IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 April 2023, accepted 8 May 2023, date of publication 15 May 2023, date of current version 24 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3276646

RESEARCH ARTICLE

An LSTM-SAE-Based Behind-the-Meter Load Forecasting Method

AYDIN ZABOLI[1], (Graduate Student Member, IEEE), VO-NGUYEN TUYET-DOAN[2], YONG-HWA KIM[3], (Member, IEEE), JUNHO HONG[1], (Senior Member, IEEE), AND WENCONG SU[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, University of Michigan-Dearborn, Dearborn, MI 48128, USA
[2]Department of Electronic Engineering, Myongji University, Yongin-si 17058, South Korea
[3]Department of AI Data Engineering, Korea National University of Transportation, Uiwang-si, Gyeonggi-do 16106, South Korea

Corresponding author: Yong-Hwa Kim (yongkim@ut.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] under Grant 2022R1F1A1074975; and in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry and Energy (MOTIE) of the Republic of Korea under Grant 20221A10100011.

ABSTRACT Nowadays, modern technologies in power systems have been attracting more attention, and households can supply a portion of or all of their electricity based on on-site generation at their location. This can be challenging for utilities in terms of monitoring and recording the data because the households' facilities can generate or consume the energy without passing it through a meter, increasing the complexity of a distribution network. The speed of transferring data to utilities is another important concern. There is a necessity to send the smart meter (SM) data of each house to a distribution management system (DMS) for more analysis in the shortest possible time. This paper presents a novel deep learning framework collaborating with sequence-to-sequence (seq2seq), long short-term memory (LSTM), and stacked autoencoders (SAEs) to forecast residential load profiles considering the photovoltaic (PV), battery energy storage system (BESS), and electric vehicle (EV) loads with more capability based on pre-defined patterns. Experimental results show that the proposed method achieves outstanding performance in the forecasting process of residential load profiles in comparison with other algorithms. Also, a smart distribution transformer can help utilities to receive the data instantly via wireless communication, which can reduce the transfer duration to every minute and make the prediction and monitoring more manageable considering the different combinations of distributed energy resources (DERs) in residential locations.

INDEX TERMS Battery energy storage system, behind-the-meter, electric vehicle, load forecasting, LSTM-based sequence-to-sequence, photovoltaic system, smart meters.

I. INTRODUCTION

Smart grid advancements have made it necessary to manage both sides of generation and demand for monitoring. Intelligent technologies for energy production have created an enthusiasm for people to utilize microgrids as on-site generation resources in their residences. To manage and curb the demand side effectively, it is indispensable to first perceive the nature of the loads on the demand side [1]. The position of an energy system in relation to SM determines whether it is a behind-the-meter (BTM) or front-of-meter (FOM) distribution system. A BTM system delivers electricity to on-site locations without passing it through a meter, whereas an FOM network transmits energy to off-site places. Before reaching an end-user, the power supplied by an FOM system must pass through a meter so that it is suitable for utilities to track the energy consumption of households. This helps utilities to provide an accurate prediction of pattern changes and facilitates the detection of privacy or cyber-attack issues for residential households.

The growing integration of DERs into power distribution networks has put pressure on utilities to enrich their

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

systematic awareness to implement load control techniques in BTM systems [2]. Consider a house with PV, BESS, and residential EV charging as a BTM model. In this case, the aforementioned facilities have different operational modes that contain the generation and stop modes for a PV system, in which the energy produced by the PV system can be hidden from utilities and does not pass through a meter for measurement. This happens for a BESS in discharging mode because it supplies energy for household appliances and an EV even though it is connected to the power grid. The surplus energy produced by DERs (e.g., rooftop solar systems) can be stored in a BESS and used locally as needed. A BESS would lead to increased self-consumption of energy produced by PV systems in some regions where electricity exported to the grid is not remunerated. Distribution system technicians may experience difficulties receiving higher variable generation from different sources of energy in areas with a high concentration of rooftop solar PV systems. Hence, residential PV, BESS, and EV charging assessments can be challenging for utilities in terms of electricity generation and consumption tracking.

When confronted with these concerns, load monitoring appears as an important aspect in a distribution grid of which one of the methods is intrusive monitoring, wherein each load/appliance has an information acquisition device. Non-intrusive load monitoring (NILM) is the process of disaggregating the collected electric power/current and converting it into individual appliances by assessing variations in the voltage and current coming into the load profile [3]. The measurements are made from the power entering the NILM, and the signal is processed and assessed, then network loads are detected. NILM techniques may be classified into two groups based on the machine learning (ML) method, including supervised and unsupervised algorithms. The supervised technique requires training to recognize the classes in order to conduct the load forecasting well; however, the unsupervised method has a lower cost [4].

The use of advanced metering infrastructure (AMI) has increased the monitoring capabilities of the grid. SMs, devices located at the customers' residences, are the main components of AMIs [2]. It is necessary for utilities to monitor the energy usage of customers during the day. Therefore, SMs can be installed in houses to send information to the distribution transformers, and then the transformers transfer data to the remote terminal unit (RTU) using a sensor remotely. SMs are very advantageous in measuring data for each house; some of their benefits include the elimination of manual readings, access to the consumption data 24/7, remote accessibility, prompt service, and high compatibility with distributed generation resources (e.g., PV and wind turbine) [5]. The RTU is a multi-functional electronic device controlled by a microprocessor that allows for remote monitoring and control of different automation systems. It is generally used in an industrial setting and performs a similar function to programmable logic circuits (PLCs) but to a greater extent. The database contains measurements of both real and reactive power injection [6].

A utility company (e.g., DTE Energy) deals with data collection from customers' SMs as quickly as possible. While the duration of data transfer is already every 15 minutes, that is not adequate and needs to be improved [7]. That means a distribution transformer reads the data from all houses (as a data collector) and sends the data to the utilities four times per hour. In this case, the utilities are not able to analyze each customer's data accurately, and the forecasting process cannot be applicable.

Deep learning algorithms have become one of the most prominent technologies, with a sophisticated and mathematically complex evolution of ML methods. Recently, this field has been getting significant attention in many research areas related to the multiple-layer structures of neural networks. It depends on stochastic optimization to improve ML capability and task performance. The LSTM model can be regarded as the most effective solution for variable-length sequences due to its property of selectively remembering patterns by having a recurrent hidden state, the activation of which at each time is reliant on that of the previous time. Hence, the LSTM algorithm has been applied extensively in time-series prediction [8]. However, the features have several highly correlated variables in the multi-time series data problems. Random initialization of a large number of neurons is a challenge of the LTSM method that will lead the learning algorithm to converge to different local minima, depending on the values of the parameter initialization [9]. Therefore, an SAE has recently been widely applied in many fields to overcome the randomized initial weight obstacle of LSTM algorithms. It also outperforms due to its ability to synthesize deep information from complex data [10], [11], [12], [13].

A. MOTIVATION AND CONTRIBUTIONS

This paper focuses on the BTM forecasting process in the presence of different combinations of PV, BESS, and EV for three houses using a seq2seq LSTM-SAE architecture. The SAE is the core structure of the model adopted to learn the shallow and deep features of the residential load profiles by exploiting multiple single-layer autoencoders in which the output feature of each layer is fed to the inputs of the successive layer. As a result, SAEs have achieved efficiency in representing invariant and abstract features [14]. Moreover, an LSTM architecture is incorporated for enhanced prediction accuracy. This study proposes the seq2seq LSTM-based model to predict the load changes, which is more accurate and efficient than other algorithms based on the performance evaluation metrics. Furthermore, this dataset is based on times-series changes for every minute that can make a better understanding of utilities to follow the energy consumption of customers. It helps to make more accurate predictions of load patterns. To recap, the final goal of this study is to assist utilities in receiving SM data instantly to provide an accurate analysis of their consumers' load patterns. The major contributions of this paper are as follows:

- An assessment of a new dataset is carried out with different DERs to infer important features linked to time. The time interval is one minute (rather than the conventional transfer time of 15 minutes). In contrast to other research, this paper considers BESS and EV consumption data in addition to the PV systems. Other papers are focused on PV forecasting in most cases. It helps utilities to make better monitoring of the consumers' energy consumption.
- A seq2seq LSTM-SAE algorithm is introduced to forecast residential load profiles. This method offers the advantages of synthesizing abstract features from complex data and solving the random weight initialization problem of the LSTM algorithm.
- The proposed model is verified through a BTM system by considering different DERs with the feature of remembering patterns. This algorithm is able to determine whether data in a sequence should be retained or discarded based on its importance. Then, it can transfer important information down a long chain of sequences to recognize patterns as a result. This method shows better accuracy and efficiency, considering the performance evaluation metrics, than previous methods.

B. PAPER ORGANIZATION

The rest of the paper is organized as follows: Section II provides a literature survey on limitations and existing problems regarding ML methods in BTM systems. Section III presents the configuration of the BTM system along with some characteristics of the distribution transformer. Section IV outlines the principles of load prediction as well as the cyber-physical system, including the data description. The proposed ML method for load forecasting is presented in Section V. Section VI discusses the results and validation of the proposed method in terms of accuracy and speed by providing performance analyses. Finally, the paper is concluded in Section VII.

