IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 13, 2020, accepted June 4, 2020, date of publication June 12, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002155

A Machine Learning Framework to Infer Time-of-Use of Flexible Loads: Resident Behavior Learning for Demand Response

MILAD AFZALAN, (Graduate Student Member, IEEE), AND FARROKH JAZIZADEH
The Charles E. Via, Jr. Department of Civil & Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA

Corresponding author: Farrokh Jazizadeh (jazizade@vt.edu)

This work was supported in part by Virginia Tech's Open Access Subvention Fund (VT OASF).

ABSTRACT Load shapes obtained from smart meter data are commonly utilized to understand daily energy use patterns for adaptive operations in applications such as Demand Response (DR). However, they do not provide information on the underlying causes of specific energy use patterns – i.e., inference on appliances' time-of-use (ToU) as actionable information. In this paper, we investigated a scalable machine learning framework to infer the appliances' ToU from energy load shapes in a collection of residential buildings. A scalable and generalized inference model obviates the need for model training in each building to facilitate its adoption by relying on training data from a set of previously observed buildings with available appliance-level data. To this end, we demonstrated the feasibility of using load shape segmentation to boost ToU inference in buildings by learning from their nearest matches that share similar energy use patterns. To infer an appliance ToU for a building, classification models are trained for inference on subintervals of load shapes from matched buildings with known ToU. The framework was evaluated using real-world energy data from Pecan Street Dataport. The results for a case study on electric vehicles (EV) and dryers showed promising performance by using 15-min smart meter load shape data with 83% and 71% F-score values, respectively, and without in-situ training.

INDEX TERMS Demand response, smart meter, distributed energy resources, segmentation, machine learning, time-of-use, non intrusive load monitoring (NILM).

I. INTRODUCTION

In recent years, conventional centralized power systems are shifting to decentralized alternatives that integrate distributed energy resources (DER) such as solar panels, district resources, storage systems, and advanced technologies for smart metering and control. These changes provide opportunities and call for efficient adaptive and responsive operations – e.g., utilization of Demand Response (DR) programs for load balancing at the neighborhood level. However, the successful implementation of adaptive operations in the residential sector requires a sound understanding of energy usage patterns such as load shapes – i.e., the variation of power demand over the span of a day. Detailed analysis of the usage patterns reveals temporal drivers of demand, which in turn enables efficient targeting of customers for customized energy programs and demand control automation [1]–[3]. Building energy load shapes are commonly characterized through data-driven segmentation methodologies by clustering smart meter data to help engage consumers for DR programs [4], [5]. However, understanding the drivers of demand variations through an in-depth analysis of the human-building interactions (HBI) at the appliance (i.e., individual load) level will improve the efficacy of managing loads for demand-supply balance as shown in previous research [1], [6].

HBI assessments center on identifying the patterns of using different flexible loads (e.g., electric vehicles) in a building. These assessments provide a statistical measure to evaluate the benefits of engaging end-users in adaptive management of loads (e.g., engagement in DR programs). By measuring the time-of-use (ToU) of flexible loads and other driving metrics such as frequency and consistency of use, operations could be moved toward intelligent and distributed load operation/scheduling with a reduced burden for end-users [7]. In other words, information from HBI patterns,

The associate editor coordinating the review of this manuscript and approving it for publication was Chin-Feng Lai.

as supplementary information to the load shapes at the aggregate level, helps manage power systems more efficiently by engaging a subset of consumers that will result in a higher gain in efficient operations. This could be achieved by using appliance-level energy use data and data-driven solutions that characterize energy consumption styles of users [1]. This information, in turn, is used to identify the subset of users to more efficiently balance the demand at peak and off-peak time and mitigate the potential rebound impact [1], [8]. Furthermore, by leveraging smart meter data, machine learning (ML) tools have been used to characterize different energy usage patterns in residential buildings and neighborhoods and to infer variations in energy lifestyles (e.g., [9]–[11]). However, leveraging building energy load shapes has not been previously suggested for the appliance ToU inference.

Determining the daily ToU of appliances can be carried out by using individual plug sensors at each device. Such an approach provides accurate measurements but calls for the distributed installation of sensors, which is intrusive and might be prohibitively complicated and expensive. On the other hand, appliances' ToU can be implicitly inferred through inference models using pattern recognition algorithms from the power data at the aggregate (i.e., whole building) level. However, an inference model suited to a specific building calls for *a priori* assumption about appliance characteristics, high-resolution data, or in-situ training, which might be an obstacle for a scalable approach. Therefore, enabling a generalized and data-driven inference approach that does not rely on specialized instrumentation and in-situ algorithm training, or specific assumptions on an individual building, could facilitate scalable adaptive operations such as DR or DER management. Research has shown that the dynamic patterns of energy consumption, driven by residents' interactional behavior, are correlated with appliance ToU [12], [13].

