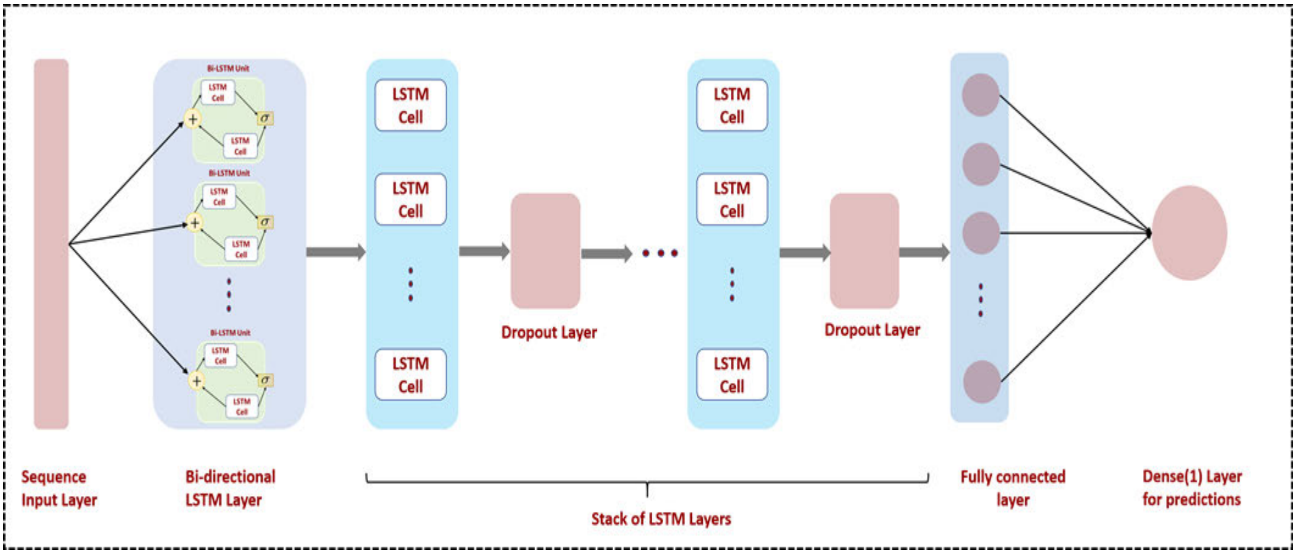


FIGURE 5. The framework of proposed hybrid stacked model.

where e_t represents the t^{th} energy consumption for time step t. Each time step is of 'k' minutes duration (k = 10 in case study 1 and k = 1 in case study 2) with the maximum value of time steps equal to j. The energy consumption problem is to forecast energy consumption value for the next time step $t+1$ using the previous 'l' time step energy consumption values and different indoor and outdoor environment features using the machine learning model 'f' which is represented as:

$$e_{t+1} = f\left(e_{1+t-l}, e_{2+t-l}, e_{3+t-l}, \ldots.., e_t, M^{t+1}\right) \quad (12)$$

where M^{t+1} represents the values of the set of features (that represent the indoor and outdoor environment variables) at time step $t+1$.

The number of lag features was set to 21 in case study 1 and set to 65 in case study 2 after the optimization of the parameter. The optimization is performed using the grid search algorithm, and the metric used for optimization is the root mean square error of predictions.

B. PROPOSED MODEL ARCHITECTURE

The framework of the proposed hybrid stacked bi-directional uni-directional LSTM with fully connected dense layers (HSBUFC) model is illustrated in Figure 5. HSBUFC model consists of three types of layers: 1) Bidirectional LSTM layer, 2) Stacked Uni-directional LSTM layers, and 3) Fully connected layers/dense layers. As discussed in the earlier section, bi-directional LSTMs make use of both forward and backward dependencies. The temporal long-term dependencies of the energy consumption values are extracted during the feature learning process in two directions by the initial layer of bi-directional LSTM. Next, LSTM layers, which are efficient in the forward dependencies, are employed in the top layers, which receive the outputs from the lower layer after learning from the extracted comprehensive and complex features. In neural network architectures, one of the most powerful ways of regularization and avoiding overfitting is the mechanism of dropout [40]. Dropout refers to eliminating a percentage of neuron units i.e. removing the ingoing and outgoing connections from the neuron units resulting in a diminished network. Dropout layers are adopted by the hybrid model within the stack of uni-directional LSTM layers to prevent overfitting. Also, the use of early stopping yields better model generalization by aiding avoidance of overfitting. Finally, the fully connected dense layers are employed to learn the representations extracted till the previous layer and the final dense layer makes predictions of energy consumption for successive future time steps. The effectiveness of the bi-directional LSTM layer and stacking of uni-directional LSTM layers is that the amalgam model can learn long-term dependencies and model implicit representation concealed in the sequential data. In energy consumption or load forecasting application, the historical consumption data are available at once. On that account, there is no rationale not to exploit future and history dependencies together at any time point while training machine learning models.