II. LITERATURE REVIEW

There has been a wide range of approaches proposed for load forecasting methods based on ML algorithms. Statistical approaches (e.g., multiple linear regression and exponential smoothing) and ML methods, including artificial neural networks (ANNs), support vector machine (SVM), and convolutional neural network (CNN) have been proposed by researchers as forecasting methods [15], [16], [17], [18], [19], [20]. Khodayar et al. [21] proposed a novel spatiotemporal graph method to determine the most important spatiotemporal aspects of the net demand using hidden features to disaggregate the historical demand into PV generation and other loads for future pattern forecasting. However, this approach provides a static description of the correlation across wind farms, which evidently fails to account for the temporal and spatial dynamics involved in correlation. In [22], a short-term household load forecasting technique was introduced that used deep learning to exploit the spatiotemporal correlation in the load data of appliances. However, the authors did not consider DERs with unpredictable and fluctuating behavior (e.g., EV and PV), which penetrate smart grids through households these days and have considerable impacts on energy consumption. Wang et al. [23] proposed a deep CNN to exploit information from SMs automatically. To detect the socio-demographic statistics of consumers, the CNN is paired with an SVM algorithm that can eliminate the overfitting issue. According to [24], the authors developed a forecasting model using a hybrid method that includes similar day selection, empirical mode decomposition, and an LSTM algorithm. Li et al. in [4] elaborated an extraction technique based on current data aggregation over time and frequency to alleviate the issue of employing a specific appliance feature by introducing a combined CNN algorithm. Rafi and colleagues proposed a serial CNN-LSTM model. The CNN module first collected the trend of the heat load data, which was subsequently flattened for input into the LSTM module [25].

To forecast the short-term energy needs of consumers, Kong et al. [26] employed the LSTM model, in which the density-based clustering methodology was used to assess how well users' energy consumption patterns match up with one another. Ahmad et al. [27] developed a forecasting model based on random forests and then compared the results to those obtained by the SVM algorithm. It has been demonstrated that the forecasting model that they have proposed has good performance in terms of mean absolute error (MAE) and root mean square error (RMSE). Nevertheless, it mostly relies on PV generation as a DER. A finite-state-machine (FSM) algorithm for extracting the root mean square (RMS) current and time value was investigated in [28] to identify the critical state and detect repeating patterns of household loads. The implementation of this technique is made more difficult by the fact that long-term load activity results in a prolonged data sample time. This makes the technique challenging to implement. Regarding the SMs data, Oprea et al. proposed an unsupervised ML method to detect anomalous data provided by SMs because there are some problems in the data extracted from the meters (e.g., missing data, fraudulent data, duplication) [29].

Massaoudi et al. proposed two integrated algorithms to create a novel computing platform for merely solar forecasting methods. Using correlation patterns between weather data and meta-learning aggregation as well as hyperparameter tuning in ML algorithms, this model provides an accurate daily prediction of PV output. However, only the values that can be reasonably anticipated for the future generation of solar power are predicted using such a method. Taking into account the intrinsic fluctuations and unpredictability of solar resources, this strategy will inevitably lead to an error in PV power forecasting, which will propel inaccurate subsequent decisions regarding energy management. Also, different combinations of DERs could be more challenging to implement for a forecasting process [30], [31]. A method for BTM PV-load decomposition and customer baseline prediction that considers the net load data, temperature, and solar irradiation was addressed for a residential distribution system. It relies on historical data rather than requiring specific information on individual PV outputs [32]. For the purpose of estimating the average temperature inside a building, Lin et al. [33] suggested a hybrid approach for the short-term load forecasting method of the individual user to analyze a dynamic model based on the temperature. This model combines less historical data and output weather conditions for the heater and air conditioning (HAC) and non-HAC loads (e.g., appliances and electric lighting) as the input. The authors used the bi-directional LSTM algorithm to forecast the load changes, but this method and input are not accurate and comprehensive, and they have the drawback of ignoring the auto-encoding process of the algorithm, which is indispensable for an accurate prediction. Furthermore, their input does not include important DERs, which are very critical and make the forecasting process complicated. To disaggregate a number of customers' net load measurements into their electric load and solar PV power, Kabir et al. represented an unsupervised approach that used a physical PV system framework (not other BTM resources such as EV and BESS) for individual solar PV generation estimation. However, they employed the hidden Markov model to estimate the joint load. A 15-minute period of load pattern variations is obtained and used by the model to predict load behavior, which cannot efficiently preserve the privacy of energy consumers [34].

In the modern power grid, the factors that contribute to uncertainty include not only the appliances, power loads, and heat loads but also the PV systems, BESSs, EVs, wind turbines, and so on. A multi-energy power grid is also becoming incredibly popular. The provision of power, heating, air conditioning, and gas should all be accompanied by equivalent forecasting systems. In the context of the LSTM forecasting method, it is important to understand how to take into consideration the connection and coordination forecasting processes of various forms of energy. Hence, a new dataset considering different forms of energy resources, including PV, BESS, and EV, with an interval of one minute was considered to enhance the prediction precision by employing the proposed algorithm.

III. REPRESENTATION OF BEHIND-THE-METER OPERATION IN DISTRIBUTION SYSTEMS

In recent years, different DERs have been utilized to generate, store, and transmit energy, making distribution systems more sophisticated. Residential buildings can generate their own energy from some DERs and then store their surplus energy in BESSs for future consumption. Hence, it can be difficult and more complicated for utilities to detect all the energy consumption by residential buildings. The three main components of BTM systems can be defined as PV systems, BESSs, and EV chargers. In this system, PV systems and

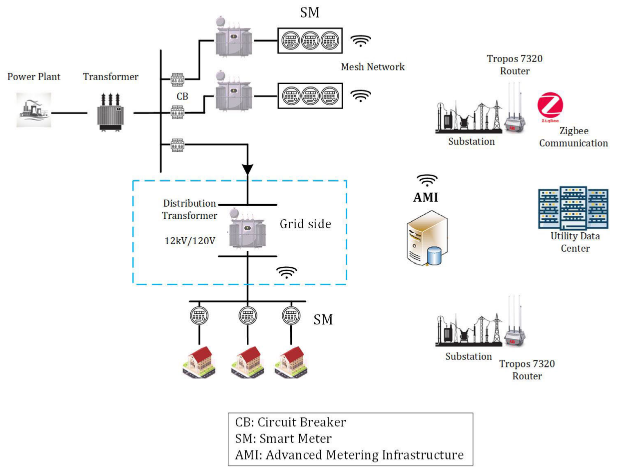


FIGURE 1. A typical distribution system with different facilities for the data transfer process.

BESSs are connected to each other as well as to the power grid using an inverter. Customers will be able to use the grid as a source of backup energy while earning credits for exporting energy to the power grid. The largest contributions of energy consumption and generation can be considered PV generation, BESS charging and discharging, and EV charging modes. Fig. 1 illustrates the general scheme of a distribution system with different components, including SMs, AMIs, power substations, and utility data centers.

As can be observed, a three-feeder power system with distribution transformers for each feeder is considered. Overcurrent relays (OCRs) are installed at the beginning of the feeder to disconnect the feeder when fault currents occur. Houses have SMs connected to each other and to the main power grid by a mesh network. Miniature circuit breakers (MCBs) are installed in houses for possible tripping, and an AMI collects data through the distribution transformer and transfers the data remotely by WiFi or ZigBee communications to the utilities for further analysis. Wireless communications can be conducted through the Tropos router, which connects all wireless facilities in a robust way. Moreover, all MCBs are closed during normal conditions. According to the given information, different combinations of three houses with defined parameters are shown in Fig. 2.

In order to transfer SMs' data to the utilities, some companies have provided a monitoring device for distribution transformers to send information to the DMS with high accuracy and reliability. GRID20/20, a company located in Richmond, VA, has presented a new solution for this concern. OptaNODE DTM, represented in Fig. 3, is a single-phase distribution transformer monitoring device that can connect to a distribution transformer. It combines precise sensing, measurement, and communication parts into a simple and well-set independent device. It can be attached to the transformer using embedded magnets and takes only a few minutes to set up [35]. Furthermore, there is another type of moni-

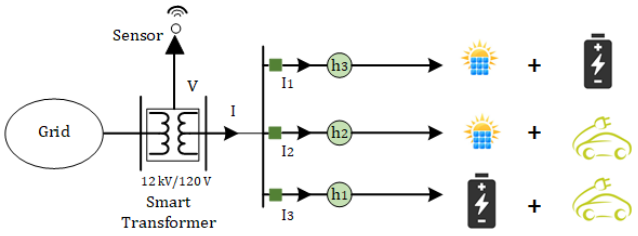


FIGURE 2. An illustration of BTM combinations with defined parameters for forecasting by an ML method.

FIGURE 3. A single-phase distribution transformer monitoring device manufactured by GRID20/20 [35].

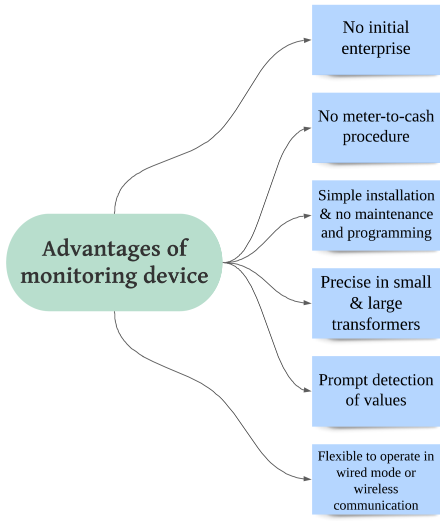


FIGURE 4. Advantages of an applicable distribution transformer monitoring device.

toring device called OptaNODE PDTM that can be utilized for three- or four-wire applications. Some advantages of this monitoring device are shown in Fig. 4.

