IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 16, 2021, accepted June 23, 2021, date of publication June 28, 2021, date of current version July 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3093053

Short-Term Energy Forecasting Framework Using an Ensemble Deep Learning Approach

MUSTAQEEM, MUHAMMAD ISHAQ, AND SOONIL KWON
National Interaction Technology Laboratory, Sejong University, Seoul 05006, Republic of Korea

Corresponding author: Soonil Kwon (skwon@sejong.edu)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019M3F2A1073179).

ABSTRACT Industrial and building sectors demand efficient smart energy strategies, techniques of optimization, and efficient management for reducing global energy consumption due to the increasing world population. Nowadays, various artificial intelligence (AI) based methods are utilized to perform optimal energy forecasting, different simulation tools, and engineering methods to predict future demand based on historical data. Nevertheless, nonlinear energy demand modeling is still unfledged for a better solution to handle short-term and long-term dependencies and avoid static nature because it is purely on historical data-driven. In this paper, we propose an ensemble deep learning-based approach to predict and forecast energy demand and consumption by using chronological dependencies. Our system initially processes the data, cleaning, normalization, and transformation to ensure the model performance. Furthermore, the preprocess data is fed to proposed ensemble model to extract hybrid discriminative features by using convolution neural network (CNN), stacked, and bi-directional long-short term memory (LSTM) architectures. We trained our proposed system on the historical data to forecast the energy demand and consumption with a different time interval. In the proposed technique, we utilized the concept of active learning based on moving windows to ensure and improve the prediction performance of the system. The proposed system could be applicable to employ energy consumption in industrial and building sectors to demonstrate their significance and effectiveness. We evaluated the proposed system by using benchmark, residential UCI, and local Korean commercial building datasets. We conducted different extensive experimentation to show the error rate and used various kinds of evaluation matrices, which indicate the lower error rate of the proposed system.

INDEX TERMS Energy analysis, electricity demand forecast, convolution neural network, deep learning, LSTM network, smart sensor system.

I. INTRODUCTION

Population growth, advancement in technologies, and socioeconomics is a greater extent, which risen a demand from the past few decades for the consumption of energy and material. Nowadays, throughout the world, the level of energy consumption has been increasing due to the vast level of population growth. The electricity demand has increased seven percent each year [1], which illustrates the detailed statistical analysis. The increase in demand requires a sufficient amount of energy to fulfill demands and satisfy the customer while keeping care of the industry. The utility corporations are responsible for improving their services continuously by facilitating better plans to maintain better energy consumption. In this domain, there are various methods to statistically and traditionally analyze and characterize energy patterns to forecast consumption and demand [2]. The recent existence techniques are broadly categorized into two main chunks [3] such as artificial intelligence and convolution neural network (CNN) methods. These convoluted methods use regression and stochastic time-series-based techniques to predict energy consumption. The stochastic methods are used for time-series data to extend the series patterns in future prediction [4]. These techniques are widely used in existing methods for linear problem solving with better and efficient outcomes. Various machine learning (ML) techniques such as a decision tree, the Bayesian, ensemble, and neural networks have become very popular with emergence of artificial intelligence (AI) [5], [6].

We were inspired by the performance of AI, which has mainly three components such as learning, validation, and testing, and researchers have utilized it in various methods

The associate editor coordinating the review of this manuscript and approving it for publication was Ehab Elsayed Elattar.

due to their capturing ability in non-linear data [6], [7]. In the learning phase, we train the AI system and then validate to ensure their performance in the validation phase and generate the mapping or coordination among the input and output variables. After, in the testing phase, we utilized a developed or trained model for energy forecasting and demand. In this technological area, some researchers were motivated by the intention toward machine learning methods such as support vector machine due to their strong theoretical background in empirical forecasting models [8]. Hence, each approach e.g. traditional and non-traditional (machine learning and artificial intelligence) have their own advantages and disadvantages. However, nowadays AI-based methods are the most popular due to their high-performance outcomes and reliability [9]. These AI-based methods such as CNN, recurrent neural network (RNN), multi-layer perceptron (MLP), and ensemble methods have been vastly used for time-series and energy forecasting problems [10]. The MLP network provides good outcomes and shows better capability as compared to traditional methods but it is not capable of historical dependencies and long-term sequence handling in time-series data. Due to these issues, the attention of the research community has diverted to CNN and RNN methods. However, these methods also failed due to long-term consequences and vanishing gradient problems [11]. Furthermore, recently there have been many techniques which were proposed by researchers to handle, predict, and model the long and short-term dependency for energy forecasting and demand at household level and region-wise [12], [13].

Due to the above-mentioned issues and facts, we propose a novel and universal framework to handle short-term energy forecasting by using an ensemble deep learning approach to overcome the above issues. Our system is capable of solving the described problems in the existing methods such as dependency handling, real-time forecasting, modeling energy demands at a different level [14]. We propose a hybrid ensemble deep learning model, which can easily handle and model the energy consumption as well as forecasting demands to upturn the future strategies, planning, and performance. The main key-points and contribution of the proposed system as follow:

- **Pre-processing:** Pre-processing is a basic step to refine the data for accurate prediction, remove the outliers, noises, and normalize with a pre-processing technique, which ensure the model performance for better energy forecasting.
- **Proposed Model:** Energy modeling, consumption, and prediction are challenging problems due to their nature and non-linearity. Simple machine learning techniques have been failed due to limitations of handling non-linearity. In this regard, we propose a hybrid deep learning approach to extract high-level spatial cues by CNN layers and handle the non-linear complex behavior, long-term dependencies, and sequences power patterns by stacked and bi-directional LSTM networks. Stacked LSTM computes features by using forward strategies and BiLSTM computes and learns features by a forward and backward stream, and combines all interpretation to generate the output for energy forecasting. In addition, the propose model gives better outcomes with recent observations, which means the model supports active learning and can be adopted for better forecasting in industrial applications.
- **Model Evaluation:** We conducted extensive experimentation using 10-fold and hold-out cross-validation techniques to illustrate the significance and effectiveness of the propose energy forecasting system over existent baseline methods. Our model achieved better results than state-of-the-art techniques with reduced error rates using mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) evaluation matrices. Due to this high performance, our model could be applicable to employ for energy consumption and prediction in industrial and building sectors.

The rest of the article is divided into sub-sections: Section 2 describes and illustrates the recent literature of the energy domain and Section 3 represents the proposed methodology and its main components in detail. Section 4 relates to the discussion, results, and experimentations to show the proposed model prediction performance. Finally, section 5 concludes the study and represents the possible future directions.