Different parameters, model parameters, and hyperparameters such as activation function in the hidden layer, batch size, dropout, learning rate, network weight initialization, number of neurons in the hidden layer, number of layers in stacking, optimization algorithm, training epochs, etc. are optimized. Mean absolute error is used as a loss function and Adam solver is utilized as a gradient descent optimization algorithm. Adam solver was selected after investigating it against Adadelta, Adagrad, Adamax, RMSProp, and stochastic gradient descent (SGD). Randomized search cross-validation is exploited to perform a search for optimum parameter values from a defined dictionary of parameters and the possible range of values. Batch size is referred to the number of training points that are utilized in one iteration

TABLE 5. Parameter settings for deep learning layers in the proposed model.

Parameter settings	Value
Batch size	128
Epoch	{50, 100}
Activation	{tanh, linear, ReLU}
Dropout after LSTM layer	25%
Loss function	Mean absolute error
Model optimizer	Adam

of training. Recurrent networks such as CNN, LSTM, etc. are very sensitive to the batch size and hence, crucial to be optimized in our case. Furthermore, there are merits and demerits in adopting the low value to batch size. Low batch size value requires less memory and classically the networks train faster with mini-batches. Nevertheless, the smaller the batch size gets, the lesser accurate the gradient estimate will be. To optimize the number of epochs, the training was performed initially by maintaining a large number for the maximum number of epochs and early stopping with the patience of 10 epochs. This method yielded a model that does not overfit and provided an approximate range for the number of epochs to start with. A few of the optimized parameter settings that are obtained after randomized search cross-validation (CV) strategy for the stacked deep learning layers of the proposed model are mentioned in Table 5.

V. EXPERIMENTAL RESULTS

In this section, several experiments were performed with a goal to obtain high accuracy energy forecasting in smart buildings using the proposed HSBUFC model and to compare the performance of the proposed model with that of other baseline models and widely employed hybrid deep learning models. The baseline models include linear regression, ELM, neural networks, uni-directional LSTM, stacked LSTM models, and bi-directional LSTM model. The commonly deployed hybrid models include LSTM encoder-decoder model, CNN-LSTM, ConvLSTM, and other ensemble models.

The experimental results have been obtained after the execution on a supercomputer with the specifications of Nodes: 1, Cores: 8, Intel Central Processing Unit (CPU) with clock rate: 2.4 GHz, Random Access Memory (RAM): 256 GB, Network speed: 40-100 Gpbs, and programming environment: python. 80% of the data was used for training, and 20% of the data was used as a testing dataset. The train and test datasets were generated using a resampling procedure of k-fold cross-validation. The value of k was chosen to be 5 as it has been empirically proven in the literature to yield better model generalization and to avoid high model bias and variance [41]. To consider the temporal dependencies

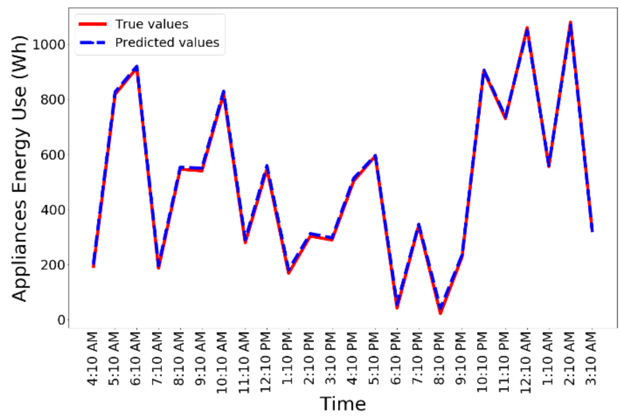


FIGURE 6. Case study 1 actual v/s predicted values for appliances energy use.