In this study, we have investigated ToU inference models that leverage neighbor (i.e., previously observed) buildings with similar energy use behavior for training an energy behavioral learning framework. To this end, we have introduced a machine learning framework for inference of appliance ToU in buildings by accounting for resident behavior reflected in their energy load shapes from smart meter data. The contributions of this work are as follows:

- Propose a scalable approach for ToU inference of flexible loads by relying on only low-resolution smart meter data – i.e., 15-min resolution, which resembles the sampling rate in practice.
- Enable a data-driven learning framework for ToU inference to identify the similarities between buildings by leveraging the aggregate energy load shapes. In contrast to methods that require specific instrumentation at the building level for in-situ training or *a priori* information on appliance models of a target building, we leverage the information in energy load shapes of a sample of buildings whose appliance ToU data is known for the purpose of training. Therefore, the framework does not require the appliance-level data in the target buildings – i.e., buildings that their appliance ToU are being predicted.

The rest of this paper is structured as follows. In section II, the related literature and research background, including the research on energy disaggregation, have been presented. In Section III, the proposed inference framework has been presented. In Section IV, the framework evaluation through a case study, as well as the results and discussions are presented. Section V includes the concluding remarks.

II. RELATED WORKS

ToU inference in buildings is becoming increasingly important for mitigating the stress on the grid and supporting distributed energy management. To this end, several studies have explored the underlying causes of variation in appliance usage profiles at different times of the day from the user interaction perspective [1], [14], [15]. The study by Cetin *et al.* [14] investigates the energy use of appliance profiles including dryers, washing machines, and refrigerators in 40 buildings to understand the variation at different times of the day. Statistical results showed electricity use varies more amongst buildings during the peak time of the day compared to off-peak time. The variations in appliance ToU amongst buildings have been studied to estimate cost saving under dynamic pricing schemes [15] and to quantify energy saving and load flexibility for different segments of users [1]. The results have shown highly variable energy demands and high load flexibility potential at different times of the day. By leveraging such information, it has been shown that associating appliance ToU with load shape analysis and segmentation in the residential sector could lead to improved implementation of DR programs [16], [17].

To infer the appliance ToU, studies have proposed different predictive models by learning from the variations on aggregate power time series as their data source [18]–[21]. Basu *et al.* [18] investigated the applicability of several classifiers, such as decision trees and Bayes networks, for ToU inference by using historical consumption information and human expert knowledge as the basis for in-situ training. Various studies have looked into the inference of appliance usage by employing different deep learning architectures including long short-term memory (LSTM) [20], denoising autoencoder [22], and convolutional neural network (CNN) [23]. Barsim and Yang [21] developed a convolutional neural network architecture to extract individual profiles of appliances and evaluated it on five residential buildings. Kelly and Knottenbelt [20] investigated the applicability of three deep learning architectures including LSTM and denoising autoencoders and investigated their performance for detecting the use of major appliances including washing machines and dishwashers for five buildings. Liu *et al.* [24] used the concept of transfer learning and leveraged a pre-trained deep learning model on an image recognition dataset for appliance classification. As a common feature, the applicability of such methods have been investigated on datasets with high-resolution data (e.g., with sampling rates of 1 Hz or higher – example relevant datasets with high-resolution data could be found in [25]), or alternative metrics such as voltage-current trajectory have been employed instead of real power measurement, which is commonly recorded by smart meters.

As an alternative to direct appliance ToU inference, classical energy breakdown (disaggregation) methodologies could be employed to extract the ToU of appliances. Through aggregate load event identification or reconstructing the time-series of individual appliances from energy breakdown and comparing thresholds with the values on the resultant time-series, ToU at different times of the day can be extracted. Energy breakdown methodologies have been widely studied – references [26]–[28] are well-known representative surveys in this field. A wide range of parameters, important in driving the outcome, has been used across studies: (1) varying levels of sampling rate in acquiring aggregate power data have been used, including high-resolution data (@60Hz) [29]–[32] and low-resolution (sampling every hour) data [33], (2) varying number of buildings for evaluation, including individual buildings [34], a few buildings (3 to 6) [22], [35], or 10 and more buildings [36], [37], and (3) different methods of training, including supervised methods that may require extensive parameter search [38], unsupervised methods, capable of automated parameter tuning [30], [32], [35], or Hidden-Markov models [39], [40]. The scalability potential of these solutions varies from one to another given their specific design and implementation [41]. A scalable solution for ToU inference ideally accounts for the following factors: it avoids in-situ training – learning from the same environment that is the subject of prediction with considerable model parameter tuning effort, and uses low-resolution (real-power) data that can be acquired with ubiquitous metering infrastructures such as smart meters and their default configurations.