In terms of communication technologies, it has several options for transferring the data, including WiFi technology based on IEEE 802.11g wireless, a global system for mobile (GSM) communication with an internal antenna, and power line communication. Moreover, it satisfies IEC 61000-4 requirements (e.g., electrostatic discharge) and immunity to electromagnetic and radiated fields, voltage interruptions, dips, and so on. Because this device does not need a meter-to-cash process, it can help utilities reduce the time and effort required to assess energy consumption on a daily basis, generate and distribute accurate bills, collect monthly payments from customers, and accurately record revenues [36]. Moreover, it has the specification of lower deployment costs and no maintenance. Also, it can decrease the efforts and costs for additional facilities that preserve the privacy of customers, so utilities are able to check and monitor the data variations every second remotely.

IV. LOAD FORECASTING

The goal of this paper is to present the aggregated load pattern forecasting for three houses with different combinations of PV, BESS, and EV using ML algorithms. There are baseline current loads in which combinations of PV, BESS, and EV can be added to the baseline load and can be read by a smart distribution transformer [37]. This current should be transferred to the DMS for further analysis that is required by the forecasting method to obtain the load pattern. These combinations have different operational modes that can be used to train an ML algorithm based on historical data. Fig. 5 illustrates an algorithm for the proposed methodology.

Based on the given flowchart, various combinations, including PV and BESS (house #1), PV and EV (house #2), and BESS and EV (house #3) loads are considered. Each combination has different operational modes, based on Table 1. Changes are applied to the baseline load, and the total current can be aggregated by a distribution transformer. This is the final current profile versus time for each minute. Utilities need to monitor households' consumption by predicting the integrated load pattern. They need to use intelligent algorithms to provide good accuracy and assessment of households' energy consumption and to preserve their privacy and security when an attack or anomalous behavior happens [18], [20].

PV generation can be estimated based on solar irradiation (W/m^2) data extracted from the National Renewable Energy Laboratory (NREL) [38] that can be converted to electric currents by suitable PV modules. Operational modes for BESSs and EVs have been considered based on historical data regarding other factors (e.g., peak times). Accordingly, the following equation can be calculated for each house. The purpose is to predict the I_{Tr} during specific times of the day for each house with high accuracy and minimum error using the proposed algorithm. There is no difference between choosing an electric current (ampere) or load power (kW) in the model because the voltage is considered constant (120 V) in the dataset. Thus, the final result will be the output currents of a distribution transformer regarding a combination of different DERs in which the load power values of PVs,

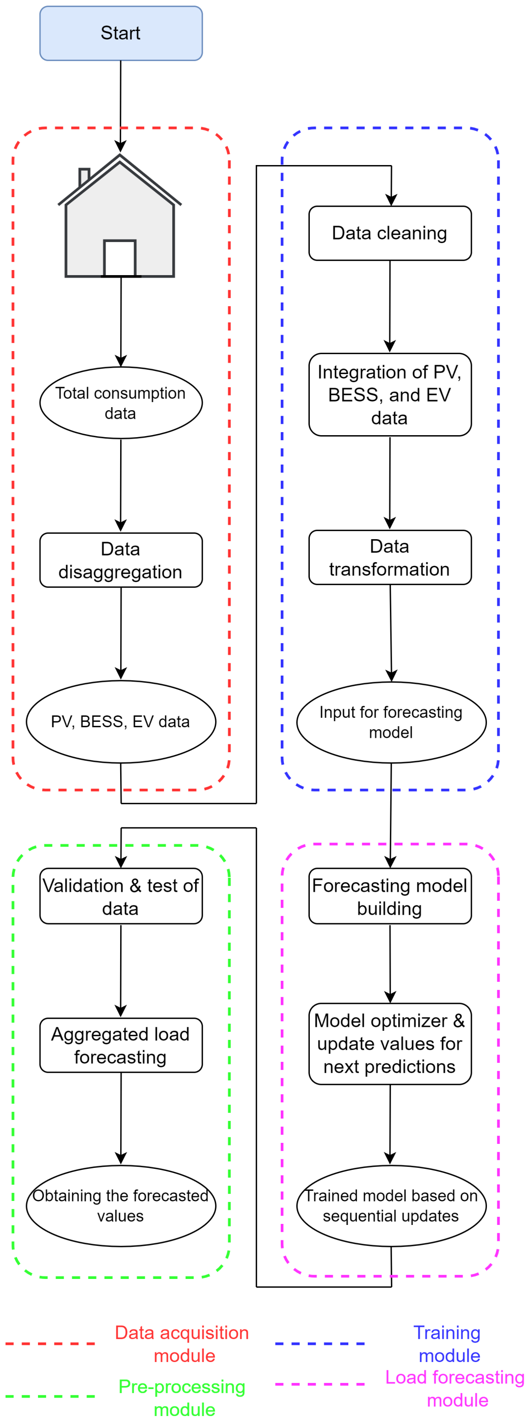


FIGURE 5. An architecture of proposed methodology.

BESSs, and EVs are converted to electric currents to be added to the baseline current (e.g., appliances and lighting) for each house. In Eq. (1), $I_{baseline}$ can be defined as the consumption current by other appliances in a house [30], [31]

$$I_{Tr} = I_{PV} + I_{BESS} + I_{EV} + I_{baseline}. \tag{1}$$

There are forecasting models for PV current due to solar radiation, BESS current, EV current based on the rated power, and residual current as Eqs. (2) – (4) [32], [39]:

$$\tilde{I}_{PV,t+h} = f_{I_{PV}}(I_{PV}, I_{BESS}, I_{EV}), \tag{2}$$
$$\tilde{I}_{BESS,t+h} = f_{I_{BESS}}(I_{PV}, I_{BESS}, I_{EV}), \tag{3}$$
$$\tilde{I}_{EV,t+h} = f_{I_{EV}}(I_{PV}, I_{BESS}, I_{EV}). \tag{4}$$

where $\tilde{I}_{PV,t+h}$, $\tilde{I}_{BESS,t+h}$, and $\tilde{I}_{EV,t+h}$ are defined as the forecasting of PV, BESS, and EV currents. Furthermore, $f_{I_{PV}}$, $f_{I_{BESS}}$, and $f_{I_{EV}}$ are denoted by the gradient boosting regression tree (as an ML method) for different currents. Now, Eqs. (5) – (8) are the vector combinations for feature data that may be used to perform an assessment:

$$I_{PV} = [I_{PV,t}, I_{PV,t+1}, I_{PV,t+2}, \ldots, I_{PV,t+n}], \tag{5}$$
$$I_{BESS} = [I_{BESS,t}, I_{BESS,t+1}, \ldots, I_{BESS,t+n}], \tag{6}$$
$$I_{EV} = [I_{EV,t}, I_{EV,t+1}, I_{EV,t+2}, \ldots, I_{EV,t+n}], \tag{7}$$
$$I_{res} = [I_{res,t}, I_{res,t+1}, I_{res+2}, \ldots, I_{res,t+n}]. \tag{8}$$

TABLE 1. Different operational modes of PV, BESS, and EV.

	Stop	Charging	Discharging (Generation)
PV	0	N/A	1
BESS	0	1	2
EV	0	1	N/A

A. CYBER-PHYSICAL SYSTEM FOR A BTM DERS

Some information regarding the given data is provided in this section. Solar irradiation for PV systems is extracted from NREL based on $Watt/m^2$ unit. According to Fig. 2, the voltage (V) level is constant in the study, and the current (I) is the aggregation of the three houses' electric currents. A 10 m^2 PV module is used and calculated the total PV currents for the 1^{st} and 2^{nd} houses. BESSs and EVs experience different operational modes during the day, as shown in Table 1. For instance, it is reasonable to consider the charging mode of EVs and BESSs from 12 a.m. to 6 a.m., which are not peak hours for energy consumption. PV operational modes are considered based on the irradiation levels, and once there is an irradiation during the day, it can be assumed to be 1; otherwise, it can be regarded as 0. The data collection is carried out based on the different parameters in which PV systems, EVs, and BESSs can be affected. The weather condition, PV array size, and irradiance changes during different days are considered for data generation of the PV resource. The peak hours are considered for data generation of EVs and BESSs to make appropriate artificial data. Because there is no comprehensive data that can include all DERs in a real-world scenario, all data are generated considering the aforementioned aspects based on one-minute intervals. There is no such residential data for these integrated DERs for the one-minute intervals. Most of the residential data are based on the 15-minute intervals in which all DERs are considered simultaneously. It makes a real-world scenario challenging without considering different types of DERs with one-minute

