An Improved Non-Schedulable Load Forecasting Strategy for Enhancing the Performance of the Energy Management in a Nearly Zero Energy Building

Nikolaos Jabbour¹, Evangelos Tsioumas¹, Student Member, IEEE, Dimitrios Papagiannis¹, Student Member, IEEE, Markos Koseoglou¹, Student Member, IEEE, and Christos Mademlis¹, Senior Member, IEEE

¹School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece.

Corresponding author: Nikolaos Jabbour (e-mail: njabbour@auth.gr).

This work has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-00399)

ABSTRACT This paper proposes an improved non-schedulable load forecasting (NLF) technique that can be utilized to enhance the performance of the energy management system (EMS) in a nearly zero energy building (nZEB). The suggested NLF is based on a long short-term memory (LSTM) framework in conjunction with a semi-supervised clustering (SSC) technique, considering the most important features which may affect the energy consumption of the non-schedulable appliances (NAs), i.e., number and identity of residents, energy consumption, weather, temperature, humidity, outdoor irradiance, correlation with other loads, day of the week and holidays. The SSC algorithm is utilized to fill the uncompleted information for the residents’ presence in the house and its output constitutes one of the inputs of the LSTM based technique which provides as output a set of forecasting sequences of the NAs’ energy consumption. Unlike the published techniques, the proposed NLF method is not only based on the modeling of the residents’ preferences and habits, but it considers them as variables which affect the nZEB’s microgrid and EMS performance. Therefore, it predicts the residents’ behavior considering its interdependence with the nZEB’s microgrid, which can considerably contribute to the enhancement of the EMS effectiveness and performance. For the implementation of the proposed NLF, no additional hardware is required, but only amendments in the EMS to consider the NLF’s outcomes. To validate the effectiveness of the proposed NLF, selective Hardware-in-the-Loop results from a real nZEB are presented.

Index Terms Forecasting, Energy Management, Buildings.

NOMENCLATURE

ABBREVIATIONS

ARMA Autoregressive Moving Average
BSS Battery Storage System
CAS Controllable Appliances
EMS Energy Management System
HiL Hardware in the Loop
IEMS Integrated Energy Management System
LSTM Long Short Term Memory
MAE Mean Absolute Error
NAs Non Schedulable Appliances
NLF Non Schedulable Load Forecasting
nZEB Nearly Zero Energy Building
PAs Programmable Appliances
RES Renewable Energy Sources
SSC Semi Supervised Clustering
SVM Support Vector Machines

SYMBOLS

\( N \) Number of prediction periods
\( M \) Number of recent past periods
\( E \) Energy consumption vector of non-schedulable appliances
\( PA_i \) \( i \)-th Programmable appliance’s activation/deactivation vector
\( CA_j \) \( j \)-th Controllable appliance’s activation/deactivation vector
\( T \) Time vector of the last \( M \) time steps
\( D \) Vector of the days of the week
\( H \) Vector of holidays
\( PR_r \) \( r \)-th resident’s presence vector for the last \( M \) time steps
\( RT_{rm} \) \( rm \)-th room’s temperature vector for the last \( M \) time steps
\( RH_{rm} \) \( rm \)-th room’s humidity vector for the last \( M \) time steps
\( OT \) Outdoor temperature vector
\( OH \) Outdoor humidity vector
\( SI \) Irradiance vector
\( \mu_k \) \( k \)-th cluster center of SSC k-means
\( J \) Objective function of SSC k-means
\( f_m \) Forget gate of LSTM algorithm
\( i_m \) Input gate of LSTM algorithm
\( o_m \) Output gate of LSTM algorithm
\( g_m \) Updated current status for the cell
\( b_f \) Forget biases of LSTM algorithm
\( b_i \) Input biases of LSTM algorithm
\( b_g \) Input node biases of LSTM algorithm
\( b_o \) Output gate biases of LSTM algorithm
\( h_{m-1} \) Intermediate output of LSTM algorithm
\( s_m \) Current status of LSTM cell
\( s_{m-1} \) Last time frame status of LSTM cell
\( w \) Weight matrices
I. INTRODUCTION

Buildings are responsible for the one-third of the global energy consumption [1]. For this reason, specific measures to increase the energy saving in buildings have been adopted by several authorities, in order to reduce the greenhouse emissions and develop a sustainable, competitive, secure and decarbonized energy system [2], such as the directives 2002/2018 [3] and 844/2018 [4] by the European Union.

It is well known that the key role for the improvement of the efficiency and performance of a nZEB belongs to the EMS, since it is responsible for the proper scheduling of the appliances and the maximum exploitation of the energy generated by renewable energy sources (RES). This can be attained by considering as many as possible variables that may affect the nZEB performance, as well as the residents’ preferences and habits. Therefore, it is apparent that the effectiveness of the EMS can be considerably increased if a forecasting technique is utilized to predict both the nZEB loads’ energy consumption and the RES’s energy production [5]. This is required by several decision making processes that may be embedded in an EMS, such as price forecasting, energy demand scheduling, and maximum exploitation of the RES [6]. The above can be smart meter data analytics that are based on big data and artificial intelligence techniques [7]. Also, since the nZEB is a low scale microgrid, EMS improved performance may be attained by adopting microgrid based techniques [8].

The load forecasting methodologies can be generally categorized into long-term (more than a year), mid-term (from a month to a year), short-term (from one day to a month) and very short-term (less than 24 hours) forecasting methods [9]. Long-term load forecasts can be used for the energy management of power systems and their financial planning [10], while the mid-term load forecasts can be used in the operational planning of the power systems, such as the maintenance scheduling and the control of a hydrothermal system of electric energy generation [11]. The short-term and very-short term forecasting can be used for the improvement of the performance of smart grids and also, as real-time forecasting techniques in order to enhance the prediction accuracy of the mid and long-term estimations [12].