II. LITERATURE REVIEW

Energy analysis is the more emerging and significant area of research and development from the last few year, due to country-level impact on the socioeconomic. Various research analytics and studies performed well on consumer dissection, contour depiction, patterns, prediction, and petition by the real-world sensor recorded forecast data [13], [15]. Many researchers developed different techniques in this domain to analyze the concept of smart metering for residential buildings. Nevertheless, in this study, we focus on long-term prediction, consumption, characterization, and active learning system for energy forecasting. In the past era, there are many tools developed for simulation to accurately predict energy utilization, which are broadly classified into three categories, engineering [16], artificial intelligence [17], and hybrid [9] methods. The dynamic relationship among variables discussed in the engineering methods based on specific internal logic and equations [16], which are also called white-box methods. The authors introduced a method in [18] to estimate the demand patterns of the building based on space heating and appliances category by utilizing various factors. Similarly, [19] developed a technique to simplify the physical characteristics of the system by using frequency features analysis. These methods worked well for energy forecasting and demand but their huge cost computation and time complexity make it problematic to generalize these techniques for real-time energy forecast applications [7].

Statistical machine learning methods implement a straightforward linear regression model [20] for modeling the pattern among predictors to deal with internal gains and solar gains [21], [22]. Similarly, [22] proposed a linear regression method to analyze the energy consumption by using conditional demand analysis. The approach shows good outcomes in the prediction analysis, but the flexibility is so poor due to the required large input for modeling with this limitation its flop and unsuitable for further development. Braun *et al.* [23] and Rezaei *et al.* [24] developed a forecasting model for the supermarket by using humidity, temperature, and relative humidity climate variables. In this method, the model generates among these variables and energy demand and predicts the variation in the future forecast for different variables such as fuels, consumption, etc. Guo *et al.* [25] proposed a machine learning method for a non-linear relationship among energy demand and other factors to forecast the energy petition. Similarly, Abedinia *et al.* [26] developed a hybrid approach by utilizing genetic algorithms by utilizing intersection theory and select features by maximum relevancy and minimum redundancy. The selected features were utilized to predict the load and estimate the price of the energy system but the model didn't predict the statistics of the system smoothly due to the non-linearity of the data [27]. Most researcher have utilized the hybrids model industrial applications in different fields and domains such as [28] for faults diagnosis and data transmitting in robotics and healthcare monitoring system [29], [30].

In this current era, various AI methods covered both artificial neural network (ANN) [31], and genetic algorithm because it is the dominant source for building an intelligent system to extract hidden cues from data and show their importance and effectiveness [32], [33]. In [34] the authors developed an ANN model for energy forecasting using multi-layer perceptron (MLP) technique with different factors and the results were encouraging than the simple ML technique. Likewise, Kialashaki and Reisel [35] proposed a multi-linear regression model using the ANN approach and predicted the energy demand for industrial sections, which proved better outcomes in input-output mapping. In [36] the authors introduced a hybrid approach using radial neural network and stochastic for short-term energy forecasting and compared the results of this system with MLP to validate the system. Furthermore, [37] elaborated and extended this approach with daily resident activities predictions and performed a comparative analysis with ML techniques, which show the capability of ANN for solving short-term energy forecasting problem.

Nowadays, deep learning (DL) is a dominant source for collecting discriminative features and providing a convenient outcome due to its deep architectures. He [38] proposed a deep-net strategy to predict and forecast the short-term energy load using CNN architecture. Similarly, the authors presented an RNN based deep learning approach [39] to forecast the short-term energy load with regional and household level aggregations. The RNN approach shows better outcomes than the shallow neural network. Furthermore, Rahman *et al.* [40] developed an RNN based model for a residential and commercial building to forecast the energy demand and compared it with the ANN model to show their effectiveness. Hence, researchers have focused on the sequential model and [41], [42] proposed the energy forecasting systems by utilizing LSTM and GRU networks. The outcomes of these models were better than RNN and other sequential models in time-series and forecasting problems. In [43] the authors presented an ensemble model with Monte-Carlo dropout and improved the generalization capability of the system in forecasting. Similarly, Kong *et al.* [44] proposed a deep-net model to estimate the short-term energy load and demands by using different independent variables. The hybrid approach was applied and identified the energy-saving amount for building by using neural networks and genetic algorithms [45] to minimize the MASE score for real-time energy demands. In [46] the authors proposed a system by using adoptive fuzzy inference, which combined the back-propagation with the least square error method and predicted the energy demand for Canada. Similarly, Duan *et al.* [47] developed a hybrid approach to forecast the building load by utilizing the max-relevance min-redundancy feature with an optimizing SVM classifier. The authors used a swarm algorithm for optimization and produced outperformed results to show the system significance and effectiveness for the energy building loads forecasting.

III. PROPOSED METHODOLOGY

The main framework including components and methodology of the proposed short-term energy forecasting using ensemble deep learning approach is explained in this section. The visual multi-level architecture representation of the system is shown in **Figure 1**. Our proposed system requires historic power consumption data to forecast the power demands for future, daily, monthly, and with different time intervals (depending on user specification). An efficient and refined energy forecasting system is important in the residential building to manage energy utilization professionally and save extra energy power. The accurate forecasting of energy is a challenging task due to noises, missing info, and weather circumstances, which yield debauched forecasting. Initially, the pre-processing module refines the input data and removes noises, biases, and fills the missing holes throughout data, then normalizes the data to ensure the model prediction. Subsequently, the refined data is split into training, validation, and testing fold to efficiently train and evaluate the system. Our proposed hybrid ensemble deep learning approach extract high-level spatial cues by CNN layers and handle the non-linear complex behavior, long-term dependencies, and sequences power patterns analyzed and extract spatiotemporal features by stacked and bi-directional LSTM networks. Stacked LSTM network computes features by using forward strategies and BiLSTM network compute and learns sequential features from a forward and backward stream to check the dependencies from both side and then recognized accordingly. After extracting all spatial and spatiotemporal cues are concatenate and combines the interpretation to generate the output for energy forecasting. Furthermore, the train-