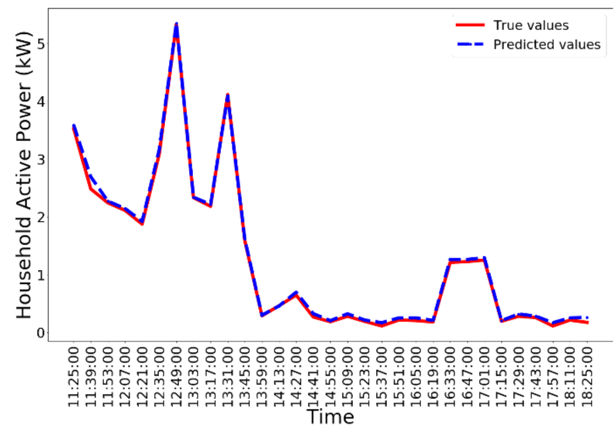


FIGURE 7. Case study 2 actual v/s predicted values for global active power.

of energy consumption on the DateTime feature, lag values of energy consumption are added as additional features to the dataset before k-fold cross-validation is performed. Model settings are created using the deep learning package of Keras (2.4.3) and the open-source software library TensorFlow (2.2.0) is employed as a backend. The proposed model architecture is built using the Keras Functional Application Programming Interface (API).

The actual and predicted values using the proposed model for appliances energy use on a random day in the testing dataset are plotted in Figure 6. The independent axis in the figure indicates the time of the day and the dependent axis depicts the appliance energy consumption. It can be observed from the figure that the predicted values follow closely with the actual energy consumption values. This shows the high accuracy and low error of the proposed hybrid deep learning prediction model. Similarly, the real and predicted values using the proposed forecasting model for global active power on a subset of records from case study 2 are demonstrated in Figure 7. The proposed model delivers substantially perfect accuracy performance at both spikes and troughs in the illustration.

TABLE 6. Proposed model performance.

Training Model	UCI appliances energy dataset (case study 1)				UCI household energy dataset (case study 2)			
	RMSE (Wh)	MAPE (%)	MAE (Wh)	R^2 score	RMSE (W)	MAPE (%)	MAE (W)	R^2 score
Linear Regression	137.27	70.3098	85.5421	0.10300	70.6600	9.661	62.97	0.99520
Extreme Learning Machine	90.109	65.8758	53.4437	0.16300	51.4800	7.647	38.76	0.99700
Neural Nets (3 hidden layers)	86.263	57.9317	48.9346	0.23200	51.0200	6.906	32.88	0.99700
Preprocessing + LSTM	17.543	7.80154	13.5716	0.99730	50.2100	5.099	36.89	0.99729
Preprocessing + Stacked LSTM (2 layers)	15.968	5.82827	12.8362	0.99675	47.4500	4.673	32.41	0.99783
Preprocessing + Stacked LSTM (3 layers)	20.837	12.2823	17.2399	0.99541	47.7700	4.443	27.69	0.99781
AREM [14]	14.385	5.27108	11.6723	0.99740	42.1570	4.088	28.35	0.99785
LSTM Encoder Decoder model	6.3320	2.55666	4.36822	0.99957	37.7616	4.948	27.72	0.99863
CNN-LSTM model	19.741	11.1782	14.9775	0.99588	37.1890	4.410	24.27	0.94800
ConvLSTM model	7.4780	2.67748	5.55012	0.99941	39.1423	6.493	24.25	0.93935
Preprocessing + Bi-directional LSTM	12.712	3.17975	10.1398	0.99829	39.3297	4.674	25.29	0.99851
Proposed Model	**5.4430**	**2.00027**	**3.45383**	**0.99968**	**29.2107**	**3.710**	**22.248**	**0.99867**

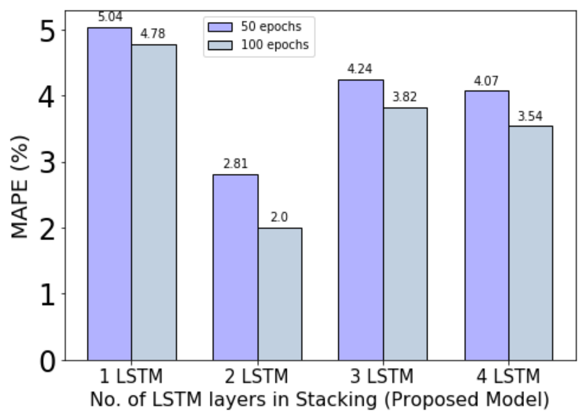


FIGURE 8. Case study 1 MAPE for different number of LSTM layers in stacking.