Several computation methods have been proposed in the technical literature to express the non-linear characteristics of the load forecasting. The most popular techniques aim to estimate the load curve by time series analysis with autoregressive moving average models (ARMA) [13] and support vector machines (SVM) based methods [14]. An improved technique for load forecasting through a combination of ARMA based method with non-Gaussian process has been proposed in [15]. Also, a grid search approach to automatically tune the model parameters of SVM and improve the load forecasting has been presented in [16].

The evolution of the artificial intelligence technology have provided effective tools that can improve the accuracy of the building’s loads forecasting such as, physical simulation models [17], [18], statistical analysis [19] and regression models [20]. Also, modern artificial intelligence techniques, such as extreme learning machines [21], artificial neural networks [22] and deep learning neural networks [23], [24], [25] have been adopted in the area of short-term load forecasting. A combination of evolutionary algorithm and artificial neural network to regulate the neural network parameters for improving the learning capability and the accuracy of the short-term load forecasting, has been presented in [26]. Another hybrid technique has been presented in [27], where convolution and LSTM neural networks were combined for short-term forecasting of building’s loads energy consumption. Finally, the combination of convolution and LSTM techniques was also adopted in [28], where the inputs of the forecasting model comprise temperature and humidity datasets, classified by seasons.

The most recent load forecasting algorithms for residential buildings that are published in the technical literature are mainly focused on the building’s total energy consumption curve forecasting. Specifically, a novel hybrid deep learning model for predicting the energy consumption in smart building was proposed in [29] and a hybrid CNN-LSTM model for short-term individual household load forecasting was presented in [30]. Total energy consumption forecasting technique for residential buildings were also proposed in [31]-[33]. However, in these techniques, the interdependence of the residents with the EMS, and the load shifting were ignored at the forecasting process.

The building’s appliances can be categorized in programmable (PAs), controllable (CAs) and non-schedulable (NAS) appliances [34]. The PAs are electric devices that their operating time can be scheduled by the residents (e.g. electric cooker, washing machine, dish washer, etc.) and the CAs are electric loads that their operation is controlled by one or more variables (e.g. heat-pump, air-conditioner, etc. where the variables are the indoor temperature and the fan speed). The NAS are appliances that their operating time cannot be planned, and their activation/deactivation either depends on the residents’ preferences without allowing to be imposed time constraints (e.g. personal computers, TV, etc.) or it is automatically controlled (e.g. lights regulated by a movement control system, refrigerator, water cooler, etc.).

The EMS can increase the energy saving and reduce the electricity expenditure in a nZEB, by properly regulating the operating time intervals of the PAs and the control variables of the CAs, considering several operating factors, such as real-time electricity price, consumed electric energy by each device, user preferences, energy generated by RES, state-of-charge and energy price of the battery storage system (BSS), weather forecast, nZEB’s construction characteristics, etc. [35]. Therefore,
the prediction of the energy consumption of the NAs is quite important information for the EMS of an nZEB to achieve reduction of the energy consumption, proper utilization of the energy generated by building’s RES and effective exploitation of the BSS. In other words, the NAs’ forecasting is a matter that does not solely refer to loads energy consumption, but it is a key factor for the EMS effectiveness and performance.

Due to the stochastic nature of the NAs, their load curve is usually neglected in the EMS for the real-time scheduling of residential appliances, as in [36]. In some other studies, the NAs’ load curve is experimentally determined after measurements of the energy consumption for a specific time period and then, its forecasting is obtained by curve fitting, as in [37] and [38]. However, these techniques do not consider the factors that affect the NAs’ operation and consequently, they cannot effectively predict their dynamic performance that could result to the accurate NAs’ forecasting.

The research methods that have been proposed to predict the NAs’ load more accurately are mainly based on the modeling of the residents’ behavior and habits. Specifically, an attempt to forecast the NAs’ energy consumption based on the resident behavior learning has been presented in [39] and by utilizing a single LSTM neural network for different household loads has been proposed in [40]. The clustering method to group residents with similar consumption behaviors has been adopted in [41] and [42]. In particular, the forecasting method of [41] was based on the load curves profiles using a dynamic time warping technique to cluster the load curve and seek a canonical shape for each set of curves, while the [42] is focused to leverage the comprehensive dataset available from a local smart grid demonstration project and explores the drivers behind the derived groups of electricity customers. Finally, a similar clustering based method with a neural network that employs weather and calendar features, and considers customer behavior similarities has been presented in [43].

From the above it is resulted that the prediction of the NAs energy consumption has been examined in the published technical literature only from the aspect of the residents’ preferences and habits, whereas the interaction with the EMS performance in respect to the time scheduling of the PAs and the control of the CAs has been ignored. Specifically, the published research efforts have been mainly focused on the modeling of the residents’ behavior for forecasting the total NAs load curve without considering the correlation of the NAs with the PAs and CAs operation, which is an important issue for an EMS. Therefore, an integrated and easily applicable method for NAs’ forecasting is required that can address the aforementioned problem from the point of views of both energy load conditions and interaction with the EMS, so as the effectiveness and performance of the EMS can be enhanced.

The scope and the contributions of this research are the following:

- A multi-objective NLF strategy is proposed, that is not only based on the modeling of the residents’ preferences and habits, as in published literature, but it considers them as variables which influence the nZEB’s microgrid operation and EMS performance. Specifically, the proposed NLF strategy can accurately predict the NAs’ energy consumption by properly utilizing the most important features which affect their operation.
- An SSC k-means algorithm is employed to lighten the burden of the residents to constantly inform the system for their presence in the house. The output of SSC k-means algorithm is one of the inputs of the LSTM neural network framework, which acts as the main forecasting technique of the NAs’ energy consumption curve for a specific forecasting period.
- The outcome of the proposed NLF system can be utilized by the EMS of a nZEB for enhancing its performance in the optimal scheduling of the PAs and the proper regulation of the CAs, as well as the maximum utilization of the energy generated by the RES and the optimal exploitation of the BSS capability.
- For the implementation of the NLF system, no additional hardware is required, since the data of the energy meters and sensors that are used for the operation of the EMS are utilized. The only amendment is the replacement of the EMS firmware, so as, the outcome of the NLF is considered in the calculations. Therefore, the proposed NLF system can be installed in any existing or new EMS of a nZEB without affecting the residents’ comfort and without increasing the installation cost.
- The effectiveness of the proposed NLF method and the enhancement of the EMS performance have been validated in a pilot nZEB and several HiL results are presented to verify the reduction of the energy exchange with the grid and the maximum exploitation of the energy generated by RES.