Time	V (volt)	PV_h1	BESS_h1	PV_h2	EV_h2	BESS_h3	EV_h3	H#1 Current (A)	H#2 Current (A)	H#3 Current (A)	I_Tr. (A)
4:10 PM	120	1	1	1	1	0	0	8.79644	6.55336	24.3665	39.7163
4:11 PM	120	1	1	1	1	0	0	8.58152	6.79671	24.3295	39.70773
4:12 PM	120	1	1	1	1	0	0	8.51745	7.12657	24.2909	39.93492
4:13 PM	120	1	1	1	1	0	0	8.42841	7.14349	24.272	39.84391
4:14 PM	120	1	1	1	1	0	0	8.35551	7.46381	24.2333	40.05262
4:15 PM	120	1	1	1	1	0	0	8.27267	8.12421	24.2028	40.59968
4:16 PM	120	1	1	1	1	0	0	8.11864	8.18975	24.1028	40.41119
4:17 PM	120	1	1	1	1	0	0	8.03727	8.27035	23.8984	40.20602
4:18 PM	120	1	1	1	1	0	0	7.98788	8.5169	23.8037	40.30848
4:19 PM	120	1	1	1	1	0	0	7.92707	8.51922	23.7964	40.24269
4:20 PM	120	1	1	1	1	0	0	7.86234	9.06818	23.7456	40.67612

FIGURE 6. A portion of feature data for different combinations of each house regarding a constant voltage.

intervals. This proposed data transfer time helps utilities to receive the residential smart meters' data every minute to have better monitoring of energy consumption during abnormal activities. A portion of the provided data displayed in the $h1$, $h2$, and $h3$ columns of Fig. 6 shows the various combinations of houses #1, #2, and #3, respectively. The next three columns demonstrate the total current values for each house, and I_{T_r} is the aggregated transformer current that must be transferred to the AMI.

Moreover, Fig. 7 illustrates the total electric current changes of each house as sampled during a day. It can be observed that houses #1 and #2 have more generations because of the presence of PV systems, and the current profiles are in the negative section most of the time (returning energy to the power grid). This originates from PV generation in these two houses during a specific time of solar irradiation. However, there is no significant generation for house #3, and it only takes advantage of BESS charging or discharging and EV charging processes without any generation. Now, the transformer current can be achieved using the electric current profiles for these houses, as demonstrated in Fig. 8.

V. PROPOSED METHODOLOGY

A utility company's operations and long-term strategy depend on reliable models for predicting electric power demand. Decisions on acquiring and generating electricity, load shifting, and infrastructure projects may all be made with the aid of a utility's load forecasting procedure. In different sectors of power systems (e.g., generation, transmission, and distribution), load predictions are essential for energy providers and financial institutions. Load forecasting is critical for utilities because of the unpredictable nature of demand and supply, as well as the impact of energy bills. Predicting short-term load flows and taking actions to avoid overloading may be done by using short-term forecasting techniques. The network's resilience and equipment failures are minimized as a result of the timely execution of these actions. When it comes to evaluating energy price contracts and other complex financial services, load forecasting would be a critical component [40].

Here, a regression problem for BTM load forecasting is defined. The architecture of the proposed model employs a seq2seq SAE using multiple LSTM models. The proposed model estimates transformer currents using a smart distribution transformer and SM data. The proposed model consists

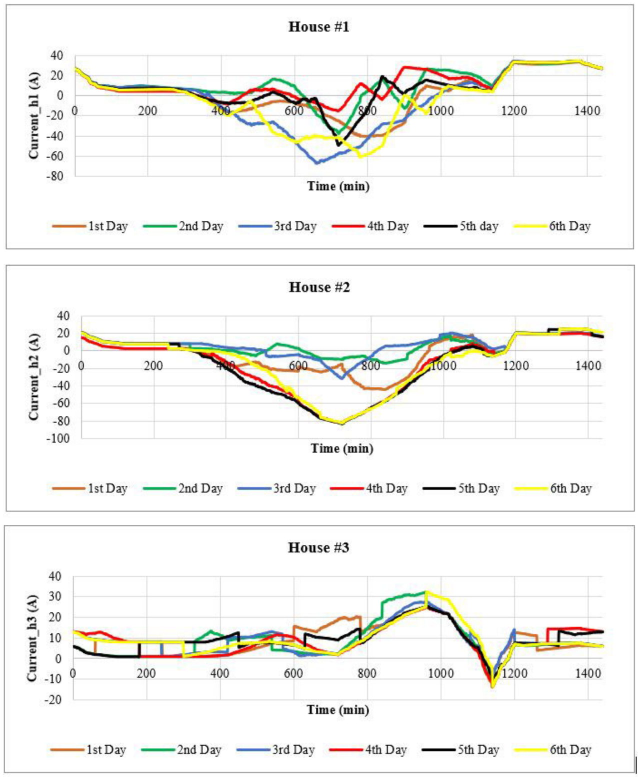


FIGURE 7. Household load profiles during 24 hours for six days.

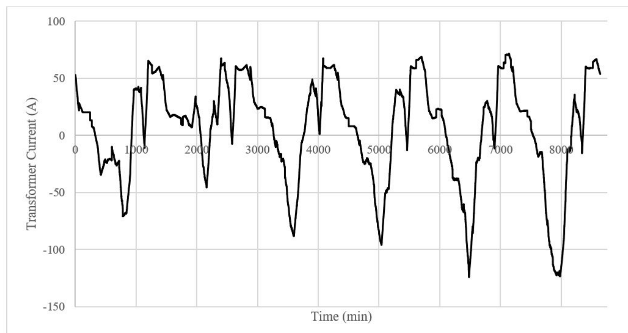


FIGURE 8. Total distribution of transformer current for a period of 24 hours.

of an LSTM encoder and decoder. The overall architecture of the proposed LSTM-SAE model is shown in Fig. 9. The encoder part compresses the given input sequence into a fixed-length vector to summarize the information of the input sequence. This fixed-length vector is known as a context vector and is used as the input of the decoder to predict the output sequence. Furthermore, a repeat vector layer and a time-distributed dense layer are added to this architecture. The repeat vector layer repeats the context vector obtained from the encoder. It is then fed to the decoder as input. This process is repeated for F steps, where F is the number of next steps that will be predicted in the future [41]. The output of the decoder with respect to each time step is mixed.

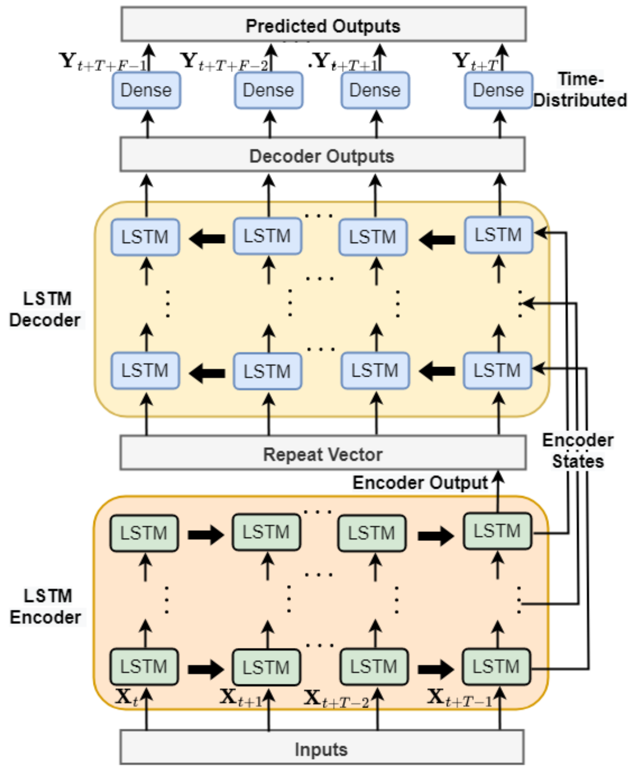


FIGURE 9. **An architecture of the stacked LSTM-SAE.**

Therefore, the LSTM decoder output is transformed directly by the time-distributed dense layer, utilizing a fully connected time-wrapped layer on each time step, and its output is separated.

An input sequence for the proposed LSTM-SAE can be written as Eq. (9):

$$\mathrm{X}_{t:t+T-1} = \mathrm{x}(t), \mathrm{x}(t+1), \ldots, \mathrm{x}(t+T-1), \tag{9}$$

where t is the time step, and $\mathrm{x}(t)$ is defined as:

$$\mathrm{x}(t) = [I_{Tr,t+T}, I_{h1,t}, I_{h2,t}, I_{h3,t}, h_t, m_t]^T. \tag{10}$$

In Eq. (10), $I_{Tr,t+T}$ is the electric currents of the transformer at time $(t+T)$; $I_{h1,t}$, $I_{h2,t}$, $I_{h3,t}$, h_t, and m_t are the electric current of houses #1, #2, #3, the hour, and minute parameters at the time t, respectively. The output sequence can be defined as Eq. (11):

$$\mathrm{Y}_{t+T:t+T+F-1} = \mathrm{y}(t+T), \mathrm{y}(t+T+1), \ldots, \mathrm{y}(t+T+F-1), \tag{11}$$

where F is the time window for the output, and $\mathrm{y}(t+T)$ at the time step $(t+T)$ is defined as Eq. (12):

$$\mathrm{y}(t+T) = [\tilde{I}_{h1,t+T}, \tilde{I}_{h2,t+T}, \tilde{I}_{h3,t+T}]^T. \tag{12}$$