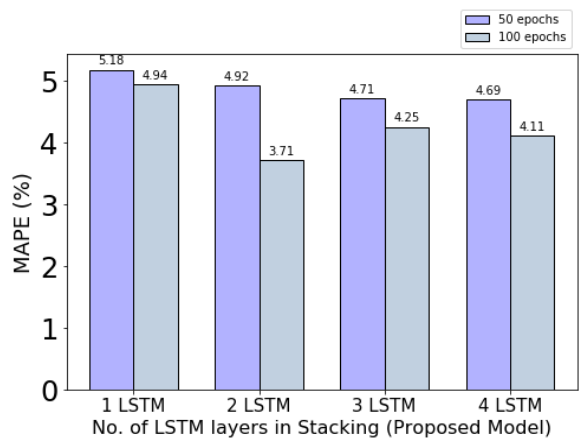


FIGURE 9. Case study 2 MAPE for different number of LSTM layers in stacking.

Further, the experiments were performed by increasing the number of LSTM layers that are stacked on top of the bi-directional LSTM layer. The MAPE results of such experiments on case study 1 and case study 2 are illustrated in Figures 8 and 9. The independent axes in the figures depict the number of LSTM layers within the stacking of the proposed model and the dependent axes depict the MAPE error value in percentage. As shown in Figures 8 and 9, the error of predictions decreases when the number of layers is increased from 1 to 2. Whereas, when the number of LSTM layers in the proposed HSBUFC model is increased beyond 2, the complexity of the model extends beyond the point where the model overfits. Therefore, the error of predictions increases. The same set of experiments were performed using the epoch value of 100, and the trend of a drop of prediction error for 2 LSTM layers in the stacking is corroborated in both cases.

The performance results of different machine learning and deep learning models are depicted in Table 6. The performance metrics used for comparison are Root Mean Square Error (RMSE), MAPE, Mean Absolute Error (MAE), and R2 score/coefficient of determination. These widely employed metrics are given by the following parametric Equations (13) to (16) [22].

$$RMSE = \sqrt{\frac{1}{n} * \sum_{j=1}^{n} (A_j - P_j)^2} \tag{13}$$

$$MAPE = \frac{100}{n} * \sum_{j=1}^{n} \left| \frac{A_j - P_j}{A_j} \right| \tag{14}$$

$$MAE = \frac{\sum_{j=1}^{n} |A_j - P_j|}{n} \tag{15}$$

$$R^2 - Score = 1 - \frac{\sum_{j=1}^{n} (A_j - P_j)^2}{\sum_{j=1}^{n} (A_j - \text{mean}(A))^2} \tag{16}$$

TABLE 7. Proposed model time performance on UCI appliances energy dataset.

Training Model	UCI appliances energy dataset		UCI household energy dataset	
	Fit time (s)	Testing time (s)	Fit time (s)	Testing time (s)
Linear Regression	0.008976	0.000997	1.037000	0.005970
Extreme Learning Machines	0.015818	0.002219	31.63213	13.77380
NN (3 hidden layers)	45.00497	0.024792	54.59372	19.11836
LSTM	18.55806	0.680055	88.95819	33.46719
Stacked LSTM (2 layers)	30.88267	1.167051	134.3528	45.47228
Stacked LSTM (3 layers)	44.53208	1.748603	188.1297	56.87649
AREM [14]	52.84724	1.720164	192.3247	57.15006
LSTM Encoder Decoder model	55.38721	0.289620	206.2960	69.87476
CNN-LSTM model	187.3263	2.116030	621.4900	173.6303
ConvLSTM model	87.48251	0.850860	222.5634	0.228140
Bi-directional LSTM	20.54402	0.983213	200.0609	47.23239
Proposed Model	36.40035	1.517856	423.8497	74.02498

Here, A_j corresponds to actual energy consumption value at time step j, P_j corresponds to predicted energy consumption value at time step j, and n is the total number of time steps. The performance metrics MAPE and R^2-score provide scale-independent measures. Hence, the short-term energy forecasting performance can be evaluated by comparing MAPE values or R^2-score.