The rest of the paper is organized as follows. The overview of the proposed NLF technique is presented in Section II and the algorithms concerning the proposed forecasting techniques are described in Section III. The testing results and discussions in a HiL system are presented in the Section IV. Finally, the conclusions are summarized in Section V.

II. OVERVIEW OF THE PROPOSED NLF TECHNIQUE

The nZEB’s architecture is based on an EMS that monitors and controls the operation of multiple energy sources and appliances. The control process is realized through a home area network, where the control signals are exchanged either via the internet or a local data network, while the monitoring operation is accomplished through smart meters [44]. The EMS is responsible for the proper scheduling of the PAs operating time and the regulation of the control variables of the CAs, so as, energy peak shaving and reduction of the electricity bill are attained and also, proper control of the energy generated by RES, so as, it can be effectively exploited. Thus, aim of the proposed NLF system is to predict the NAs’ energy demand curve for a specific period ahead considering the EMS operation, so as, to improve its performance through the proper regulation of the PAs, CAs, RES and energy storage in batteries.

Fig. 1 illustrates the schematic layout of a nZEB’s microgrid and the overview of the proposed NLF with an EMS. As can be seen, the nZEB is controlled by a combined system of EMS with NLF that constitute an Integrated EMS (IEMS). The NLF
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3126900, IEEE Access

FIGURE 1. Schematic representation of the microgrid and overview of the proposed NLF in a nZEB’s IEMS (combination of EMS with NLF).

One page of the document contains the following text:

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
c) the short-term LSTM neural network framework, that is the calculation part and provides the prediction of the NAs’ energy consumption for the \( N \) time intervals.

The features and details of the performance of the three parts of the NLF are described below.

### A. Formulation and feature scaling of the input data

The NLF algorithm is initiated with the data preprocessing of the input variables that are historical data of the recent past \( M \) time intervals. They are operating parameters and measurement data collected by the smart sensors and the energy meters, which are already installed in the nZEB, since they are used for the operation of the EMS. They are formulated in vectors with discrete values for each \( \Delta \) time interval, as defined below.

The energy consumption vector of the NAs for the \( M \) last time steps is defined as

\[
E = [e_1 \ e_2 \ ... \ e_M]
\]

and the vectors of the PAs’ and CAs’ activation/deactivation status for the last \( M \) time steps are defined as

\[
PA^t = [PA^t_1 \ PA^t_2 \ ... \ PA^t_M]
\]

\[
CA^t = [CA^t_1 \ CA^t_2 \ ... \ CA^t_M]
\]

where \( t \) is the PA identification number \( t \in \{1 \ 2 \ ... \ N_{PA} \} \), \( N_{PA} \) is the total number of the PAs, \( j \) is the CA identification number \( j \in \{1 \ 2 \ ... \ N_{CA} \} \) and \( N_{CA} \) is the total number of the CAs. The time vector of the last \( M \) time steps that is used as the time base for reporting the appliances’ operation is specified as

\[
T = [t_1 \ t_2 \ ... \ t_M]
\]

and the vector of the days of the week, of the last \( M \) time steps, is given by

\[
D = [d_1 \ d_2 \ ... \ d_M]
\]

where the \( d_1, d_2, ..., d_M \) take values from 0 to 6 for the seven days of the week, e.g. 0 for Sunday, 1 for Monday, etc. Since the potential operation of the appliances may differ depending on whether each referred day is a normal day or holiday, the vector of holidays for the last \( M \) time steps, is defined as

\[
H = [h_1 \ h_2 \ ... \ h_M]
\]

and it is represented in binary mode by strings of 0/1, i.e. 0 for normal days and 1 for holidays.

An important parameter for the NFL framework is the vector of each resident’s presence for the last \( M \) time steps, that is given by

\[
PR' = [PR'_1 \ PR'_2 \ ... \ PR'_M]
\]

where \( r \) is the identification number of the residents. Specifically, \( PR'_1, PR'_2, ..., PR'_M \) take binary values of 0/1, e.g. 0 when the resident is not in the house, 1 when he/she is in the house, or null values in the cases that there is no information for residents’ presence. The null values are filled by the SSC \( k \)-means algorithm, which will be presented in the Subsection III-B.

Finally, the vectors of the rooms’ temperature and humidity for the last \( M \) time steps are considered by the NFL system, since they affect the energy consumption of the CAs, and are defined respectively as

\[
RT'' = [RT''_1 \ RT''_2 \ ... \ RT''_M]
\]

\[
RH'' = [RH''_1 \ RH''_2 \ ... \ RH''_M]
\]

where \( rm \) is the identification number of the rooms. For similar reason with the above, the vectors of the outdoor temperature, humidity and irradiance for the last \( M \) time steps, are given, respectively, by

\[
OT' = [OT'_1 \ OT'_2 \ ... \ OT'_M]
\]

\[
OH' = [OH'_1 \ OH'_2 \ ... \ OH'_M]
\]

\[
SI' = [SI'_1 \ SI'_2 \ ... \ SI'_M]
\]

Since the input data are variables of several physical characteristics, they are scaled in the range of 0/1 using the correct base value for each input variable. Specifically, the \textit{min-max normalization} is adopted for the input variables \( E, RT'', RH'', OT, OH, \) and \( SI' \), while the \textit{one-hot encoding} is applied for the \( PA^t, CA^t, T, D, H, PR' \).