In Eq. (12), $\tilde{I}_{h1,t+T}$, $\tilde{I}_{h2,t+T}$, and $\tilde{I}_{h3,t+T}$ are current estimates for houses #1, #2, and #3, correspondingly. Therefore, the proposed LSTM-SAE model predicts $\mathrm{Y}_{t+T:t+T+F-1}$ from observations $\mathrm{X}_{t:t+T-1}$. The pseudo-code for the proposed method presented in Algorithm 1 is explained as follows:

Algorithm 1 Pseudo-Code for Stacked LSTM-SAE Architecture

```
Input: Sequence time-series data X_{t:t+T-1} of length T in (9).
<LSTM Encoder>
Input: X_{t:t+T-1} of length T
for n = 1 : N layers do
    Random initial states;
    for t ∈ T time step do
        Step 1: Running LSTM architecture;
        Step 2: Update cell memory state vector and
        hidden state vectors.
    end for
    Save final states in each layer.
end for
Output: Summary state of final LSTM architecture
and final states in each layer.
Using a repeat vector from the output of the encoder in the forward F steps.
<LSTM Decoder>
Input: Repeat vector of length F.
for n = 1 : N layers do
    Initial states are updated from the final states of
    the encoder, respectively;
    for f ∈ F time step do
        Step 1: Running LSTM architecture;
        Step 2: Update cell memory state vector and
        hidden state vectors;
        Step 3: Save final outputs of LSTM architecture.
    end for
end for
Output: Summary final output of each LSTM
architecture in the F time step.
<Time distributed>:
Running fully connected dense layer for
three houses.
Output: Predict currents for houses #1, #2, and #3,
Y_{t+T:t+T+F-1} for F future observations in (11).
```

- The encoder layer of the LSTM-SAE processes the input sequence $\mathrm{X}_{t:t+T-1}$ of length T.
- The cell memory state vectors and the hidden state vectors are updated through the LSTM architecture at each time step. The encoder output summarizes the input sequence through the final LSTM state.
- The repeat vector layer is utilized to repeat the context vector received from the encoder to feed it into a decoder input and initial cell memory state vectors of the decoder layer apply vectors from the final encoder state at the respective time step.
- After that, the decoder learns features from the initial cell memory state vectors and the repeat vectors wrapped in the forward F steps that can be used to predict F future observations.
- Time distributed dense layer adopts a fully connected dense layer on each decoder output [9].

During the training phase, the adaptive moment estimation (Adam) algorithm (one of the best stochastic optimizer techniques for deep learning models) is chosen as the optimizer, with a constant learning rate of $lr = 0.001$ [42]. The parameters are learned via a back-propagation algorithm by minimizing the mean squared error (MSE) between the prediction and the ground truth.

A specific region's historical load data is gathered, and null values are evaluated. The variation of the load pattern acquired from the preceding record must be used to compensate for the missing data. As a result, information quality may be attained with minimal influence on predicting performance. Following that, a dataset must be segmented into test and training sets in order to assess the suggested modeling approach. This data frame may be reformed to define a model that can forecast the current profile for the next week or month. After the decoding process and time-distributed layer, the estimated current and predicted output can be achieved. Additionally, it is necessary to update the fit model function to integrate all features from the previous time steps as the input for the next step and consider every last observation as input data for the next time step [25].

VI. RESULTS AND DISCUSSION

In this section, the performance evaluations of the proposed LSTM-SAE model for the BTM load forecasting approach are presented. The input sequence is described by combining six features, including the electric current of the transformer; the electric current of houses #1, #2, and #3; the hour and minute parameters to acquire the appropriate characteristic according to the designed LSTM-SAE algorithm. Based on the purpose of this paper for the prediction of total electric currents in households for a specific area, the most important feature is the total electric currents of distribution transformers which is the aggregation of electric currents from houses #1, #2, and #3. Because of the uncertain behavior of users in households in energy consumption, the assessment of these currents is considered at time $(t + T)$ to check future changes. From that, the currents for houses #1, #2, and #3 are predicted. The simulation obtained a dataset containing 31,680 observations of minute-wise speed within the 22-date range. This dataset was divided into three subsets using 23040, 4320, and 4320 splits for training, validation, and testing, respectively. For the training and validation stages, a sliding window with a size of $T = 20$ and a recurrent step size of 1 is applied. The forecasting interval chosen is $F = 20$ minutes.

Table 2 depicts the optimized progression of the hyperparameters within the boundary range of minimum and maximum values and whether the parameter values were an integer or real. During the training phase, the optimization is executed to derive hyperparameters optimized through grid search, including the initial learning rate $\{0.1, 0.01, 0.001\}$, batch size $\{32, 64, 128, 256, 512\}$, number of layers $\{1, 2, 3\}$, and embedding size of the LSTM algorithm $\{32, 64, 128, 256, 512\}$. The model tunes different combinations of parameters,

TABLE 2. Maximum and minimum bound of hyperparameter optimization.

Hyperparameter	Minimum	Maximum	Type
Number of layers	1	8	Integer
Embedding Size	32	512	Integer
Initial Learning rate	0.001	0.1	Real
Batch size	32	512	Integer
Number of epochs	50	500	Integer

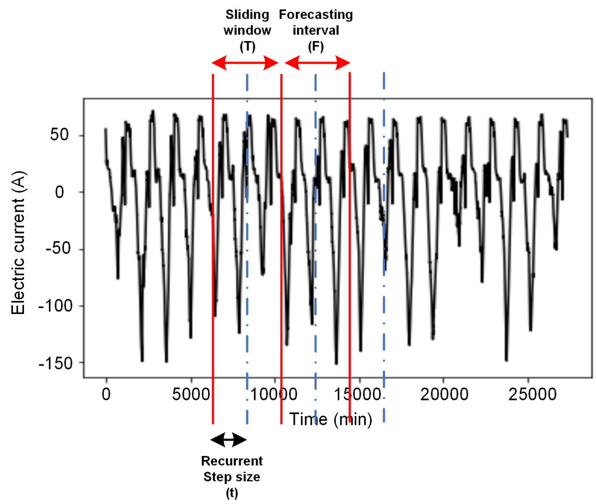


FIGURE 10. An architecture of a repeating LSTM block including four layers.

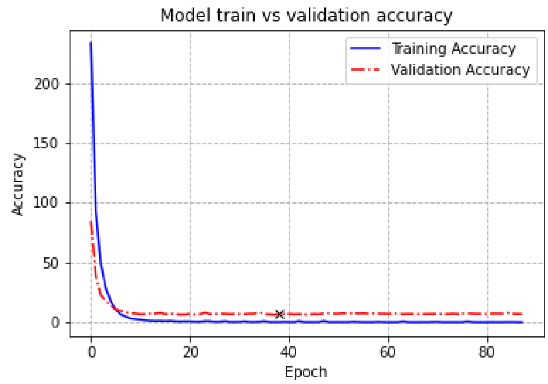
and the final proposed model is the best combination. Trials are carried out to alleviate the consequences of random initial values on the network during the tests, and then an average is made for the results. Such averaging is needed to enhance the robustness of the proposed model.

Using the BTM load dataset, the proposed LSTM-SAE model was compared with the conventional LSTM, Recurrent Neural Network (RNN), Feedforward Neural Network (FNN), and Deep Belief Network (DBN) algorithms to evaluate its predictive efficacy [43], [44]. During the training process, the hyperparameters of the models are derived by conducting multiple tests to choose the best combination for the models. Considering the conventional LSTM and the LSTM-SAE models, better results were obtained when using 3 hidden layers with 512, 256, and 128 hidden nodes at each layer, respectively. The number of epochs was 50 neurons, with a small batch size of 64 training samples. The outputs of the conventional LSTM and the LSTM-SAE models were also applied to the time-distributed dense layer with 3 nodes to produce the final output sequence predicting BTM loads for houses #1, #2, and #3 at each time step.

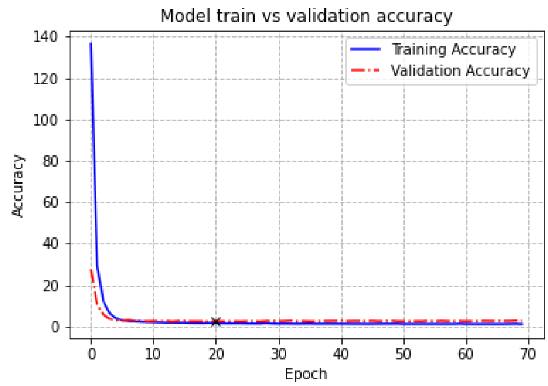
Overfitting problems often occur in deep learning models without sufficient training samples. Therefore, data augmentation algorithms have been adopted to increase the number of training samples to improve generalization accuracy and provide high reliability. Here, the training and validation processes use overlapping data with a sliding window with a size of $T = 20$, and the forecasting interval chosen is $F = 20$ minutes. A recurrent step size (t) of one minute is applied where $t = (0, 1, 2, \ldots)$. This process is depicted in Figure 10.