As shown in Table 6, in case study 1, linear regression attained an RMSE of 137.274 Wh and MAPE of 70.30%, while ELM achieved 90.109 Wh RMSE and 65.875% MAPE, and Neural Networks with 3 hidden layers attained 86.263 Wh RMSE and 57.931% MAPE. The improvement in model accuracy was observed after the inclusion of lag values of appliance energy use, data preprocessing and optimization of the number of lag features to be included. With preprocessing, LSTM attained 17.543 RMSE and 7.801% MAPE while Stacked LSTM (2 layers) achieved 15.968 RMSE and 5.828% MAPE and stacked LSTM (3 layers) attained 20.83 Wh RMSE and 12.28% MAPE. Results of LSTM models with different stacking depths showed that increasing the complexity does not necessarily enhance the model's accuracy. Our proposed model has an RMSE of 5.443 Wh and MAPE of 2.00% for appliances energy use prediction, while the single-layer bi-directional LSTM model, after preprocessing, has the RMSE value of 12.71 Wh and MAPE of 3.1797%. The better performance of the proposed model against the different invariants of unidirectional LSTM layers indicates that the prediction of current values benefits from the information in the forward and backward pass (effect of bidirectionality) existent in bi-directional LSTMs. Also, the proposed model has been evaluated in comparison to the most widely used hybrid models such as the LSTM Encoder-Decoder model, CNN-LSTM hybrid, and ConvLSTM model with superior results for our proposed hybrid model. And, we have included the proposed Averaging Regression Ensemble Model (AREM) ensemble model from our previous work in performance evaluation.

In case study 2, linear regression attained an RMSE of 70.66 W and MAPE of 9.661%, while ELM achieved 51.48 W RMSE and 7.647% MAPE and Neural Networks with 3 hidden layers attained 51.02 W RMSE and 6.906% MAPE. Our proposed model attained 29.21 W RMSE and the lowest MAPE of 3.7% whilst the LSTM layer based model achieved 50.21 W RMSE and MAPE of 5.099%. The better performance of the proposed model in both the case studies indicates that the forecasting model benefits from the two-direction information and training present in bi-directional LSTMs. The best results in case studies 1 and 2 are highlighted in Table 6, and the best performing values correspond to the proposed hybrid model involving bi-directional LSTM, multiple layers of LSTM, and fully connected layers.

Table 7 presents the time cost of different training models developed for the energy use and global active power predictions task in case study 1 and case study 2 respectively. The time cost is provided for fitting/training times, and testing times. The simple single-layered models of LSTM and Bidirectional LSTM have low fit times among all the models. It is evident from Table 7 that the fitting time is directly dependent on the number of layers in a model. The proposed model consisting of two layers of LSTM, one layer of bi-directional LSTM, and two layers of fully connected networks have a fit time of 36.4 secs in case study 1 and 423.8 secs which are highly competent to the other models in the table. As a tradeoff between the accuracy and fitting time, it can be concluded that the proposed model is the best model out of the tested models for the application requirements.

TABLE 8. Results of multi-step week-ahead daily forecasting.

Training Model	UCI appliances energy dataset			UCI household energy dataset		
	RMSE (kWh)	MAPE (%)	MAE (kWh)	RMSE (kW)	MAPE (%)	MAE (kW)
Linear Regression	11.31660	55.540	10.15917	1510.31	59.97	1524.60
Extreme Learning Machine	8.227196	39.308	5.464549	598.973	46.98	330.320
Neural Nets (3 hidden layers)	8.144832	38.572	5.426364	598.694	46.90	330.320
LSTM	7.532940	39.676	5.582667	663.082	41.81	647.206
Stacked LSTM (2 layers)	7.419587	37.920	5.334749	452.088	25.70	431.305
Stacked LSTM (3 layers)	7.810895	40.136	5.646407	645.140	28.80	519.824
AREM [14]	7.278559	37.313	5.075018	419.035	24.32	401.763
LSTM Encoder Decoder model	8.033106	36.308	5.107945	395.306	40.60	319.353
CNN-LSTM model	7.825592	35.288	4.964348	393.194	37.51	298.060
ConvLSTM model	8.674451	39.609	5.572306	405.804	37.60	318.887
Bi-directional LSTM	8.227171	32.389	4.804583	393.246	27.23	294.561
Proposed Model	**7.093705**	**31.308**	**4.404549**	**391.137**	**20.82**	**290.42**

Short-term energy forecasting ranges to a maximum of two weeks. Hence, to demonstrate the accuracy of the proposed model for multi-step ahead short-term energy forecasting, we have performed a week-ahead average daily energy forecasting for subsequent weeks. The testing errors for the week-ahead daily energy forecasting are provided in Table 8. The same widely used hybrid models, ensemble AREM model, and different invariants of LSTM have been investigated for multi-step forecasting on datasets from case study 1 and case study 2. The performance comparison is presented in Table 8. As shown in Table 8. the proposed model provides the least error and the highest accuracy even in the case of multi-step ahead forecasting. In case study 1, the average daily MAPE is improved by 4.032% utilizing the proposed hybrid deep learning model when compared to the LSTM-based model. In case study 2, the average daily MAPE for the LSTM-based model is 27.2% and the least error is for the proposed hybrid deep learning model with an average MAPE value of 20.828%.