### B. Determining the residents’ presence in the house utilizing the SSC \( k \)-means algorithm

This algorithm aims to alleviate the pressure on the residents to inform constantly and on time their presence in the house. Thus, they can do this later or they may overlook it for several times, since the SSC \( k \)-means algorithm can fill the empties and provide to the LSTM a complete sequence of the residents’ presence. The SSC \( k \)-means algorithm is a semi-supervised machine learning technique, as per [45], and a separate SSC \( k \)-means algorithm is activated for each resident of the building.
Algorithm 1: SSC k-means algorithm for residents’ presence in the house

Input: Data set $X = \{X_1, X_2, ..., X_l\}$, number of clusters $l$ ~ total labeled data.
Output: Resident’s presence array $PR'$.

Method:
1. Initialize: $\mu^*_k$, $k = 1, 2, ..., l$, eq. (15).
2. Each data point $X_i$ is assigned to the cluster center $\mu^*_k$.
3. Each cluster center $\mu^*_k$ is updated by the mean of its constituent points.
4. Repeat 2 and 3 until $k$-means objective function $J$, eq. (14), is optimized.
5. For each cluster $k$ and its data points $X_i$, find the data points $dp_k$ closest to the centroid, with $\arg\min_a X_i$.
6. Match the data points $dp_k$ to the total labeled data (0 indicates that resident is not in the house while 1 indicates that resident is in the house).
7. Define the resident’s presence array $PR'$.

The inputs to the SSC k-means clustering algorithm are a subset of the variables that are defined in the “Formulation & feature scaling” step of the NLF. However, they are reported here as a dataset of $M$ Euclidean variables $X_1, X_2, ..., X_{M}$ of the last $M$ time steps. Each Euclidian variable $X_m$, where $m \in \{1, 2, ..., M\}$, is defined as a $Y$-dimensional vector of the input vectors $E, PA, CA, T, D, H, RT, RH, OT, OH,$ and $SI$ for the exact $m$ time step. Thus, the following formula that gives the Euclidian variable $X_i$ for the $1^\text{st}$ time step is

$$X_i = \left[ e_1 P_{A_i}^1 \cdots P_{A_i}^{N_w} C_{A_i}^1 \cdots C_{A_i}^{N_w} t_1 d_1 \cdots S_1 \right]$$

The SSC k-means algorithm partitions the data set into multiple clusters denoted by $k \in \{1, 2, ..., l\}$, where each cluster is linked with the dataset which denotes that the resident $r$ is in the house or the dataset which denotes that the resident $r$ is not in the house. For each resident $r$ and data point $X_m$, a set of indicator variables $R'_{r,m} \in \{0, 1\}$ is used to describe the assignment of the data point to the specific cluster, that are formulated in vectors

$$R'_{r,m} = \left[ R'_{r,1,m} \ R'_{r,2,m} \cdots R'_{r,l,m} \right]$$

For example, if the $X_m$ is assigned to cluster 2, then $R'_{r,2,m}$ takes the value 1. Also, for each resident $r$, a set of $Y$ dimensional vectors $\mu^*_r$ is introduced to represent the center of the $k$ cluster. According to the confirmed residents’ labeled data and the EMS labeled data, the cluster’s data point $dp_k$ that is closest to the respective cluster’s center is linked with the dataset which denotes that the resident $r$ is in the house, or the dataset which denotes that the resident $r$ is not in the house. Thus, the $PR'$ for each resident of (7) is filled with the appropriate binary values. The number of the clusters $l$ is proportional to the number of the total labeled data, which is the sum of residents’ and EMS labeled data.

The goal of the SSC k-means clustering algorithm is to find, for each $r$ resident, the exact values for the $R'_{r,m}$ and $\mu^*_r$ that minimize the objective function

$$J' = \sum_{m=1}^{M} \sum_{k=1}^{l} R'_{r,m} \|X_m - \mu^*_r\|^2$$

The $J$ function can be considered as a distortion measure and it can be minimized by applying an iterative procedure for the optimization variables $R'_{r,m}$ and $\mu^*_r$, considering that $R'_{r,m}$ takes the value 1 for whichever variable $k$ minimizes the $\|X_m - \mu^*_r\|^2$, while the $\mu^*_r$ is selected by

$$\mu^*_r = \frac{\sum R'_{r,m} X_m}{\sum R'_{r,m}}$$

The proposed SSC k-means algorithm is described in a pseudocode form in Algorithm 1. As aforementioned, the SSC algorithm is utilized to fill the uncompleted information for the residents’ presence in the house. Thus, if the data provided by the residents are complete and reliable, the SSC is bypassed, and they are directly provided to the LSTM. It should be noted that the information provided by the residents, whenever it is done, improves the accuracy of the SSC algorithm through the updated input data vectors.

The data that the residents provide to the SSC k-means algorithm to inform the proposed algorithm for their presence in the house are defined as residents’ labeled data. However, since the proposed algorithm may be integrated in the EMS of a nZEB, some of the PAs or/and CAs may indirectly provide information for the presence of the residents. For example, the oven, the vacuum cleaner, and the dishwasher usually are not activated without at least one resident’s presence in the house. Thus, the activation of one of these appliances can be considered as a signal of the resident’s presence in the house. These data are defined as EMS based labeled data. Moreover, since the proposed NLF algorithm considers the EMS based labeled data, the required residents’ labeled data may be further decreased. Therefore, by properly considering the information that the PAs, CAs and EMS provides, satisfactory accuracy in the outcome of the SSC k-means algorithm can be attained with reduced requirement by the residents to constantly inform for their presence in the house.