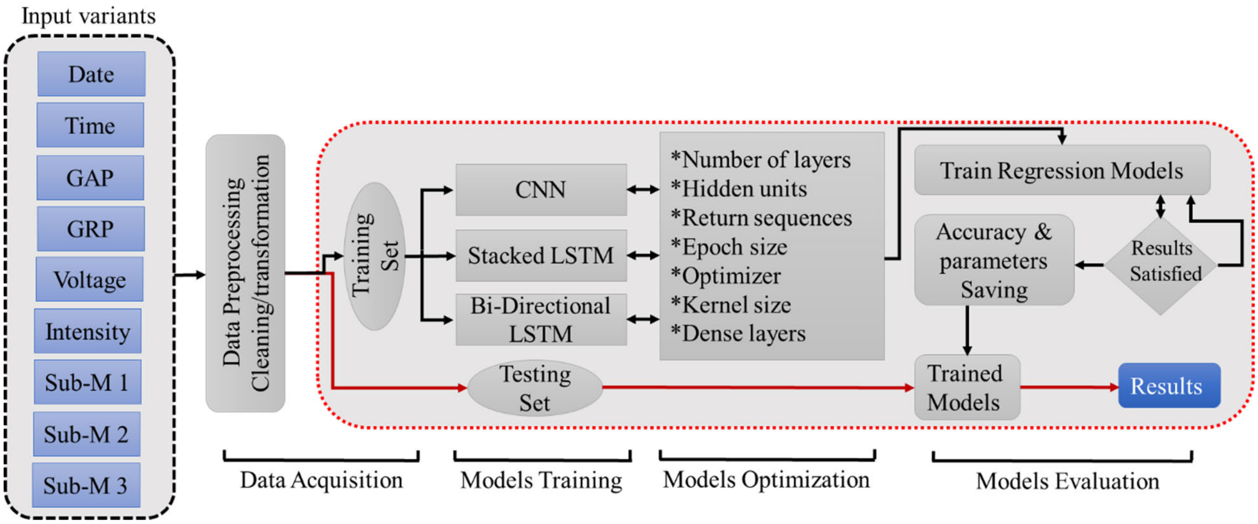

FIGURE 1. Proposed energy forecast system architecture with related comments and modules such as data acquisition, model training, model optimization parameters, and trained model evaluations with detailed input and output.

ing data pass from the proposed ensemble system to model and forecast the energy demands by using CNN, stacked, and bi-directional LSTM networks. The moving window-based strategy is used in feature mapping to actively forecast the current demand and consider real-time observation. The suggested system is implemented to forecast and estimate the power consumption and demand for residential buildings as well as the obtained outcomes compared with existing baseline state-of-the-art regression models.

A. DATA PRE-PROCESSING

The collected UCI energy dataset has various discrepancies [48] for example, missing values, incomplete data, noises, and raw format due to real-time susceptibility. These errors and discrepancies in the un-process data produce confusion and might be an indication of poor data analysis. Hence, the pre-processing step toward data refining is very important for real-world datasets that ensure the performance and reliability of the system to discover knowledge from the real-world data. Usually, the data pre-processing step includes basic sub-step or phases to apply on raw data for refining, which are the following:

- **Cleaning of data**: In this phase, the system refines the data by filling missing values, removing noises, detecting outliers, and determining discrepancies inside the raw data [5]. This data analysis phase ensures the model prediction performance due to utilizing the most refine data.
- **Transformation of data**: In this phase, the system utilizes different techniques for data integration, for example, it integrates multiple files into a single format [5] as well as scaling the data attributes by following specific properties.
- **Reduction of data**: In this phase, the system removes the redundancy from the data and captures the properties as well as provides the reduced representation of refining data either by reducing attributes or by sampling.
- **Discretization of data:** In this phase, the system utilizes the binning technique to reduce the values of a variable by splitting the attribute range into intervals. Discretization follows the hierarchies' concept to refine data and make it suitable for the specific task [5].

Mostly these phases are used in the pre-processing step to refine and analyze data for efficient and accurate prediction.

Therefore, various sub-phases can be used efficiently depending upon the data formatting, approach, and requirement of input for the proposed model.

B. SEQUENCE LEARNING

Sequence learning, recurrent neural networks (RNNs) is one of the dominant sources to find out temporal correlations and performs well to decide the present state-based historical cues in time-series prediction/data [49]. The RNNs units consist of hidden state cells, which contribute to future events by using temporal cues with high outcomes as compared to traditional techniques [50]. RNN is capable of short-term dependencies and stores a lot of info regarding previous cues but the RNNs are not able to model the long-term dependencies, which create complications in training called vanishing gradient [11]. Due to this problem, the RNNs are unable to compute long-term sequences, and result in the researchers developing long-short term memory (LSTM) in 1997 [51]. LSTM is a variant of RNNs that overcome the limitations of RNNs by replacing the hidden layers of RNNs with memory cells to model the long-term dependencies. LSTM has different gates such as input, output, and forget gate along with activation function to model and learn the behavior of time-based relations. The working phenomena of the LSTM network is defined mathematically in **Equations** (1-5) [52]

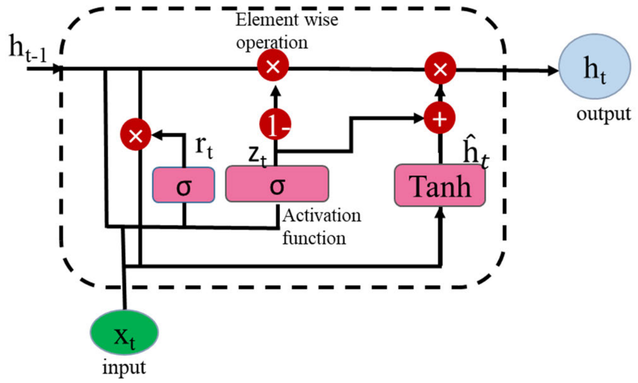

FIGURE 2. Internal structure and mechanism of LSTM cell.

and visually show the internal mechanism in **Figure 2.**

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_t + b_o) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

In the above equations the sigmoid function is represented by σ gates, and the memory cell is illustrated by i, f, c, and o, respectively. The weight matrices of a memory cell to gate unit is represented diagonally by $w_{xi,f,c,o}$ respectively in order to catch the temporal cues. In this regard, we utilized two variants of LSTM in this study to investigate and forecast the energy consumption and future demands, which are further explained in the upcoming sections.