Since this study has been motivated by DER integration and automated DR, we looked at the class of loads that are controllable and energy-intensive. Recent efforts have reported reasonable accuracies for energy breakdown methods. For example, disaggregation of EV loads [42], [43], dryer, washing machine, dishwasher [40], or AC [24] has reached accuracies in the range of ~70-95% in several efforts. However, the aforementioned studies have focused either on high-resolution data (1-minute or higher sampling rates), required considerable model parameter selection, and reported the performance on small samples (i.e., a few buildings or a short duration of 1 day to a few days).

A number of studies focused on developing scalable solutions, for which learning the data representation from a sample of labeled data is carried out to identify the breakdown energy for another set of buildings. Using monthly aggregate data, Batra *et al.* [37] estimated the monthly energy breakdown of appliances for around 50 buildings by matching one building with similar ones in which the energy breakdown data is available. They used different features such as monthly energy usage, statistical attributes such as skew and kurtosis, and external attributes such as temperature and building size for classification with K nearest neighbors. The results show an improved performance over a benchmark Hidden Factorial Markov (HMM) model for a variety of appliances including HVAC, washing machine, dryer, with accuracies ranging between 40-75%. The authors later developed a Matrix Factorization approach as a scalable solution for energy breakdown, and they evaluated the performance of the method over 500 houses [44]. However, the scalable solutions in [37], [44] only estimated the monthly usage (i.e., one consumption estimation per month), without providing insight into the variation of daily appliance ToU, which is essential for dynamic load management applications.

A summary of related example works is presented in Table 1. These studies mainly relied on higher-resolution data or in-situ training as common features, which hindered their adoption as a scalable approach for ToU inference. In contrast, in this research, we have investigated the feasibility of an ML framework for inferring ToU of major flexible loads from smart meter data with the 15-minute resolution that accounts for the following features: (1) integrating a resident behavior learning component that leverages energy load shapes for identification of a training dataset to increase the efficiency of the machine learning training process, (2) inferring ToU for unseen buildings which have not contributed to the training process, and (3) investigation of appliance ToU is performed over the hourly basis with application to dynamic load management (e.g., demand-response).

III. APPLIANCE TOU INFERENCE FRAMEWORK

We propose this framework to investigate a scalable approach for inferring appliance ToU from aggregate load shapes in buildings (called *target buildings*) with smart meter data using default resolution of 15-minute intervals. Target buildings are environments that are new to the framework and do not contribute to model training. The premise of this framework centers on learning from buildings (called *training sample buildings*) with similar energy behavior patterns, reflected in the characteristics of their aggregate daily load shapes. The training sample buildings are instrumented with plug metering devices at individual load level in addition to commonly accessible smart meters. The training sample buildings exclude the target building and could be identified as the best matches (in terms of energy behavior patterns) with the target building. The framework is illustrated in Figure 1. We have elaborated on the components of this framework in the following sections.

A. DATA REQUIREMENTS AND PROCESSING

The data from the *training sample buildings* includes: (1) power time-series at 15-min intervals (resembling the sampling rate by smart meters [46]), (2) appliance ownership information, (3) and power time-series from plug meters (i.e., sub-metering at appliance level) with the same resolution as the aggregate-level power data. Considering

TABLE 1. Summary characteristics of related example research efforts.