The accuracy of the SSC k-means algorithm in completing the potential lack of information for the residents’ presence in the house with respect to the percentage of the confirmed resident’s labeled data, is validated in a case study that is illustrated in Fig. 3. To evaluate the importance of the PAs and CAs as

**TABLE 1**

| MODEL OF THE EXAMINED NZEB |
|-----------------------------|
| **Area** | 120 m² |
| **City** | Polytakra, Greece |
| **PV nominal power** | 2.7 kW |
| **WT nominal power** | 2 kW |
| **BSS energy capacity** | 5 kWh |
| **PAs** | Dishwasher, oven, and washing machine |
| **CAs** | Heat pump and air conditioner |
of the proposed algorithm, it is compared the accuracy of the algorithm for the cases that the inputs of the PAs, CAs and EMS labeled data are considered or not. In the examined case study, the parameter $M$ is equal to 96 and an electric oven, a washing machine and a dishwasher are considered as PAs, while the heat pump and the air conditioner are considered as CAs. The model of the examined nZEB can be seen in Table I.

As can be seen in Fig. 3, the accuracy of the algorithm is increased as the percentage of the confirmed resident’s labeled data is increased, as it may expected. However, since there is high correlation between the operation of the PAs and CAs with the resident’s presence in the house, the same accuracy of the SSC $k$-means algorithm can be attained with reduced confirmed resident’s data when the PAs and CAs activations/deactivations are provided as inputs. Specifically, accuracy of 93% is accomplished for confirmed resident’s data of 14%, for the case that the PAs and CAs are not provided as inputs (blue curve), while the same accuracy is attained with confirmed resident’s data of 6%, if the PAs and CAs are given as inputs (dotted red curve). Moreover, further reduction in the confirmed resident’s data can be achieved if, in addition to the PAs and CAs as inputs, the labeled EMS data are given as inputs to the SSC $k$-means algorithm. As can be observed, the same accuracy as above is achieved with confirmed resident’s data of only 2.8%, if both the PAs-CAs and EMS labeled data are given as inputs to the algorithm (black curve). Therefore, not only less residents’ data are required in the latter case, but also, their comfort is less affected, or on the other hand, higher accuracy is attained with the same percentage of confirmed resident’s labeled data.

It should be noted that a minimum percentage of the confirmed resident’s labeled data is required, so as the accuracy of the outcome of the SSC $k$-means algorithm is satisfactory. This depends on the resident’s lifestyle pattern variations and a typical value may range between 2.5% to 3.5% for a mean resident.
The contribution of the SSC k-means algorithm on the accuracy of the proposed NLF system is summarized at the Table II that compares the mean absolute error (MAE) of the NAs’ load forecasting in kWh for the cases of the LSTM with and without the SSC k-means algorithm. As can be observed, the MAE is 0.37 kWh for the case of the LSTM without the SSC k-means algorithm; however, it is reduced to 0.31 when the SSC k-means algorithm is used.

| TABLE II | NAS’ LOAD FORECASTING MAE SUMMARY (30-DAYS) |
|----------|------------------------------------------|
| Technique | MAE (kWh) |
| LSTM with SSC k-means algorithm | 0.31 |
| LSTM without SSC k-means algorithm | 0.37 |

Note that the SSC k-means algorithm can always fill the uncompleted information for the residents’ presence in the house, even the case that a minimum percentage of confirmed residents’ labeled data is not provided, but with the penalty of reduced accuracy. Also, the LSTM can give prediction outcome for any values are chosen, for the number and the duration of the time intervals of the prediction and the recent past periods. Therefore, the operation of the NLF method is stable and the sensitivity and consequently the accuracy of the obtained results depend on the fulfillment of the aforementioned limitations.

IV. HIL RESULTS

The model of the examined nZEB microgrid is housed in a dSPACE DS1104 PCI controller board and the IEMS (EMS with the proposed NLF) runs in a PC with an i7 8700k processor and an RTX 2070 Super as GPU. The communication between the HiL plant and the PC is conducted by digital and analog signals that command the BSS and the appliances, while the measurements are obtained by utilizing the RS232 serial interface.

The examined microgrid plant is a pilot nZEB of area 120m² in Polykastro, Greece, which is inhabited by two adults and a child. The RES of the nZEB consists of PVs and a WT of nominal power 2.7 kW and 2 kW, respectively (Table I). The data were collected for a period of 1-year (from 1-March-2019 to 1-March-2020) and the sampling time was Δ = 30min. The BSS consists of a Li-ion battery pack of energy capacity 5 kWh. Three PAs are considered, i.e. a dishwasher, an oven, and a washing machine.

The EMS control strategy of [34] has been adopted in this paper that aims to the proper scheduling of the PAs and control of the CAs, as well as the optimal management of the energy generated by the RES by properly selecting to self-consume it within the nZEB, inject it to the grid, or temporary store it in batteries. The above tasks of the EMS are performed by considering the forecast of the NAs’ load curve that is estimated by the proposed NLF system.

The data are split into three subsets. The first subset is used for the training of the forecasting models of the proposed NLF, the second subset (which is the validation subset) is used to evaluate the performance of the different models, and finally, the third subset (which is the test subset) is used for the evaluation of the results. According to the consistent finding in [46] for LSTM algorithms, multiple layers always work better than a single layer, and the number of hidden nodes should be sufficiently large. Thus, it has been decided to implement the LSTM...
algorithm with two hidden layers and 200 hidden nodes for each layer.

To demonstrate the effectiveness of the NLF and demonstrate the operating improvements of the IEMS, three scenarios are illustrated, for one day that was chosen arbitrarily on a period of 1 year. For the 1-year considered period and for all the examined scenarios, the residents have directly provided to the system information for their presence in the house at an average percentage of 6% of the total cases (10% total labeled data and 4% EMS’s labeled data). Specifically, the performance of the IEMS (combined EMS and NLF) on the tested nZEB is illustrated in Fig. 7 that can be compared with the ideal case where the EMS runs knowing a priori the real energy consumption of the NAs (Fig. 8) and the case that the EMS operates considering a predefined fixed curve for the NAs’ energy consumption (Fig. 9). Note that the case of Fig. 8 is not realistic, but it can be used as a reference to evaluate the IEMS performance of Fig. 7. Also, Fig. 7 is compared with the case that is usually applied in the literature, where the NAs’ loads curve is considered constant (Fig. 9) to validate the advantages of the NLF. For all the above examined cases, the energy generated by the RES is the same, in order to have a common comparison basis (1st diagram of all figures). Thus, different energy transaction with the grid and the BSS the SoC performance of the Fig. 7 are very close to the ideal case of Fig. 8 (6th to 7th diagrams). The energy exchanged with the grid is slightly higher with the NLF compared to the ideal case, due to some discrepancies in NA’s forecast (8th diagrams).