C. STACKED LSTM

Stacked LSTM networks follow the stack architecture by placing the multiple LSTM layers consecutively like multilayers fully connected structure as shown in **Figure 3**. Stacking multiple LSTM layers increases the depth of the model and leads to greater model complexity [15]. In the stacked network the earlier layer outputs should be sequential and must be used as input for the upcoming layer. One of the big pros of a stacked LSTM network is to provide the output for each time-stamp, not a single output for all time-stamps [9]. The mathematical representation of the stacked LSTM network is represented in Equations (6-11) to model the L^{th} layer as listed below.

$$f_t^L = \sigma(w_{fh}^L h_{t-1}^L + w_{fx}^L h_t^{L-1} + b_f^L) \quad (6)$$

$$i_t^L = \sigma(w_{ih}^L h_{t-1}^L + w_{ix}^L h_t^{L-1} + b_i^L) \quad (7)$$

$$\varsigma_t^L = \sigma(w_{\varsigma h}^L h_{t-1}^L + w_{\varsigma x}^L h_t^{L-1} + b_\varsigma^L) \quad (8)$$

$$c_t^L = f_t^L . c_{t-1}^L + i_t^L . \varsigma_t^L \quad (9)$$

$$o_t^L = \sigma(w_{oh}^L h_{t-1}^L + w_{ox}^L h_{t-1}^L + b_o^L) \quad (10)$$

$$h_t^L = o_t^L . \tanh(c_t^L) \quad (11)$$

The above equations show the connection between layers in a stacked network where the output of the L-1^{th} layer is h_t^{L-1} that holds an input for the next L^{th} layer. In the stacked network the interconnection among input-output shows only the relation between two consecutive layers.

D. BI-DIRECTIONAL LSTM

Simple LSTM networks process the information in a single direction as usual and pay no attention to future handled substantial. To overcome the limitation of traditional LSTM Graves and Schmidhuber [53] introduced the bi-directional LSTM network in 2005. The basic concept of bi-directional LSTM is to split the standard LSTM into two states, forward and backward as illustrated in **Figure 3**. The forward state output is not utilized as an input for the backward state and vice-versa. The connections of forward layers follow the similar phenomena of stacked LSTM, which is explained in the previous section. The backward layer's hidden sequences of bi-directional LSTMs are iteratively computed from time t = 1 to time T. The layers of the bi-directional LSTM network can be expressed mathematically at time t [54]–[56], which is shown in **Equations** (12-18).

$$f_t^{\leftarrow L} = \sigma(w_{\leftarrow fh}^L h_{t+1}^L + w_{\leftarrow fx}^L h_t^{L-1} + b_{\leftarrow f}^L) \quad (12)$$

$$i_t^{\leftarrow L} = \sigma(w_{\leftarrow ih}^L h_{t+1}^L + w_{\leftarrow ix}^L h_t^{L+1} + b_{\leftarrow i}^L) \quad (13)$$

$$\varsigma_t^{\leftarrow L} = \sigma(w_{\leftarrow \varsigma h}^L h_{t+1}^L + w_{\leftarrow \varsigma x}^L h_t^{L+1} + b_{\leftarrow \varsigma}^L) \quad (14)$$

$$c_t^{\leftarrow L} = f_t^{\leftarrow L} . c_{t+1}^{\leftarrow L} + i_t^{\leftarrow L} . \varsigma_t^{\leftarrow L} \quad (15)$$

$$o_t^{\leftarrow L} = \sigma(w_{\leftarrow oh}^L h_{t+1}^L + w_{\leftarrow ox}^L h_t^{L-1} + b_{\leftarrow o}^L) \quad (16)$$

$$h_t^{\leftarrow L} = o_t^{\leftarrow L} . tanh(c_t^{\leftarrow L}) \quad (17)$$

The bi-directional LSTM network provides a cumulative output of forward and backward layer such as $h_t^{\leftarrow L}$ and $h_t^{\rightarrow L}$ respectively. So the **Equation 18** represents it accordingly.

$$y_t = w_{\rightarrow hy h_t^{\rightarrow}} + w_{\rightarrow hy h_t^{\leftarrow}} + b_y \quad (18)$$

The bi-directional LSTM network trains the model and upgrades the weights by utilizing a forward and backward pass. In the forward pass, the model runs all inputs by using time $1 \leq t \leq T$ to find the predicted outcomes: the time t = 1 to T accomplish forward pass for forward state and time t = T to 1 accomplish backward pass for output neurons. Therefore, the backward pass finds the derivative, which is utilized in the forward pass, and upgrades the weights accordingly [57]. The visual representation of the stacked and bi-directional LSTM network is illustrated in **Figure 3**."

IV. EXPERIMENTAL EVALUATIONS AND DISCUSSION

In this section of the article, we provide a detailed discussion of the conducted experimentations and evaluate the performance of the proposed system to show the robustness and effectiveness of the system over state-of-the-art methods. We utilized two standard datasets called IHPC [58], and local Korean power/energy consumption. We conducted extensive experimentation using power consumption data to check the model prediction performance for industrial applications and, as a result, we obtained a good outcome. Based on reported

TABLE 1. The detailed description and remarks of each attribute of the household power consumption dataset.

S. No	Attributes	Units	Remarks
01	Date	day/month/year	Date variable shows the day, month, and year as an integer.
02	Time	Hours/minutes/sec	The time range for hour values is 0 to 23, for minute values i1 to 60 and for second as well.
03	Global Active Energy	Kilowatts	The overall average of active power of each minute.
04	Global Reactive Energy	Kilowatts	The overall average of re-active power of each minute.
05	Voltage	Volts	The whole/average voltage of every minute.
06	Global Intensity	Amperes	Overall intensity of current for every minute.
07	Sub-metering 1	Watt-hour	The level of active kitchen energy of kitchen stuff.
08	Sub-metering 2	Watt-hour	The level of active energy for laundry system.
09	Sub-metering 3	Watt-hour	The level of active energy for living room.

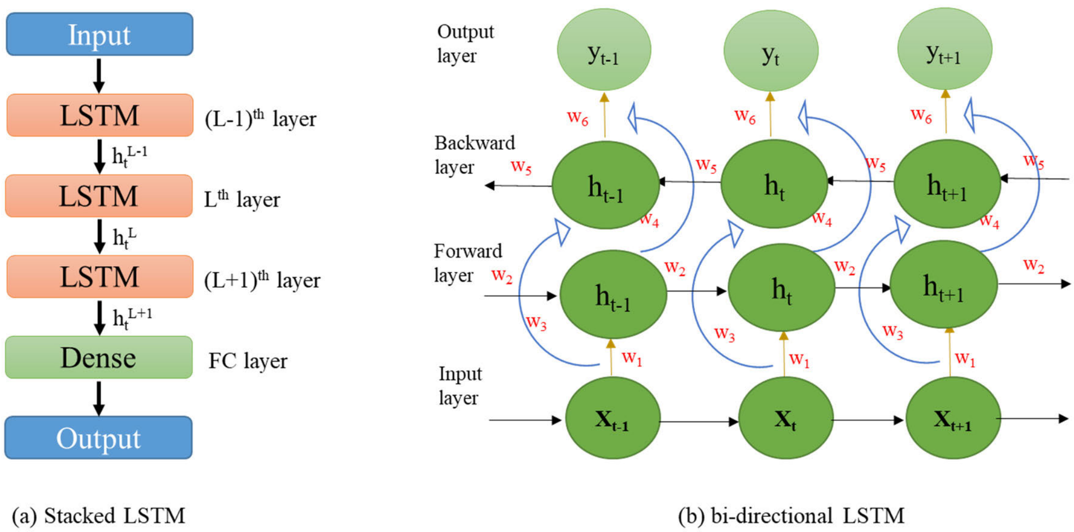

FIGURE 3. Structures of stacked and bi-directional LSTM network.

performance, our system is applicable for industrial applications. The detail about datasets is illustrated in the upcoming next section.