VI. CONCLUSION

Improving the accuracy of energy consumption forecasting at the level of buildings will hugely impact the generation and scheduling of energy resources and efficient utilization of renewable energy resources. This paper proposed a novel hybrid deep learning model that amplifies the merits of unidirectional LSTMs, bidirectional LSTMs, and stacking of RNNs on energy consumption forecasting accuracy. Bidirectional LSTMs are used to recognize the underlying energy consumption patterns in both directions and forecast energy consumption values with high accuracy. The obtained accuracy is irrespective of the high uncertainty and stochasticity of individual household load demand. Superior accuracy performance was demonstrated using two real energy consumption datasets in individual residential smart buildings. Dropout regularization and early stopping were employed to prevent overfitting in the proposed hybrid deep learning model. The proposed model has been evaluated against the widely employed hybrid models including CNN-LSTM, ConvLSTM, LSTM encoder-decoder model, other invariants of stacked LSTM models, and our proposed ensemble AREM model from previous work. The superior performance of our proposed model in both case studies and in multi-step forecasting corroborates the merits of employing bidirectional training. The proposed model is not limited to short-term energy forecasting. The work can be extended to medium-level forecasting and long-term forecasting of energy consumption. Future work will focus on including additional contributing factors such as household occupancy data. The scalability of the proposed model will be evaluated on big data. In the future, we will explore to speed up the model training time to facilitate real-time big data analytics for energy forecasting applications. Future research calls for parallelization of bidirectional LSTMs to achieve distributed computing and training of energy forecasting models.

REFERENCES

[1] A. Vishwanath, V. Chandan, and K. Saurav, "An IoT-based data driven precooling solution for electricity cost savings in commercial buildings," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7337–7347, Oct. 2019, doi: 10.1109/JIOT.2019.2897988.

[2] M. S. Ahmed, A. Mohamed, H. Shareef, R. Z. Homod, and J. A. Ali, "Artificial neural network based controller for home energy management considering demand response events," in *Proc. Int. Conf. Adv. Electr., Electron. Syst. Eng. (ICAEES)*, Nov. 2016, pp. 506–509, doi: 10.1109/ICAEES.2016.7888097.

[3] R. Z. Homod, H. Togun, H. J. Abd, and K. S. M. Sahari, "A novel hybrid modelling structure fabricated by using takagi-sugeno fuzzy to forecast HVAC systems energy demand in real-time for basra city," *Sustain. Cities Soc.*, vol. 56, May 2020, Art. no. 102091, doi: 10.1016/j.scs.2020.102091.

[4] Z. Han, M. Cheng, F. Chen, Y. Wang, and Z. Deng, "A spatial load forecasting method based on DBSCAN clustering and NAR neural network," *J. Phys. Conf. Ser.*, vol. 1449, no. 1, p. 12032, 2020, doi: 10.1088/1742-6596/1449/1/012032.

[5] T. Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Nov. 2015, pp. 879–887. [Online]. Available: https://ieeexplore.ieee.org

[6] J. Ponocko and J. V. Milanovic, "Forecasting demand flexibility of aggregated residential load using smart meter data," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5446–5455, Sep. 2018, doi: 10.1109/TPWRS.2018.2799903.

[7] M. S. Ahmed, A. Mohamed, T. Khatib, H. Shareef, R. Z. Homod, and J. A. Ali, "Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm," *Energy Buildings*, vol. 138, pp. 215–227, Mar. 2017, doi: 10.1016/j.enbuild.2016.12.052.

[8] W. Xu, H. Hu, and W. Yang, "Energy time series forecasting based on empirical mode decomposition and FRBF-AR model," *IEEE Access*, vol. 7, pp. 36540–36548, 2019, doi: 10.1109/ACCESS.2019.2902510.

[9] Z. Tan, J. Zhang, Y. He, Y. Zhang, G. Xiong, and Y. Liu, "Short-term load forecasting based on integration of SVR and stacking," *IEEE Access*, vol. 8, pp. 227719–227728, 2020, doi: 10.1109/access.2020.3041779.

[10] K. Park, J. Jeong, D. Kim, and H. Kim, "Missing-insensitive short-term load forecasting leveraging autoencoder and LSTM," *IEEE Access*, vol. 8, pp. 206039–206048, 2020, doi: 10.1109/access.2020.3036885.

[11] X. Shao, C. Pu, Y. Zhang, and C. S. Kim, "Domain fusion CNN-LSTM for short-term power consumption forecasting," *IEEE Access*, vol. 8, pp. 188352–188362, 2020, doi: 10.1109/access.2020.3031958.