As can be seen, the PAs operate in different time intervals (4th diagram of all figures), since the EMS provides different time scheduling for each examined case. The same also occurs for the power provided/absorbed by the grid (5th diagram of all figures) and the power and the SoC of the BSS (6th and 7th diagrams of all figures). Thus, different energy transaction with the grid is observed for each of the examined cases that is illustrated in the 8th diagram of all figures and it is the sum of the absolute values of the energy consumed and injected into the grid, $E_{\text{Grid}}$. In the 5th diagrams, the positive values of the power $P_{\text{Grid}}$ are referred to energy consumed by the grid, while the negative values correspond to energy supplied to the grid. In the 6th diagrams, the positive values of the power $P_{\text{RES}}$ are referred to energy recovered by the batteries while the negative values correspond to energy stored to the batteries.

In Fig. 7, comparing the two curves of the 3rd diagram, it is concluded that the NLF system can successfully predict the NAs’ energy consumption, since the blue curve which corresponds to the predicted NAs’ energy consumption is close to the real NAs’ energy load (red curve). Also, comparing the EMS performance of the Fig. 7 with that of the Fig. 8 which corresponds to the reference ideal case, it is resulted that satisfactory performance of both the energy management of the RES and BSS is accomplished, since both the total energy traded with the grid and the BSS the SoC performance of the Fig. 7 are very close to the ideal case of Fig. 8 (6th to 7th diagrams). The energy exchanged with the grid is slightly higher with the NLF compared to the ideal case, due to some discrepancies in NA’s forecast (8th diagrams).

The advantages of the proposed NLF can be revealed by comparing the Fig. 7 (proposed NLF) with the Fig. 9 (method that is usually applied in the literature, where the NAs’ loads curve is considered constant). As can be observed in the 2nd diagram of Fig. 9, although the power load of the nZEB is low at the time interval between 10:00 to 11:30, the SoC of the BSS has reached the maximum value of 80% (7th diagram) and consequently the only option is to provide the excess energy generated by the RES to the grid. Contrarily, the SoC of BSS in the
The case of Fig. 7 has reached the maximum value for a much smaller period and therefore, less energy is traded with the grid. Therefore, enhanced energy management in the BSS is attained with the proposed NLF (Fig. 7) compared with the case where a fixed NAs’ load curve is considered (Fig. 9). Specifically, the energy exchanged with the grid at the end of the examined day is higher by 5kWh for the case of Fig. 9 compared with the case of the proposed NLF (Fig. 7) and 7kWh compared with the ideal case of Fig. 8 (please compare the 8th diagrams of the above figures). Considering that the price of selling the electric energy to the grid is lower than the price of buying energy from the grid, the economic impact is lower in the case of the proposed NLF technique, compared to the case of fixed load curve.

The above conclusion has also been validated for a short period of one month (Fig. 10) and a long time period of six months (Fig. 11). As can be seen in Fig. 10, the energy exchanged with the grid for the proposed NLF load forecasting technique is 501kWh, whereas it has been increased to 559kWh for the case of the NAs’ fixed load curve scenario. Similar results are obtained for the longer period of six months (Fig. 11). In this case, the energy exchange with the grid of the proposed IEMS is 4,126 kWh, while it has been increased to 4,798 kWh for the case of the NAs’ fixed load curve. Therefore, due to the higher buying energy price against the selling energy price to the grid, the proposed NLF system provides economic benefits for the building that corresponds to the lower energy exchange with the grid of 58kWh. Moreover, since the increase of energy exchange between building and electric grid increases the energy losses, the proposed NLF technique results to the increase in the efficiency of the nZEB’s microgrid. Finally, the advantage of the proposed NLF technique in reducing the energy exchange with the grid, contributes to the enhancement of the stability and reliability of electric power system. The energy exchange with the grid for the three examined case studies of very short, short and long period of time are reported in Table III.

From the above it is concluded that the NLF can contribute to the improvement of the EMS with respect to both the energy transaction with the grid and the proper control of the BSS for the maximum exploitation of the energy generated by the RES with several economic and technical benefits.
V. CONCLUSIONS

In this paper, an improved NLF system for accurate forecasting of the NAs’ energy consumption is proposed, that is based on a LSTM recurrent neural network framework in conjunction with a semi-supervised clustering technique. The LSTM is the main technique for the NAs’ forecasting, while the SSC $k$-means is an auxiliary algorithm that aims to alleviate the pressure on the residents to constantly inform the system for their presence in the house. Through the proposed SSC $k$-means algorithm, not only less confirmed residents’ labeled data are required, but residents’ comfort is less affected. Also, the proposed LSTM technique with the SSC $k$-means achieves lower mean absolute error compared with the case of the LSTM without SSC $k$-means. Since, the proposed NLF technique can be utilized by the EMS of a nZEB, the stability and the reliability of the building microgrid can be enhanced by reducing the energy exchange with the grid, as well as the proper control of the BSS, attaining maximum exploitation of the energy generated by the RES with several economic and technical benefits. For the implementation of the NLF, no extra energy meters and sensors or any additional hardware are required, but only amendments in the EMS firmware so as, the outcome of the NLF is considered in the calculations. The feasibility and effectiveness of the proposed NLF in the EMS's performance has been validated in a HiL system of a pilot nZEB and selective results are presented to demonstrate the operating improvements.

As a future work is the extension of the proposed methodology on a master/slave topology for application in block of nZEBs and large area microgrids. Specifically, a master algorithm will control the outcomes of the IEMSs that act as slave algorithms, in order to maximize the exploitation of the RES and BSS of the block of nZEBs and the microgrid.