A. INDIVIDUAL HOUSEHOLD POWER CONSUMPTION (IHPC) AND LOCAL KOREAN DATASET

Individual household power consumption is a UCI machine learning repository dataset that is publically available by using free licensing, which contains four-year data of electric power consumption [58]. The dataset contains minute information of consuming power of a four-year period, which consists of more than two million instances on the record. The dataset has some attributes and their measurement units to show the active and re-active energy. The active power shows the sub-metering active consumed energy of each minute in watts. The collection and quantity measurement of the consumed energy is performed by installed sensors. We train our system by using this dataset and test the proposed system for two hours to predict the next two hours prediction of power consumption. The detailed description of the IHPC dataset and its attributes is described in **Table 1** with some useful hints and remarks.

Additionally, we evaluated our system by a local Korean dataset, which is most similar to the UCI dataset. The UCI data is recorded from residential buildings while the Korean data is recorded from commercial buildings. Furthermore, the UCI data have three sub-meters and each sub-meter used individual consumption sensors while the Korean dataset has only one consumption sensor to measure the power consumption. Thus, the UCI dataset has single-minute information and the Korean dataset has fifteen-minute information in a single value. The rest of the attributes and purpose of both datasets is the same. So we evaluated the proposed system by this data as well. We utilize different evaluation metrics to evaluate the system performance, which is explained in the coming section.

B. EVALUATION METRICS

We utilized various assessment metrics such as mean square error (MSE), root means square error (RMSE), mean average percentage error (MAPE) and mean absolute error (MAE) to evaluate the performance of the proposed system. A mathematical representation of all these assessment metrics is expressed in **Equation (19-22)**. The MAE metric reports the

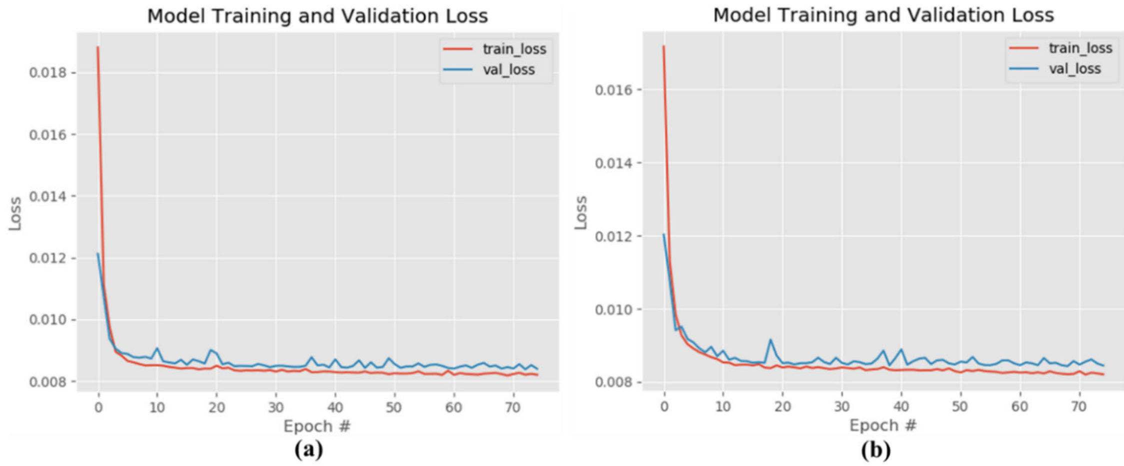

FIGURE 4. The model training and validation losses, (a) represent the UCI residential building dataset and (b) represents the Korean commercial building dataset.

difference of the predicted variables in percentage and variation among predicted and testing variables are represented by RMSE. The testing and predicted variables average square value is represented by MSE and the prediction accuracy in percentage is illustrated by MAPE.

$$MSE = \frac{1}{n}\sum\nolimits_{1}^{n}(y-\hat{y})^2 \tag{19}$$

$$MAE = \frac{1}{n}\sum\nolimits_{1}^{n}|y-\hat{y}| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{1}^{n}(y-\hat{y})} \tag{21}$$

$$MAPE = \frac{100}{n}\sum\nolimits_{t-1}^{n}\left|\frac{A_t-F_t}{A_t}\right| \tag{22}$$

We train our proposed model by using both UCI and Korean data and testing as well. We split the data in 75:25 % ratio for model training and testing. Our model utilized 75 % of data during training and the remaining 25 % is used for testing. The model training and validation loss is mentioned in **Figure 4** for both datasets. We conducted an ablation study to select the best model architecture for the specified task and the numerical results of the conducted experiments are reported in **Table 2**. All these are standard equations and are frequently used for the measurement of time-series and forecasting problems, which analyze the errors among actual and predicted values as well as show the model robustness and effectiveness over provided data for the specific task.

C. MODEL PERFORMANCE EVALUATIONS

We conducted extensive experimentations on both UCI and local Korean power consumption datasets to validate and evaluate the proposed model's robustness and effectiveness for a real-world problem. During the model testing, we validated the system for hourly as well as for daily data and obtained the prediction results for each that show the actual and predicted energy forecasting and consumptions. The visual results of the proposed system over hourly and daily energy consumption and their predictions for hours and day

TABLE 2. Ablation study for best model selection based on the prediction performance using suggested dataset.

Input Features	Architecture	MAE	MSE	RMSE
Time-series data	Regressions model	1.20	1.11	1.25
	CNN model	0.40	0.35	0.80
	LSTM model	0.36	0.33	0.54
	CNN+LSTM model	0.33	0.29	0.47
	CNN+Stacked+BiLSTM	**0.31**	**0.21**	**0.35**

are illustrated in **Figure 5**. The results of the proposed method are closely related to the actual consumed energy level as well as the other native characteristics of energy. Hence, it shows better results than the other approaches as reported in **Table 2**. The proposed model can effectively and easily handle the irregular tendencies of energy consumption. Hence, the time-series data has more complex patterns. Our proposed model handles them well and moderates the error at each interval compared to other model structures as discussed in **Table 2**.