Fig. 11 shows the convergence of the proposed LSTM-SAE and conventional LSTM methods over epochs with the training and validation sets. The experiments were built using Python 3 under the framework of Keras on Tensorflow [45], [46]. The training process of all models was conducted on a PC computer with an AMD Ryzen Threadripper 2990WX 32-core processor with 80GB RAM and an NVIDIA GeForce RTX 2080 Ti GPU. The best performances for each model are analyzed. Here, early stopping is adopted in determining optimum parameters for predictive models [47], which quits the training process by monitoring validation loss to prevent overfitting issues in the training dataset. After 50 iterations, if the validation loss error diminishes again, the training process proceeds; otherwise, the training process is paused and the parameters are recorded. The proposed LSTM-SAE model achieved the minimum validation loss at 20 epochs while the conventional LSTM model achieved the minimum validation loss at 38 epochs. This indicates the improvement in forecasting performance and computational cost of the proposed method in employing residential electric loads along with the BTM system.

Table 3 shows a comparison of training and testing times for the DBN, FNN, RNN, LSTM, and proposed LSTM-SAE methods for three different forecast intervals (10 minutes, 15 minutes, and 20 minutes). It can be observed that the training time of the proposed LSTM-SAE model was slower than that of the other benchmark models. This was because the design of the LSTM-SAE utilized the composition of the encoder and decoder to overcome the problem of random initial values in the network. The testing time of the proposed LSTM-SAE model took a similar amount of time to make predictions compared to FNN, RNN, and LSTM. The LSTM-SAE model performed well not only in terms of prediction accuracy but also in terms of training and testing time, indicating that it can be a practical and efficient solution for the BTM system.



(a)



(b)

FIGURE 11. A loss improvement over epochs: (a) conventional LSTM and (b) proposed LSTM-SAE algorithms.

A. PERFORMANCE EVALUATION OF THE PREDICTION METHOD

The suggested BTM forecasting method is evaluated using the following metrics [48], [49]. Eq. (13) represents the error function of the network, which is called the **MSE**. When comparing predicted and actual values, the MSE index is able to calculate the differences in the square error. More accurate network predictions may be made with a lower MSE.

$$MSE = \frac{1}{n} \times \sum_{i=1}^{n} (x_{m,i} - x_{f,i})^2. \tag{13}$$

Mean absolute error (MAE) determines the average magnitude of errors in a series of forecasts without taking into account the direction of the errors. It is the average of the absolute changes between the predicted and actual interpretations over the test sample, in which all single deviations are given similar weights.

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n} | x_{m,i} - x_{f,i} |. \tag{14}$$

RMSE is an error indicator dependent on the scale, which is also used in the forecasting process and is represented as Eq. (15):

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (x_{m,i} - x_{f,i})^2}. \tag{15}$$

Mean absolute percentage error (MAPE) is the error criterion that is the most commonly used error metric in the forecasting domain, where n is the forecast horizon or the size of the sample, $x_{m,i}$ is the actual measured value at time i, and $x_{f,i}$ is the forecasted value for the same time.

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \frac{| x_{m,i} - x_{f,i} |}{x_{m,i}}. \tag{16}$$

Table 4 provides a comparison of the evaluation performance of the proposed LSTM-SAE model, conventional LSTM, RNN, FNN, and DBN for forecast intervals of 10, 15,

TABLE 3. Training and testing time comparisons of the models for different forecast intervals.

Time (min)	Model	DBN	FNN	RNN	Conventional LSTM	Proposed LSTM-SAE
10	Training-time	*9m 23s*	*9m 34s*	*57m 30s*	*12m 12s*	*20m 28s*
	Test time on the test set	*13ms*	*270ms*	*394ms*	*398ms*	*384ms*
15	Training-time	*9m 49s*	*9m 54s*	*53m 10s*	*13m 37s*	*20m 23s*
	Test time on the test set	*13ms*	*173ms*	*354ms*	*357ms*	*355ms*
20	Training-time	*9m 35s*	*10m 24s*	*1h 21 m 51s*	*14m 59s*	*22m 9s*
	Test time on the test set	*12ms*	*210ms*	*346ms*	*344ms*	*313ms*

TABLE 4. The performance metric comparisons of the models for different forecast intervals.

Metric/Model		DBN	FNN	RNN	Conventional LSTM	Proposed LSTM-SAE
10 mins	MSE	443.039	1.658	2.948	2.951	1.178
	MAE	15.529	0.866	1.022	0.971	0.631
	RMSE	21.048	1.288	1.717	1.718	1.085
	MAPE	328.558%	52.781%	35.403%	30.751%	23.200%
15 mins	MSE	444.628	3.230	4.618	5.494	1.733
	MAE	15.595	1.233	1.248	1.306	0.733
	RMSE	21.086	1.797	2.149	2.344	1.316
	MAPE	337.005%	71.285%	45.316%	34.195%	30.819%
20 mins	MSE	445.003	5.460	7.024	8.671	4.543
	MAE	15.647	1.634	1.560	1.618	1.070
	RMSE	21.095	2.337	2.650	2.945	2.131
	MAPE	344.906%	99.183%	61.192%	56.449%	43.525%

and 20 minutes including three houses. From this table, it is evident that the proposed LSTM-SAE model outperformed the other ML models for most of the evaluation metrics and forecast intervals. For instance, in terms of MSE, the LSTM-SAE model achieved the lowest values (i.e., 1.178, 1.733, and 4.543) compared to the other models for the 10, 15, and 20 mins time intervals, respectively. Similarly, for MAE, RMSE, and MAPE, the LSTM-SAE model outperformed the other models, achieving the lowest values for most of the evaluation metrics and forecast intervals. Therefore, based on this table, it can be inferred that the proposed LSTM-SAE model is superior to the other models evaluated in this study, in terms of its ability to accurately predict the outcome for the various energy resources of the BTM system in houses.

The MSE, MAE, RMSE, and MAPE indexes for the electric load and combination of BTM resources for each residential location are reported in Table 5. As can be inferred from this table, the LSTM-SAE model outperforms the DBN, FNN, RNN, and traditional LSTM algorithms. Especially, the MAPE indicator of the LSTM-SAE model has been perceived as effective in forecasting recurrent load at 38.223% which is much lower than that of the DBN, FNN, RNN, and traditional LSTM models at 147%, 92.224%, 109.451%, and 63.084% for house #1, respectively. Meanwhile, the indexes of MAPE of the LSTM-SAE model have also shown a capability of estimating with accuracy rates of 78.875% and 13.476% for house #2 and house #3, correspondingly, which are significantly superior to other algorithms. These results corroborate the premise of this work in which an LSTM-SAE structure is helpful for the forecasting process of the BTM system in residents based on different combinations of DERs which has many complexities and fluctuations in load profiles. These combinations can make the forecasting process more challenging in comparison with only PV systems as a resource.

To dive into the behavioral difference between conventional LSTM and the proposed LSTM-SAE, Figs. 12–14 illustrate the forecasts for the BTM loads estimated by conventional LSTM and LSTM-SAE models for houses #1, #2, and #3, respectively. As can be seen, both methods produce smooth series. To be more specific, the MAPE values are calculated for two different time ranges zoomed in the figures where the accuracy is low due to the large variation in the load profiles at that time. It shows that the proposed LSTM-SAE model performs better than the other baseline models for predicting the electric currents related to houses #1, #2, and #3.

The LSTM-SAE algorithm provides more accurate forecasting compared with other algorithms in all error metrics. These four evaluation indexes are obtained based on the prediction of electric currents (voltage is considered as a constant) for each house as can be observed in Figs. 12–14. The proposed algorithm shows an enhanced accuracy (e.g., the MSE index, of twice that of the conventional LSTM for

TABLE 5. The performance metrics of the proposed LSTM-SAE models for three houses.

Metric/Model		DBN	FNN	RNN	Conventional LSTM	Proposed LSTM-SAE
House #1	MSE	486.194	7.128	9.804	12.628	6.915
	MAE	17.226	1.913	2.082	2.008	1.243
	RMSE	22.050	2.670	3.131	3.554	2.630
	MAPE	147.231%	92.224%	109.451%	63.084%	38.223%
House #2	MSE	784.783	6.208	8.436	10.688	5.018
	MAE	23.806	1.715	2.171	1.857	1.222
	RMSE	28.014	2.492	2.904	3.269	2.240
	MAPE	697.770%	171.710%	176.503%	89.681%	78.875%
House #3	MSE	64.033	3.045	6.691	2.697	1.695
	MAE	5.912	1.274	1.972	0.990	0.744
	RMSE	8.002	1.745	2.587	1.642	1.302
	MAPE	189.719%	33.615%	68.736%	16.582%	13.476%