ACKNOWLEDGMENT

This work has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-00399).

REFERENCES

[1] H. C. Akdag and T. Beldek, "Waste management in green building operations," *Int. J. Supply Chain Manage.*, vol. 6, no. 3, pp. 174-180, 2017.

[2] F. Spertino and et.al., "Toward the Complete Self-Sufficiency of an nZEBs Microgrid by Photovoltaic Generators and Heat Pumps: Methods and Applications," *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 7028 - 7040, 2019.

[3] "EU Directive 2018/2002 of 11 December 2018, amending directive 2012/27/EU on energy efficiency, [Online], Available at:https://eur-lex.europa.eu/eli/dir/2018/2002/oj".
[4] "Towards nearly zero-energy buildings Definition of common principles," European Commission, https://ec.europa.eu/energy/sites/ener/files/documents/nzeb_full_report.pdf, 2013.

[5] B. Naghibi, M. A. Masoum and A. D. Deilami, "Effects of V2H Integration on Optimal Sizing of Renewable Resources in Smart Home Based on Monte Carlo Simulations," IEEE Power and Energy Technology Systems Journal, vol. 5, no. 3, pp. 73-84, 2008.

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," Int. J. Forecasting, vol. 32, no. 3, p. 914-938, 2016.

[7] Y. Yang, W. Li, T. A. Gulliver and S. Li, "Bayesian Deep Learning-Based Probabilistic Load Forecasting in Smart Grids," IEEE Trans. on Ind. Inform., vol. 16, no. 7, pp. 4703 - 4713, 2020.

[8] T. Khalili, S. Nojavan and K. Zare, "Optimal performance of microgrid in the presence of demand response exchange: A stochastic multi-objective model," Computers & Electrical Engineering, vol. 74, pp. 429-450, 2019.

[9] K. Yan, L. Wei, J. Zhiwei, Q. Meng and D. Yang, "A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households," IEEE Access, vol. 7, pp. 157633-157642, 2019.

[10] J. Xie, T. Hong and J. Stroud, "Long-Term Retail Energy Forecasting With Consideration of Residential Customer Attrition," IEEE Trans. on Smart Grid, vol. 6, no. 5, pp. 2245 - 2252, 2015.

[11] E. Gonzalez-Romera, M. Jaramillo-Moran and D. Carmona-Fernandez, "Monthly electric energy demand forecasting based on trend extraction," IEEE Trans Power Syst., vol. 21, no. 4, pp. 1946 - 1953, 2016.

[12] K. Liu, S. Subbarayan, R. R. Shoultz, M. T. Manry, F. I. L. C. Kwan and J. Naccarino, "Comparison of very short-term load forecasting techniques," IEEE Trans. Power Syst., vol. 11, no. 2, pp. 877-882, 1996.

[13] C. Dinesh, S. Makonin and I. V. Bajić, "Residential Power Forecasting Using Load Identification and Graph Spectral Clustering," IEEE Trans. on Circuits and Systems II: Express Briefs, vol. 66, no. 11, pp. 1900 - 1904, 2019.

[14] Y.-M. Wi, S.-K. Joo and K.-B. Song, "Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment," IEEE Trans. Power Syst., vol. 27, no. 2, 2012.

[15] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," IEEE Trans. on Power Systems, vol. 18, no. 2, pp. 673 - 679, 2003.

[16] Y. Yang, J. Che, C. Deng and L. Li, "Sequential grid approach based support vector regression for short-term electric load forecasting," Appl. Energy, vol. 238, pp. 1010-1021, 2019.

[17] K. J. Beven, M. J. Kirkby, N. Schofield and A. F. Tagg, "Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments," J. Hydrol., vol. 69, pp. 119-144, 1984.

[18] A. H. Neto and F. A. S. Fiorelli, "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption," Energy Buildings, vol. 40, no. 12, 2008.

[19] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," IEEE Trans. Power Syst., vol. 16, no. 3, pp. 498-505, 2001.

[20] Y.-S. Lee and L.-I. Tong, "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming," Knowl.-Based Syst., vol. 24, no. 1, pp. 66-72, 2011.

[21] M. Rafei, T. Niknam, J. Aghaei, M. Shafie-Khah and J. P. S. Catalão, "Probabilistic Load Forecasting Using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine," IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 6961 - 6971, 2018.

[22] T. K. Chau, S. S. Yu, T. Fernandez, H. H.-C. Lu and M. Small, "A Load-Forecasting-Based Adaptive Parameter Optimization Strategy of STATCOM Using ANNs for Enhancement of LFOD in Power Systems," IEEE Trans. Ind. Inf., vol. 4, no. 6, pp. 2463 - 2472, 2018.

[23] T. Ouyang, Y. He, H. Li, Z. Sun and S. Baek, "Modeling and Forecasting Short-Term Power Load With Copula Model and Deep Belief Network," IEEE Trans.Emerg. Top. in Comp. Intel., vol. 3, no. 2, pp. 127 - 136, 2019.

[24] 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.

[25] K. Auranzegb, M. Alhusseins, K. Javaid and S. I. Haider, "A Pyramid-CNN Based Deep Learning Model for Power Load Forecasting of Similar-Profile Energy Customers Based on Clustering," IEEE Access, vol. 9, pp. 14992 - 15003, 2021.

[26] P. Singh and P. Dwivedi, "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," Appl. Energy, vol. 217, pp. 537-549, 2018.

[27] K. Yan, X. Wang, Y. Du, N. Jin, H. Huang and H. Zhou, "Multi-step short-term power consumption forecasting with a hybrid deep learning strategy," Energies, vol. 11, no. 11, p. 3089, 2018.

[28] M. Kim, W. Choi, Y. Jeon and L. Liu, "A hybrid neural network model for power demand forecasting," Energies, vol. 12, no. 2, p. 931, 2019.