D. DISCUSSION AND COMPARATIVE ANALYSIS

The comparative analysis and discussion of the proposed system are illustrated in this section over suggested datasets with state-of-the-art baseline models using similar of input data. The obtained results of the proposed system are compared with deep learning as well as machine learning techniques using daily and hourly data, which are reported in **Table 3**. We tested our system on hourly energy date and the predicted result of the proposed system is reported in **Table 3** and the daily data results of the system are mentioned in **Figure 6** with SOTA baseline methods. Our proposed model secured better results as compared to other recent methods and recorded the reduced error rate of MSE, MAE, RMSE, and MAPE metrics. We utilized the same evaluation matrices, which are used frequently in contrast to the measurement of energy forecasting. The final outcomes of the baseline

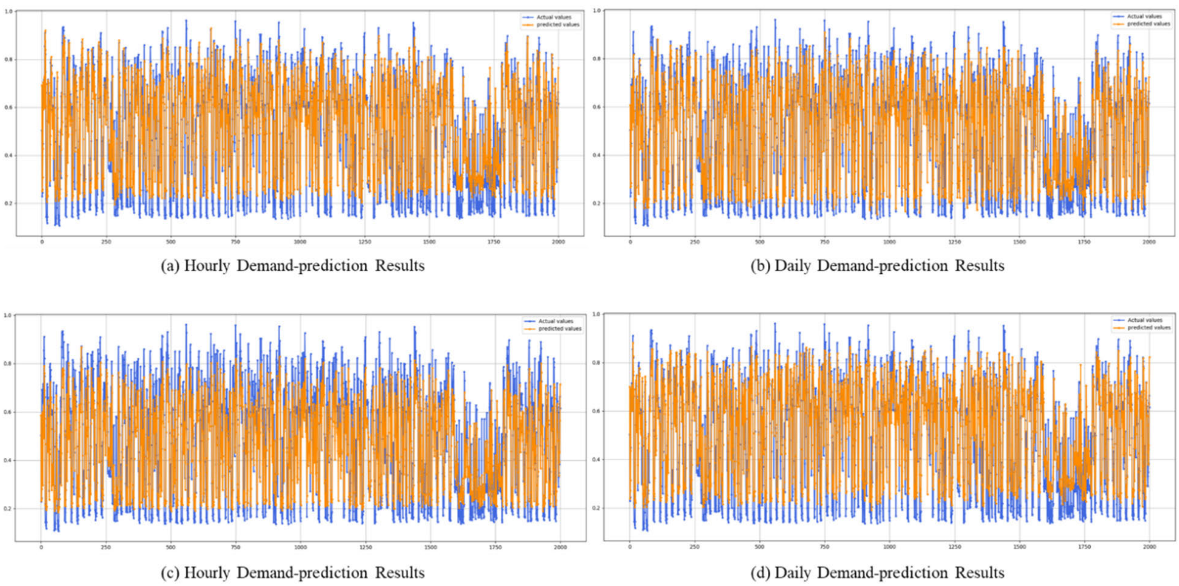

FIGURE 5. The visual representation of average daily and hourly electricity forecasting and consumption outcomes over benchmark UCI residential (Fig, a, b) and local commercial (Fig, c, d) building datasets.

TABLE 3. Statistical comparison of the proposed model with SOTA baseline models using household power consumption dataset using deep learning and machine learning methods.

Approach	Reference	MSE	MAE	RMSE	MAPE
Deep learning	[59]	0.35	0.33	0.59	-
	[60]	0.38	0.39	-	-
	[61]	-	-	0.74	-
	[62]	0.29	0.39	0.54	-
Machine learning	[63]	-	-	0.36	-
	[64]	-	1.12	1.25	-
	[65]	-	-	-	0.82
Proposed	**##**	**0.21**	**0.31**	**0.35**	**0.78**

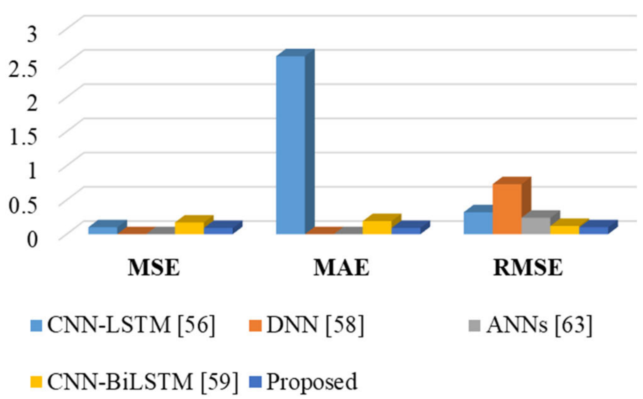

FIGURE 6. The comparison of the proposed system with SOTA baseline methods using daily energy prediction outcomes with reduced error rate using MSE, MAE, and RMSE metrics.

methods and the proposed technique are mentioned and summarized in **Table 3** and **Figure 6**.

Where **Table 3** denoted the error rate of the SOTA baseline methods using standard evaluation matrices and the hyphens (-) representing the missing values, which means the authors didn't use that metric. As far as we know deep learning is a dominant source nowadays due to achieving high performance in many fields. Hence, in this domain, deep learning achieved good outcomes as well, which is shown in the above table. Recent deep learning approaches achieved better results but they are missing some important information in during modeling the energy patterns due to these missing cues that model performance is a bit low. During the literature study we assume this limitation and designed an ensemble model to cover all aspect and easily model the energy pattern and recognize spatial and spatiotemporal cues as well. So cause of the ensemble framework designing is the inspiration of the performance of deep learning in the forecasting of time-series problems, that why's we propose an ensemble deep learning approach and secure good results as compared to other baseline methods. The performance of the proposed system over daily prediction consumption is illustrated in below **Figure 6**.

V. CONCLUSION AND FUTURE DIRECTION

Electricity/power demand and forecasting in an accurate and reliable way are of great significance but, unfortunately, it has gained smaller interest as compared to other domains. However, researchers have done some significant improvements in the recent few years and developed robust, accurate, and efficient forecast models using AI and DL. These methods, AI and DL performed well and effectively handle the non-linear problems in time-series data due to their characteristics and importance. In this paper, we propose an ensemble deep learning-based approach to predict and forecast energy demand and consumption by using chronological dependencies. Our system initially processes the data, cleaning, normalization, and transformation to ensure model performance. Furthermore, the preprocessed data is fed into an ensemble model to extract hybrid discriminative features by using CNN, stacked, and bi-directional LSTM architectures. We evaluated the proposed system by using a benchmark, UCI residential, and local Korean commercial building datasets and compared the achieved results with SOTA methods, which are shown in **Table 3**.