[12] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018, doi: 10.1109/TPWRS.2017.2688178.

[13] Z. Kong, C. Zhang, H. Lv, F. Xiong, and Z. Fu, "Multimodal feature extraction and fusion deep neural networks for short-term load forecasting," *IEEE Access*, vol. 8, pp. 185373–185383, 2020, doi: 10.1109/access.2020.3029828.

[14] D. Syed, S. S. Refaat, H. Abu-Rub, O. Bouhali, A. Zainab, and L. Xie, "Averaging ensembles model for forecasting of short-term load in smart grids," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Los Angeles, CA, USA, Dec. 2019, pp. 2931–2938, doi: 10.1109/BigData47090.2019.9006183.

[15] Y. Wang, Q. Xia, and C. Kang, "Secondary forecasting based on deviation analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 500–507, May 2011, doi: 10.1109/TPWRS.2010.2052638.

[16] 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.

[17] D. Syed, A. Zainab, S. S. Refaat, H. Abu-Rub, and O. Bouhali, "Smart grid big data analytics: Survey of technologies, techniques, and applications," *IEEE Access*, early access, Nov. 27, 2020, doi: 10.1109/access.2020.3041178.

[18] A. Moradzadeh, A. Mansour-Saatloo, B. Mohammadi-Ivatloo, and A. Anvari-Moghaddam, "Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings," *Appl. Sci.*, vol. 10, no. 11, p. 3829, May 2020, doi: 10.3390/app10113829.

[19] M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-Khah, and J. P. S. Catalao, "Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6961–6971, Nov. 2018, doi: 10.1109/TSG.2018.2807845.

[20] X. Zheng, X. Ran, and M. Cai, "Short-term load forecasting of power system based on neural network intelligent algorithm," *IEEE Access*, early access, Sep. 2, 2020, doi: 10.1109/access.2020.3021064.

[21] D. Syed, S. S. Refaat, H. Abu-Rub, and O. Bouhali, "Short-term power forecasting model based on dimensionality reduction and deep learning techniques for smart grid," in *Proc. IEEE Kansas Power Energy Conf. (KPEC)*, Jul. 2020, pp. 1–6, doi: 10.1109/KPEC47870.2020.9167560.

[22] D. Syed, S. S. Refaat, and H. Abu-Rub, "Performance evaluation of distributed machine learning for load forecasting in smart grids," in *Proc. Cybern. Informat. (K&I)*, Jan. 2020, pp. 1–6, doi: 10.1109/KI48306.2020.9039797.

[23] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020, doi: 10.1109/ACCESS.2020.3009537.

[24] P. W. Khan and Y.-C. Byun, "Genetic algorithm based optimized feature engineering and hybrid machine learning for effective energy consumption prediction," *IEEE Access*, vol. 8, pp. 196274–196286, Oct. 2020, doi: 10.1109/access.2020.3034101.

[25] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, "Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model," *IEEE Trans. Ind. Appl.*, vol. 56, no. 6, pp. 7185–7192, Nov. 2020, doi: 10.1109/TIA.2020.3025742.

[26] Y. Chen, G. Fu, and X. Liu, "Air-conditioning load forecasting for prosumer based on meta ensemble learning," *IEEE Access*, vol. 8, pp. 123673–123682, 2020, doi: 10.1109/ACCESS.2020.2994119.

[27] M. Tan, S. Yuan, S. Li, Y. Su, H. Li, and F. H. He, "Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2937–2948, Jul. 2020, doi: 10.1109/TPWRS.2019.2963109.

[28] Z. Cao, C. Wan, Z. Zhang, F. Li, and Y. Song, "Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 1881–1897, May 2020, doi: 10.1109/TPWRS.2019.2946701.

[29] S. Motepe, A. N. Hasan, and R. Stopforth, "Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms," *IEEE Access*, vol. 7, pp. 82584–82598, 2019, doi: 10.1109/ACCESS.2019.2923796.

[30] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

[31] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy Buildings*, vol. 140, pp. 81–97, Apr. 2017, doi: 10.1016/j.enbuild.2017.01.083.

[32] A. Zainab and D. Syed, "Deployment of deep learning models on resource-deficient devices for object detection," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 73–78, doi: 10.1109/ICIoT48696.2020.9089651.

[33] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "Performance evaluation of deep recurrent neural networks architectures: Application to PV power forecasting," in *Proc. 2nd Int. Conf. Smart Grid Renew. Energy (SGRE)*, Nov. 2019, pp. 1–6, doi: 10.1109/SGRE46976.2019.9020965.