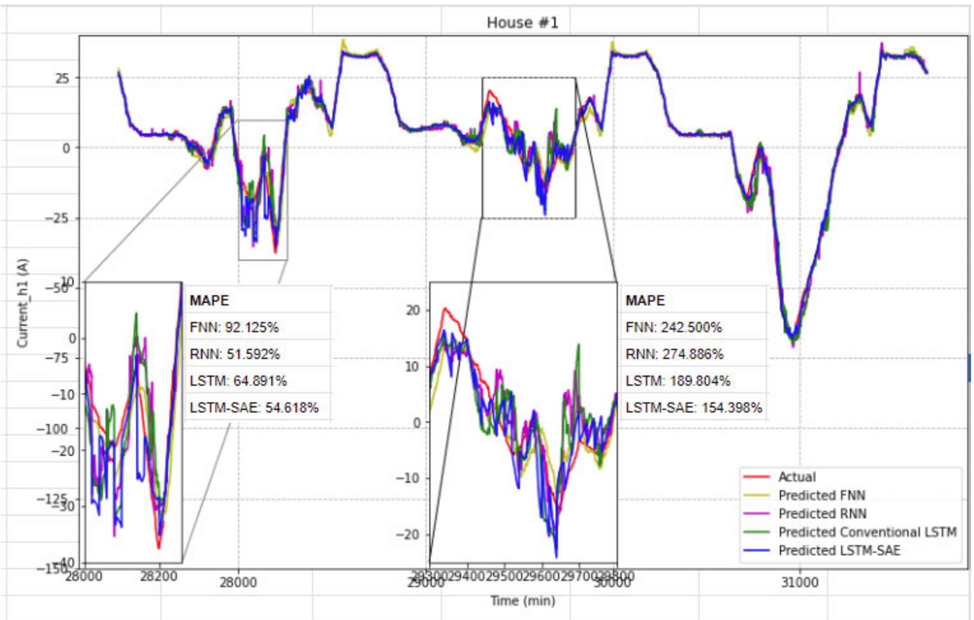


FIGURE 12. The forecasting results of different algorithms in comparison with the proposed LSTM-SAE models for house #1.

house #2). An error reduction (based on the MSE index) of the proposed LSTM-SAE model for house #2, in comparison with FNN, RNN, and conventional LSTM models, can be written as follows, respectively:

$$\text{Improvement}_{FNN}\ (\%) = \frac{6.208 - 5.018}{6.208} \times 100 = 19.17\%$$
$$\text{Improvement}_{RNN}\ (\%) = \frac{8.436 - 5.018}{8.436} \times 100 = 40.52\%$$
$$\text{Improvement}_{LSTM}\ (\%) = \frac{10.688 - 5.018}{10.688} \times 100 = 53.05\%$$

This observation indicates that not only the LSTM-SAE structure learning component contributes a significant part to the overall accuracy, but an SAE also has contributed crucially to the overall performance thanks to the learning strategy of initialization parameter values.

There are two main reasons for this proposed algorithm's better performance in terms of the power system's perspective. First, better accuracy in the forecasting process than in previous methods is demonstrated. Unlike most of the papers, which focus on PV generation forecasting as a BTM load, this research proposed different combinations of resources with variable energy consumption. Therefore, residential electric loads along with these BTM systems are proposed as a real-world scenario in this paper. Second, the sampling period for data is one minute for this research, so it can provide more accurate data with high reliability in the forecasting procedure, while most of the other research

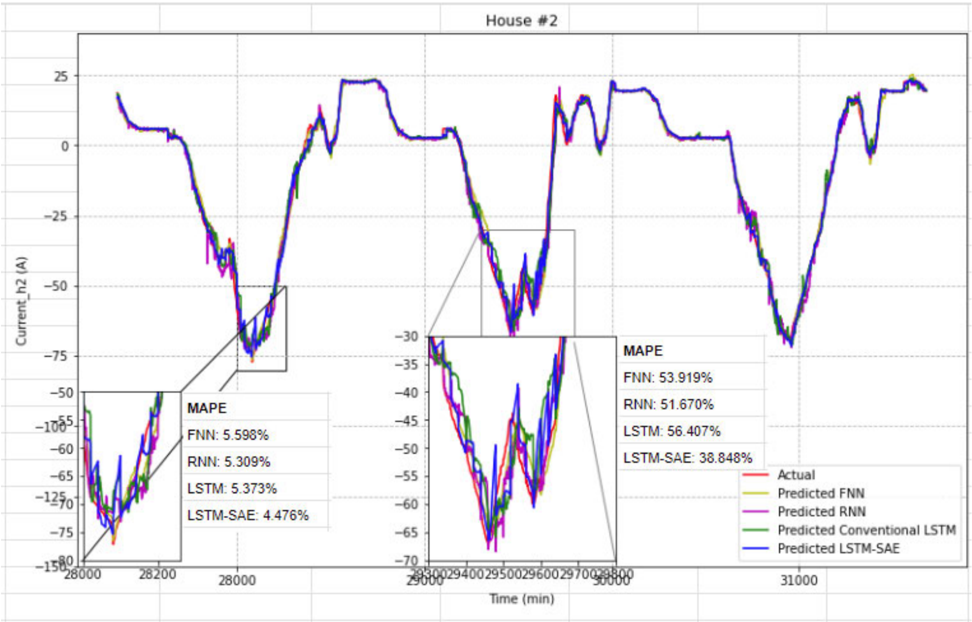


FIGURE 13. The forecasting results of different algorithms in comparison with the proposed LSTM-SAE models for house #2.

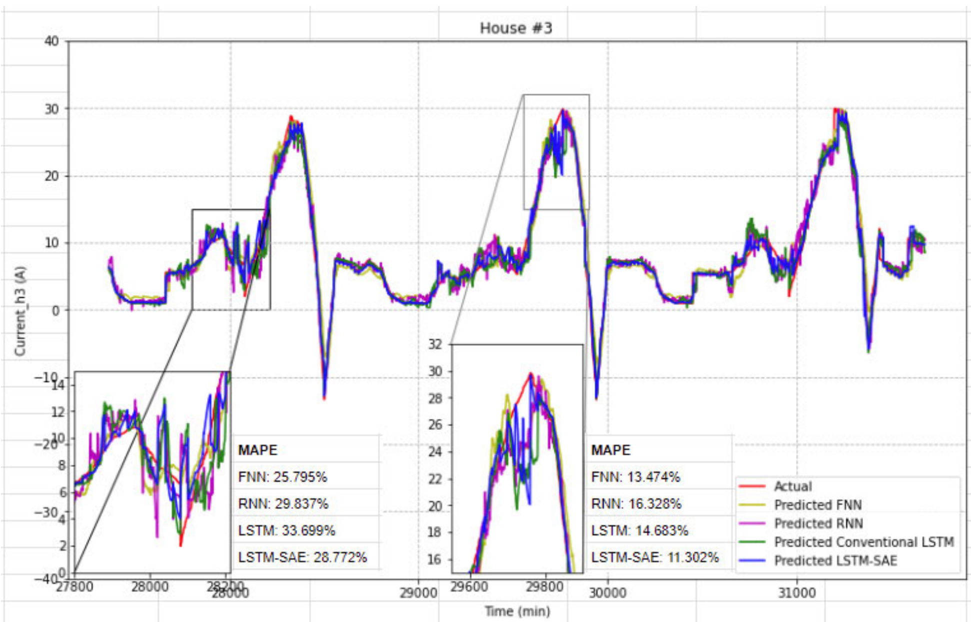


FIGURE 14. The forecasting results of different algorithms in comparison with the proposed LSTM-SAE models for house #3.

deals with a 15-minute time period. A sampling procedure for every minute is very helpful for utilities to conveniently estimate and predict their customers' energy consumption, and also can be useful for customers by preserving their privacy regarding any possible anomaly in their SM data. Considering Table 5, there is a slight difference between the MAPE indexes for the house #3. That is because there is no PV generation in this house, and BESS and EV are regarded as BTM loads in which BESS has a discharging mode in this case.

VII. CONCLUSION

In this paper, a more complex case of a BTM load is presented that concentrates on more than PV generation for forecasting analysis. Alternative combinations of PV, BESS, and EV loads for each household are assessed, in which different factors affect the residential feature data. Furthermore, a forecasting process based on data extraction for every minute is another contribution of this research. Finally, an LSTM-SAE algorithm is proposed based on stacked decoding to predict the total load currents of each house with good accuracy.

The results demonstrate better performance and accuracy compared with the traditional LSTM algorithms, even though there is a variational BTM load (e.g., PV, BESS, and EV charging) with fluctuating load patterns. A purpose for future work is to forecast different combinations of residential BTM loads, considering complex features of PV, BESS, and EV components, by extracting data for every second with high accuracy.

REFERENCES

[1] E. Proedrou, "A comprehensive review of residential electricity load profile models," *IEEE Access*, vol. 9, pp. 12114–12133, 2021.

[2] Y. Yuan and Z. Wang, "Mining smart meter data to enhance distribution grid observability for behind-the-meter load control: Significantly improving system situational awareness and providing valuable insights," *IEEE Electrific. Mag.*, vol. 9, no. 3, pp. 92–103, Sep. 2021.

[3] M. A. Devlin and B. P. Hayes, "Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data," *IEEE Trans. Consum. Electron.*, vol. 65, no. 3, pp. 339–348, Aug. 2019.

[4] K. Li, B. Yin, Z. Du, and Y. Sun, "A nonintrusive load identification model based on time-frequency features fusion," *IEEE Access*, vol. 9, pp. 1376–1387, 2021.

[5] (2021). *Smart Meters*. DTE Energy, Detroit, MI, USA. [Online]. Available: https://www.newlook.dteenergy.com/wps/wcm/connect/dte-web/home/service-request/common/natural-gas/rates/advanced-meters

[6] R. Dutta, S. Chakrabarti, and A. Sharma, "Topology tracking for active distribution networks," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 2855–2865, Jul. 2021.