Aimed at future exertion, numerous non-linear exogenous structures of data for example weather/climate circumstances or changes, monetary attributes, and variables can be explored for the inclination investigation of power feasting patterns. Additionally, several optimization methods can be considered and designed to increase the forecast precision of the suggested system.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no acknowledged competing commercial and financial interests or individual affiliations that could affect the work stated in this article.

ACKNOWLEDGMENT

The authors read and agreed with the submitted form of the article.

REFERENCES

[1] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep. 2007.

[2] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: Literature survey and classification of methods," *Int. J. Syst. Sci.*, vol. 33, no. 1, pp. 23–34, Jan. 2002.

[3] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, Aug. 2003.

[4] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[5] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *Morgan Kaufmann Data Manage. Syst.*, vol. 5, no. 4, pp. 83–124, 2011.

[6] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Eds., *Machine Learning: An Artificial Intelligence Approach* (Symbolic Computation). Berlin, Germany: Springer, 1983, doi: 10.1007/978-3-662-12405-5.

[7] Z. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. Baik, "Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework," *Sensors*, vol. 20, no. 5, p. 1399, Mar. 2020.

[8] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 671–682, May 2006.

[9] M. Sajjad, S. U. Khan, N. Khan, I. U. Haq, A. Ullah, M. Y. Lee, and S. W. Baik, "Towards efficient building designing: Heating and cooling load prediction via multi-output model," *Sensors*, vol. 20, no. 22, p. 6419, Nov. 2020.

[10] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, May 2014.

[11] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, Apr. 1998.

[12] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Appl. Energy*, vol. 123, pp. 168–178, Jun. 2014.

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

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

[15] 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.

[16] N. Fumo, "A review on the basics of building energy estimation," *Renew. Sustain. Energy Rev.*, vol. 31, pp. 53–60, Mar. 2014.

[17] R. Kumar, R. K. Aggarwal, and J. D. Sharma, "Energy analysis of a building using artificial neural network: A review," *Energy Buildings*, vol. 65, pp. 352–358, Oct. 2013.

[18] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the U.K," *Energy Buildings*, vol. 37, no. 6, pp. 663–671, Jun. 2005.

[19] A. Boodi, K. Beddiar, Y. Amirat, and M. Benbouzid, "Simplified building thermal model development and parameters evaluation using a stochastic approach," *Energies*, vol. 13, no. 11, p. 2899, Jun. 2020.

[20] T.-T. Ha, V. Feuillet, J. Waeytens, K. Zibouche, S. Thebault, R. Bouchie, V. Le Sant, and L. Ibos, "Benchmark of identification methods for the estimation of building wall thermal resistance using active method: Numerical study for IWI and single-wall structures," *Energy Buildings*, vol. 224, Oct. 2020, Art. no. 110130.

[21] S. Fathi, R. Srinivasan, A. Fenner, and S. Fathi, "Machine learning applications in urban building energy performance forecasting: A systematic review," *Renew. Sustain. Energy Rev.*, vol. 133, Nov. 2020, Art. no. 110287.

[22] T. Alajmi and P. Phelan, "Modeling and forecasting end-use energy consumption for residential buildings in Kuwait using a bottom-up approach," *Energies*, vol. 13, no. 8, p. 1981, Apr. 2020.

[23] M. R. Braun, H. Altan, and S. B. M. Beck, "Using regression analysis to predict the future energy consumption of a supermarket in the U.K.," *Appl. Energy*, vol. 130, pp. 305–313, Oct. 2014.

[24] M. Rezaei, S. K. Chaharsooghi, A. H. Kashan, and R. Babazadeh, "A new approach based on scenario planning and prediction methods for the estimation of gasoil consumption," *Int. J. Environ. Sci. Technol.*, vol. 17, no. 6, pp. 3241–3250, Jun. 2020.

[25] Q. Guo, Y. Feng, X. Sun, and L. Zhang, "Power demand forecasting and application based on SVR," *Procedia Comput. Sci.*, vol. 122, pp. 269–275, Jan. 2017.

[26] O. Abedinia, N. Amjady, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 62–74, Jan. 2017.

[27] O. Abedinia, M. Bagheri, M. S. Naderi, and N. Ghadimi, "A new combinatory approach for wind power forecasting," *IEEE Syst. J.*, vol. 14, no. 3, pp. 4614–4625, Sep. 2020.

[28] J. Long, J. Mou, L. Zhang, S. Zhang, and C. Li, "Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots," *J. Manuf. Syst.*, Sep. 2020.

[29] J. Zhong, D. Wang, J. Guo, D. Cabrera, and C. Li, "Theoretical investigations on kurtosis and entropy and their improvements for system health monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, Oct. 2020.

[30] Mustaqeem and S. Kwon, "Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network," *Int. J. Intell. Syst.*, pp. 1–20, May 2021.

[31] H.-X. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012.

[32] A. K. Dwivedi and S. M. Satapathy, "Mining patterns in open source software using software metrics and neural network models," *Int. J. Syst. Eng.*, vol. 10, no. 4, pp. 397–409, 2020.

[33] D. Guérillot and J. Bruyelle, "Geochemical equilibrium determination using an artificial neural network in compositional reservoir flow simulation," *Comput. Geosci.*, vol. 24, no. 2, pp. 697–707, Apr. 2020.

[34] L. Ekonomou, "Greek long-term energy consumption prediction using artificial neural networks," *Energy*, vol. 35, no. 2, pp. 512–517, Feb. 2010.

[35] A. Kialashaki and J. R. Reisel, "Development and validation of artificial neural network models of the energy demand in the industrial sector of the united states," *Energy*, vol. 76, pp. 749–760, Nov. 2014.

[36] Mustaqeem and S. Kwon, "CLSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network," *Mathematics*, vol. 8, no. 12, p. 2133, Nov. 2020.