[29] D. Syed, H. Abu-Rub, A. Ghraeb and S. S. Refaat, "Household-Level Energy Forecasting in Smart Buildings Using a Novel Hybrid Deep Learning Model," IEEE Access , vol. 9, pp. 33498 - 33511, 2021.

[30] M. Alhusseins, K. Auranzegb and S. I. Haider, "Hybrid CNN-LSTM Model for Short- Term Individual Household Load Forecasting," IEEE Access , vol. 8, pp. 180544 - 180557, 2020.

[31] 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.

[32] J. Han, L. Yan and Z. Li, "A Task-based Day-ahead Load Forecasting Model for Stochastic Economic Dispatch," IEEE Transactions on Pow. Syst., vol. Early Access, pp. 1-1, 2021.

[33] J. Wang, X. Chen, F. Zhang, F. Chen and Y. Xin, "Building Load Forecasting Using Deep Neural Network with Efficient Feature Fusion," Journal of Modern Power Systems and Clean Energy, vol. 9, no. 1, pp. 160 - 169, 2021.

[34] E. Tsioumas, N. Jabbour, M. Koseoglou and C. Mademlis, "A
Novel Control Strategy for Improving the Performance of a Nearly Zero Energy Building,” IEEE Trans. Pow. Electr., vol. 35, no. 2, pp. 1513 - 1524, 2020.

[35] Z. Zhao, W. C. Lee, Y. Shin and K.-B. Song, "An optimal power scheduling method for demand response in home energy management system," IEEE Trans. Smart Grid, vol. 4, no. 2, p. 1391–1400, 2013.

[36] Z. Wu, S. Zhou, J. Li and X.-P. Zhang, "Real-Time Scheduling of Residential Appliances via Conditional Risk-at-Value," IEEE Trans. Smart Grid, vol. 5, no. 3, pp. 1282 - 1291, 2014.

[37] I.-Y. Joo and D.-H. Choi, "Optimal household appliance scheduling considering consumer's electricity bill target," IEEE Trans. Cons. Electr., vol. 63, no. 1, pp. 19 - 27, 2017.

[38] E. Proedrou, "A Comprehensive Review of Residential Electricity Load Profile Models," IEEE Access, vol. 9, pp. 12114 - 12133, 2021.

[39] 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, 2018.

[40] 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, 2017.

[41] T. Teeraratkul, D. O'Neil and S. L., "Shape-based approach to household electric load curve clustering and prediction," IEEE Trans. Smart Grid, vol. 9, no. 5, pp. 5196-5206, 2017.

[42] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar and M. E. Webber, "Clustering analysis of residential electricity demand profiles," Appl. Energy, vol. 135, pp. 461-471, 2014.

[43] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," IEEE Trans. Smart Grid, vol. 6, no. 2, pp. 911-918, 2015.

[44] L. Wang, Z. Wang and R. Yang, "Intelligent Multiagent Control System for Energy and Comfort Management in Smart and Sustainable Buildings," IEEE Trans. on Smart Grid, vol. 3, no. 2, pp. 605-617, 2012.

[45] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[46] G. H. e. al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82-97, 2012.

Nikolaos Jabbour was born in Thessaloniki, Greece, on December 4, 1988. He received the Dipl. Eng. and Ph.D. degrees in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece, in 2012 and 2018, respectively. Since 2018, he has been with the Electrical Machines Laboratory, School of Electrical and Computer Engineering, Aristotle University of Thessaloniki as Postdoctoral Researcher. His research interests include control optimization, electrical machines and drives, power electronics, embedded systems, renewable energy systems, smart grids and energy management in nearly zero energy buildings.

Evangelos Tsioumas was born in Komotini, Greece, on June 13, 1988. He received the Diploma degree from the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, in 2011, where he is currently working toward the Ph.D. degree in the area of energy management optimization in microgrids.

His research interests include energy management in power systems, control optimization in microgrids, large-scale electric energy storage systems, and embedded systems design.

Dimitrios Papagiannis was born in Krana Elassonas, Greece, on December 24, 1990. He received the Diploma degree from the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, in 2015, where he is currently working toward the Ph.D. degree in the area of optimization of vehicles’ active suspension systems.

His research interests are in the areas of performance improvement of active suspension systems in vehicles, vehicle dynamics, electric motor drives, power electronics and embedded systems design.

Markos Kosoglou was born in Thessaloniki, Greece, on July 11, 1992. He received the Diploma degree from the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, in 2018, where he is currently working toward the Ph.D. degree in the area of battery management system optimization.

His research interests include energy management in power systems, battery management systems, and embedded systems design.

Christos Mademlis (S’96–A’00–M’03–SM’11) was born in Arnea Chalkidikis, Greece, on February 7, 1964. He received the Diploma degree in Electrical Engineering (1st class Hons.) and the Ph.D. degree in electrical machines from the Aristotle University of Thessaloniki, Greece, in 1987 and 1997, respectively.

Since 1990, he has been with the Electrical Machines Laboratory, Faculty of Electrical and Computer Engineering, Aristotle University of Thessaloniki as Research Associate (1990-2001), Lecturer (2001-2006), Assistant Professor (2007–2014), Associate Professor (2014-2019) and currently Professor (since 2019). In 2013-2016, he has been the Head of the Department of Electrical Energy and since 2010, he has been the Director of the Electrical Machine Laboratory, Faculty of Engineering of the Aristotle University of Thessaloniki. He is author and coauthor of more than 75 peer-reviewed technical papers. He was the founder (2013), and for the period 2013-2017 the Faculty Advisor, of the Student Formula SAE racing team of the Aristotle University of Thessaloniki. In 2017, he was awarded a Fulbright Grant as a Visiting Professor at the APED, Electrical and Computer Engineering Dept., University of Connecticut, USA.

His research interests include the areas of electrical machines and drives, especially in design and control optimization, renewable energy sources, electric vehicles, energy saving systems and energy management in nearly zero energy buildings.