[34] N. Somu, G. Raman M R, and K. Ramamritham, "A deep learning framework for building energy consumption forecast," *Renew. Sustain. Energy Rev.*, vol. 137, Mar. 2021, Art. no. 110591, doi: 10.1016/j.rser.2020.110591.

[35] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schluter, and H. Ney, "A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2462–2466, doi: 10.1109/ICASSP.2017.7952599.

[36] *UCI Machine Learning Repository: Appliances Energy Prediction Data Set*. Accessed: Jan. 22, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction

[37] *UCI Machine Learning Repository: Individual Household Electric Power Consumption Data Set*. Accessed: Jan. 22, 2021. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption

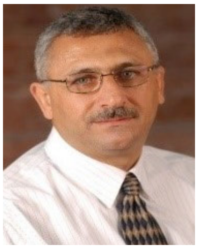
[38] J. Sessa and D. Syed, "Techniques to deal with missing data," in *Proc. 5th Int. Conf. Electron. Devices, Syst. Appl. (ICEDSA)*, Sarawak, Malaysia, Dec. 2016, pp. 1–4, doi: 10.1109/ICEDSA.2016.7818486.

[39] S. V. Khond, "Effect of data normalization on accuracy and error of fault classification for an electrical distribution system," *Smart Sci.*, vol. 8, no. 3, pp. 117–124, Jul. 2020, doi: 10.1080/23080477.2020.1799135.

[40] P. Baldi and P. Sadowski, "Understanding dropout," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 2814–2822, doi: 10.17744/mehc.25.2.xhyreggxdcd0q4ny.

[41] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?" *Comput. Statist.*, pp. 1–23, Jun. 2020, doi: 10.1007/s00180-020-00999-9.

DABEERUDDIN SYED (Graduate Student Member, IEEE) received the B.E. degree in electrical and electronics engineering (EEE) from the University College of Engineering, Osmania University, Hyderabad, India, and the M.Sc. degree in data science and engineering (DSE) from Hamad Bin Khalifa University, Doha, Qatar. He is currently pursuing the Ph.D. degree in electrical engineering (EE) with Texas A&M University (TAMU), College Station, TX, USA. He has a work experience as an Oracle DBA with Oracle Pvt. Ltd., as a Teaching Assistant in courses of machine learning (learning from data), electrical circuits, and python programming. His research interests include big data analytics, machine learning, deep learning, cybersecurity, and smart grids.



HAITHAM ABU-RUB (Fellow, IEEE) received the dual Ph.D. degrees. He is currently a Full Professor. He attained extensive teaching and research experience in diverse nations of USA, Germany, Poland, Palestine, and Qatar. He has been working with Texas A&M University at Qatar (TAMU-Q), Doha, Qatar, for the past 17 years. He was a Chair of the Electrical and Computer Engineering (ECEN) Program with TAMU-Q for five years. He is currently working as the Managing Director (MD) with the Smart Grid Center, Extension in Qatar (SGC-Q). He has published more than 400 research articles in prestigious journals and conferences and has five books, and six book chapters to his credit. His primary research interests include power electronic converters, smart grid, renewable energy, and electric drives. He received multiple recognitions, international, and national awards. He was a recipient of the German Alexander von Humboldt Fellowship and the American Fulbright Scholarship

ALI GHRAYEB (Fellow, IEEE) received the Ph.D. degree in electrical engineering (EE) from The University of Arizona, USA. He has worked as a Professor with the Department of Electrical and Computer Engineering (ECE), Concordia University, Canada. He is currently working as a Professor with the Department of Electrical and Computer Engineering (ECE), TAMU-Q, Qatar. His principal research interests include massive MIMO, wireless communications, physical layer security, and visible light communications. He has served on the Editorial Board for several IEEE and non-IEEE journals.

SHADY S. REFAAT (Senior Member, IEEE) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering (EE) from Cairo University, Egypt, in 2002, 2007, and 2013, respectively. He attained industrial work experience of over 12 years. He served as a Senior Electrical Engineer, a Team Leader, and a Design Engineer for several engineering projects. He is currently working as an Associate Research Scientist with TAMU-Q. He has published over 105 research journal and conference articles. His main research interests include on big data, condition monitoring, electrical machines, energy management systems, fault detection, reliability of power systems, fault-tolerant systems, and smart grid. He is also serving as a member for the SGC-Q and the Institution of Engineering and Technology (IET).