[37] K. Gajowniczek and T. Ząbkowski, "Electricity forecasting on the individual household level enhanced based on activity patterns," *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0174098.
[38] W. He, "Load forecasting via deep neural networks," *Procedia Comput. Sci.*, vol. 122, pp. 308–314, Jan. 2017.
[39] H. Shi, M. Xu, Q. Ma, C. Zhang, R. Li, and F. Li, "A whole system assessment of novel deep learning approach on short-term load forecasting," *Energy Procedia*, vol. 142, pp. 2791–2796, Dec. 2017.
[40] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Appl. Energy*, vol. 212, pp. 372–385, Feb. 2018.
[41] J. Lago, F. De Ridder, and B. De Schutter, "Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms," *Appl. Energy*, vol. 221, pp. 386–405, Jul. 2018.
[42] A. Arif, N. Javaid, M. Anwar, A. Naeem, H. Gul, and S. Fareed, "Electricity load and price forecasting using machine learning algorithms in smart grid: A survey," in *Proc. Workshops Int. Conf. Adv. Inf. Netw. Appl.* Islamabad, Pakistan: COMSATS Univ. Islamabad, 2020, pp. 471–483.
[43] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.
[44] 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, Mar. 2017.
[45] M. De Wit, "Heat air and moisture model for building and systems evaluation," *Bouwstenen*, vol. 100, p. 100, 2006.
[46] G. Zahedi, S. Azizi, A. Bahadori, A. Elkamel, and S. R. W. Alwi, "Electricity demand estimation using an adaptive neuro-fuzzy network: A case study from the Ontario province—Canada," *Energy*, vol. 49, pp. 323–328, Jan. 2013.
[47] M. Duan, A. Darvishan, R. Mohammaditab, K. Wakil, and O. Abedinia, "A novel hybrid prediction model for aggregated loads of buildings by considering the electric vehicles," *Sustain. Cities Soc.*, vol. 41, pp. 205–219, Aug. 2018.
[48] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining* (Intelligent Systems Reference Library). Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-10247-4.
[49] I. B. Shiroky, A. Papangelo, N. Hoffmann, and O. V. Gendelman, "Nucleation and propagation of excitation fronts in self-excited systems," *Phys. D, Nonlinear Phenomena*, vol. 401, Jan. 2020, Art. no. 132176.
[50] K. Singh, S. Shastri, A. S. Bhadwal, P. Kour, M. Kumari, A. Sharma, and V. Mansotra, "Implementation of exponential smoothing for forecasting time series data," *Int. J. Sci. Res. Comput. Sci. Appl. Manag. Stud.*, vol. 8, no. 8, 2019.
[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[52] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
[53] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.
[54] Mustaqeem and S. Kwon, "MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach," *Expert Syst. Appl.*, vol. 167, Apr. 2021, Art. no. 114177.
[55] Mustaqeem, M. Sajjad, and S. Kwon, "Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM," *IEEE Access*, vol. 8, pp. 79861–79875, 2020.
[56] Mustaqeem and S. Kwon, "1D-CNN: Speech emotion recognition system using a stacked network with dilated CNN features," *Comput., Mater. Continua*, vol. 67, no. 3, pp. 4039–4059, 2021.
[57] T. Xia, Y. Song, Y. Zheng, E. Pan, and L. Xi, "An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation," *Comput. Ind.*, vol. 115, Feb. 2020, Art. no. 103182.
[58] G. Hebrail and A. Berard, "Individual household electric power consumption data set," UCI Mach. Learn. Repository, Paris, France, Tech. Rep., 2012.
[59] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.
[60] J.-Y. Kim and S.-B. Cho, "Electric energy consumption prediction by deep learning with state explainable autoencoder," *Energies*, vol. 12, no. 4, p. 739, 2019.
[61] D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2016, pp. 7046–7051.
[62] T. Le, M. T. Vo, B. Vo, E. Hwang, S. Rho, and S. W. Baik, "Improving electric energy consumption prediction using CNN and bi-LSTM," *Appl. Sci.*, vol. 9, no. 20, p. 4237, Oct. 2019.
[63] P. Chujai, N. Kerdprasop, and K. Kerdprasop, "Time series analysis of household electric consumption with ARIMA and ARMA models," in *Proc. Int. MultiConf. Eng. Comput. Scientists*, 2013, pp. 295–300.
[64] R. Rajabi and A. Estebsari, "Deep learning based forecasting of individual residential loads using recurrence plots," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–5.
[65] D. Wu, B. Wang, D. Precup, and B. Boulet, "Multiple kernel learning-based transfer regression for electric load forecasting," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1183–1192, Mar. 2020.

MUSTAQEEM received the bachelor's degree in computer science from the Institute of Business and Management Sciences (IBMS), Agricultural University Peshawar, Pakistan, and the master's degree from Islamia College, Peshawar, Pakistan, with a focus on video analysis (content base video retrieval/action recognition). He is currently pursuing the Ph.D. degree in software engineering with the Interaction Technology Laboratory, Sejong University, Seoul, Republic of Korea. He is working as a Researcher at the Interaction Technology Laboratory (IT Lab). His major research interests include audio digital signals processing, energy managements and forecasting, smart energy consumption and prediction speech processing, speech synthesis and diarization, emotion recognition, and image and video processing. He served as a Professional Reviewer for several well-reputed journals, such as IEEE ACCESS, *Sensors* (MDPI), *Remote Sensing* (MDPI), *Applied Sciences* (MDPI), *Applied Soft Computing* (Elsevier), *Expert System with Applications* (Elsevier), *CMC-Computers, Materials & Continua* (TSP), and *International Journal of Intelligent Systems* (Wiley).

MUHAMMAD ISHAQ received the bachelor's degree in electrical engineering from the Sarhad University of Science and Technology (SUIT), Peshawar, Pakistan. He is currently pursuing the master's degree with the Software Engineering Department, Sejong University, Seoul, Republic of Korea. He is working as a Researcher at the Interaction Technology Laboratory (IT Lab). His major research interests include smart energy consumption and prediction, audio digital signals processing, speech processing, speech synthesis, and diarization.

SOONIL KWON received the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, USA, in 2000 and 2005, respectively. He is currently a Professor with the Department of Software, College of Software Convergence, Sejong University, Seoul, Republic of Korea. He is the Head of the Interaction Technology Laboratory, Sejong University, where students are involved in research projects under his supervision, such as social data analysis, audio analysis, multi-modal data analysis and speech emotion recognition, speech synthesis, speaker recognition, and speaker diarization. His research interests include speech recognition, human–computer interaction, affective computing, and speech and audio processing. He served as a Professional Reviewer for several well-reputed journals, such as the *IEEE Communications Magazine*, *Sensors*, *Information Fusion*, *Information Sciences*, the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), *MBEC*, *MTAP*, *SIVP*, and *JVCI*.