[7] (2021). *Leading the Way to a Cleaner, Safer and Smarter Energy Future*. DTE Energy, Detroit, MI, USA. [Online]. Available: https://newlook.dteenergy.com/wps/wcm/connect/dte-web/home/about-dte/common/about-dte/about-dte

[8] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1724–1734.

[9] S. Banerjee, M. Rath, T. Swain, and T. Samant, "Music generation using time distributed dense stateful char-RNNs," in *Proc. IEEE 7th Int. Conf. Converg. Technol. (ICT)*, Apr. 2022, pp. 1–5.

[10] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2438–2442, Dec. 2015.

[11] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du, and S. Marshall, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, Apr. 2016.

[12] A. Sagheer and M. Kotb, "Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, Dec. 2019.

[13] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.

[14] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[15] I. Keka and B. Çiço, "Pattern detection of load profiles based on regression model with multiple variables," in *Proc. 7th Medit. Conf. Embedded Comput. (MECO)*, Jun. 2018, pp. 1–4.

[16] A. Abbas and B. Chowdhury, "A data-driven approach for providing frequency regulation with aggregated residential HVAC units," in *Proc. North Amer. Power Symp. (NAPS)*, Oct. 2019, pp. 1–6.

[17] Y. Chakhchoukh, P. Panciatici, and L. Mili, "Electric load forecasting based on statistical robust methods," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 982–991, Aug. 2011.

[18] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Estimation of behind-the-meter solar generation by integrating physical with statistical models," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2019, pp. 1–6.

[19] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1739–1747, May 2018.

[20] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3255–3264, May 2018.

[21] M. Khodayar, G. Liu, J. Wang, O. Kaynak, and M. E. Khodayar, "Spatiotemporal behind-the-meter load and PV power forecasting via deep graph dictionary learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4713–4727, Oct. 2021.

[22] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, "A deep learning method for short-term residential load forecasting in smart grid," *IEEE Access*, vol. 8, pp. 55785–55797, 2020.

[23] Y. Wang, Q. Chen, D. Gan, J. Yang, D. S. Kirschen, and C. Kang, "Deep learning-based socio-demographic information identification from smart meter data," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2593–2602, May 2019.

[24] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, p. 1168, Aug. 2017.

[25] S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated CNN and LSTM network," *IEEE Access*, vol. 9, pp. 32436–32448, 2021.

[26] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[27] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465–474, Dec. 2018.

[28] L. Du, Y. Yang, D. He, R. G. Harley, and T. G. Habetler, "Feature extraction for load identification using long-term operating waveforms," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 819–826, Mar. 2015.

[29] S.-V. Oprea, A. Bara, F. C. Puican, and I. C. Radu, "Anomaly detection with machine learning algorithms and big data in electricity consumption," *Sustainability*, vol. 13, no. 19, p. 10963, Oct. 2021.

[30] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, H. Abu-Rub, and F. S. Oueslati, "An effective hybrid NARX-LSTM model for point and interval PV power forecasting," *IEEE Access*, vol. 9, pp. 36571–36588, 2021.

[31] M. Massaoudi, H. Abu-Rub, S. S. Refaat, M. Trabelsi, I. Chihi, and F. S. Oueslati, "Enhanced deep belief network based on ensemble learning and tree-structured of Parzen estimators: An optimal photovoltaic power forecasting method," *IEEE Access*, vol. 9, pp. 150330–150344, 2021.

[32] K. Pan, C. Xie, C. S. Lai, D. Wang, and L. L. Lai, "Photovoltaic output power estimation and baseline prediction approach for a residential distribution network with behind-the-meter systems," *Forecasting*, vol. 2, no. 4, pp. 470–487, Nov. 2020.

[33] X. Lin, R. Zamora, C. A. Baguley, and A. K. Srivastava, "A hybrid short-term load forecasting approach for individual residential customer," *IEEE Trans. Power Del.*, vol. 38, no. 1, pp. 26–37, Feb. 2023.

[34] F. Kabir, N. Yu, W. Yao, R. Yang, and Y. Zhang, "Joint estimation of behind-the-meter solar generation in a community," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 682–694, Jan. 2021.

[35] (2021). *OptaNODE DTM*. GRID20/20, Richmond, VA, USA. [Online]. Available: https://grid2020.com/products

[36] (2021). *Cost Effective Solution for Electric Distribution Utilities*. GRID20/20, Richmond, VA, USA. [Online]. Available: https://www.itron.com/-/media/feature/products/documents/spec-sheet/grid2020_optanode_specsheet.pdf

[37] *Fault Detection in Power Microgrid*. Accessed: Feb. 5, 2022. [Online]. Available: https://github.com/AgHarsh/Fault-Detection-in-Power-Microgrid.git

[38] (2021). *PVWatts Calculator*. National Renewable Energy Laboratory, Golden, CO, USA. [Online]. Available: https://pvwatts.nrel.gov/pvwatts.php

[39] J.-W. Cha and S.-K. Joo, "Probabilistic short-term load forecasting incorporating behind-the-meter (BTM) photovoltaic (PV) generation and battery energy storage systems (BESSs)," *Energies*, vol. 14, no. 21, p. 7067, Oct. 2021.

[40] E. A. Feinberg and D. Genethliou, *Load Forecasting*. Boston, MA, USA: Springer, 2005, pp. 269–285, doi: 10.1007/0-387-23471-3_12.

[41] G. Gong, X. An, N. K. Mahato, S. Sun, S. Chen, and Y. Wen, "Research on short-term load prediction based on Seq2seq model," *Energies*, vol. 12, no. 16, p. 3199, Aug. 2019.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[43] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[44] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
[45] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
[46] *Keras-Team*. Accessed: Mar. 30, 2022. [Online]. Available: https://github.com/fchollet/keras
[47] L. Prechelt, "Early stopping-but when?" in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 1998, pp. 55–69.
[48] M. Mansoor, F. Grimaccia, S. Leva, and M. Mussetta, "Comparison of echo state network and feed-forward neural networks in electrical load forecasting for demand response programs," *Math. Comput. Simul.*, vol. 184, pp. 282–293, Jun. 2021.
[49] P. Jia, H. Zhang, X. Liu, and X. Gong, "Short-term photovoltaic power forecasting based on VMD and ISSA-GRU," *IEEE Access*, vol. 9, pp. 105939–105950, 2021.

AYDIN ZABOLI (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the Babol Noshirvani University of Technology, Babol, Iran, in 2012, and the M.S. degree in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2015. He is currently pursuing the Ph.D. degree in electrical, electronics, and computer engineering with the University of Michigan-Dearborn, Dearborn, MI, USA. From 2015 to 2021, he was a Research Assistant with the Amirkabir University of Technology. His research interests include the security of smart grids and autonomous vehicles, transportation electrification, power system planning, and load forecasting.

VO-NGUYEN TUYET-DOAN received the B.S. degree in statistics from Ton Duc Thang University, Ho Chi Minh City, Vietnam, in 2018. She is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Myongji University, South Korea. Her research interests include statistics, data analysis, pattern recognition, computer vision, image processing, and deep learning techniques.

YONG-HWA KIM (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2001 and 2007, respectively. From 2007 to 2011, he was a Senior Researcher with the Korea Electrotechnology Research Institute (KERI), Gyeonggi-do, South Korea. From 2011 to 2013, he was an Assistant Professor with the Division of Maritime Electronic and Communication Engineering, Mokpo National Maritime University, South Korea. From 2013 to 2021, he was a Professor with the Department of Electronic Engineering, Myongji University, South Korea. Since April 2021, he has been a Faculty Member with the Department of AI Data Engineering, Korea National University of Transportation, South Korea. His research interests include communication systems, digital signal processing, and artificial intelligence for communications, radar systems, and the smart grids.

JUNHO HONG (Senior Member, IEEE) received the Ph.D. degree with Cyber-Security of Substation Automation System in Electrical Engineering from Washington State University, Pullman, in 2014. From 2014 to 2019, he was with ABB, where he provided technical project leadership and supports strategic corporate technology development/productization in the areas related to cyber-physical security for substation, power grids control and protection, renewable integration, and utility communications. He has been working on cyber-security of energy delivery systems with the Department of Energy (DOE), as a Principal Investigator (PI) and a Co-PI in the areas of substation, microgrid, HVDC, FACTS, and high power EV charger. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Michigan-Dearborn. He serves in Cigre WG D2.50 "Electric Power Utilities' cyber-security for contingency operations."

WENCONG SU (Senior Member, IEEE) received the B.S. degree (Hons.) from Clarkson University, Potsdam, NY, USA, in 2008, the M.S. degree from Virginia Tech, Blacksburg, VA, USA, in 2009, and the Ph.D. degree from North Carolina State University, Raleigh, NC, USA, in 2013. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Michigan-Dearborn, USA. He has authored one book, four book chapters, and more than 120 research articles in prestigious international journals and peer-reviewed conference proceedings. His current research interests include power systems, transportation electrification, machine learning applications, and cyber-physical systems. He is a Fellow of IET. He was a recipient of the 2015 IEEE Power and Energy Society (PES) Technical Committee Prize Paper Award and the 2013 IEEE Industrial Electronics Society (IES) Student Best Paper Award. He is an Editor of IEEE TRANSACTIONS ON SMART GRID, IEEE ACCESS, IEEE POWER ENGINEERING LETTERS, and IEEE DATAPORT.