Resolution	Method	Performance	Major appliances	Potential application	Ref
6 seconds (real power)	Deep learning	F-score of 49%-72%	Washing machine, dishwasher	NILM/ToU	[20]
1 Hz (real power)		F-score of 41%-80%	Washing machine, dishwasher	ToU	[21]
1 Hz (real power)		F-score of ~80%-90%	Dryer, dishwasher	NILM/ToU	[22]
1 Hz (real power)	Hidden Markov Model (HMM)	F-score of 84%	Washer/Dryer	NILM/ToU	[39]
1 min	Hidden Markov Model (HMM)	Energy accuracy of 52%-75%	Dryer, dishwasher	NILM	[40]
30 kHz (current-voltage)	Transfer learning	F-score of 87%-100%	AC, washing machine	ToU	[24]
1 min	Heuristic algorithm	Normalized mean square error of 19%	EV	ToU	[43]
1 min – 5 min	Independent component analysis (ICA)	F-score of 68%-94%	EV	ToU	[42]
1 min	Factorial HMM/Iterative dynamic time warping	F-score of ~70%-90%	Dryer, heat pump	NILM/ToU	[45]
1 hour	Sparse coding	Energy accuracy of 55%	Dryer, dishwasher	NILM	[33]

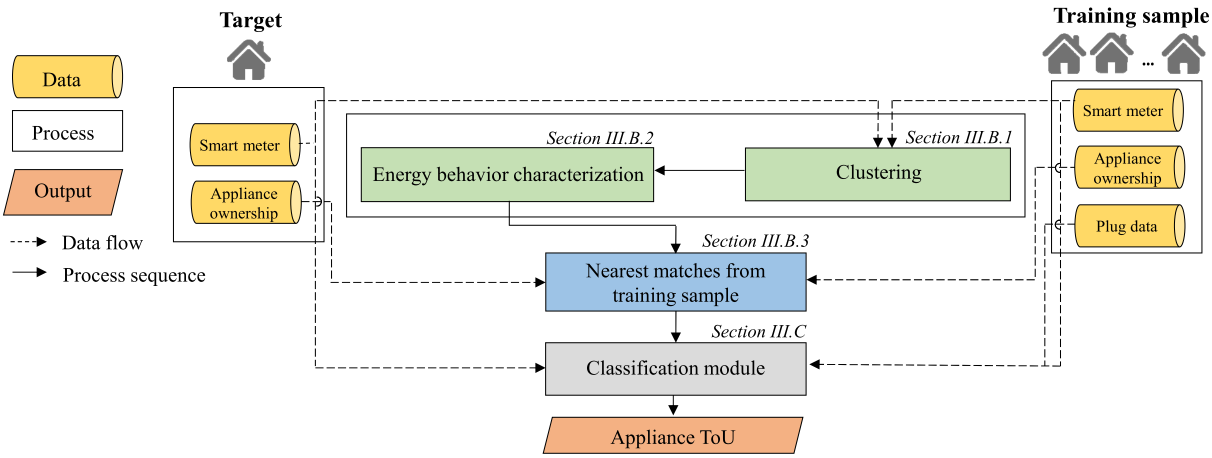


FIGURE 1. Schematic framework for appliance ToU inference.

the data collection campaigns and public release of large-scale energy data sets from residential buildings in recent years, datasets with such characteristics are available for hundreds of buildings [47]. On the other hand, the data sets in *target buildings*, in which the framework makes the inference for, include (1) power time-series from smart meters at 15-min intervals and (2) appliance ownership information. Accordingly, the process of identifying a training sample for a given target with similar behavior (which we refer to as the matching process) calls for aggregate power time-series and appliance ownership information, which are either available or could be acquired. Smart meters have been deployed in half of the United States and considered as a scalable platform for data collection, and appliance ownership data can be obtained through a one-time inquiry from homeowners or could be automated.

B. CHARACTERIZATION OF ENERGY BEHAVIOR PATTERNS FOR THE TRAINING SAMPLE

Identifying the training sample buildings for a given target building relies on energy behavior patterns identification. Therefore, the framework characterizes the energy behavior patterns by leveraging aggregate daily load shapes in order to find potential candidate buildings that are similar to the target. In this process, characterization is defined as understanding the variations of daily energy use patterns, which could be achieved by segmentation of daily load shapes through clustering techniques.

1) SEGMENTATION OF DAILY LOAD SHAPES (CLUSTERING)

Let $P_{i,n}(t) = \{p_1, p_2, p_3, \ldots, p_T\}$ be the daily load shapes in the form of power time-series collected through a smart meter. Here, T is the total number of samples per day, $i \in \{1, 2, \ldots, I\}$ is the building index (I is the total number of buildings), and $n \in \{1, 2, \ldots, N\}$ is the index for historical days (N is the number of historical days). Therefore, considering M, as the total number of load shapes ($M = I = \times N$), the load shape library could be clustered into K clusters.

Different clustering techniques can be employed for the segmentation of energy load shapes, including K-means, hierarchical clustering, and customized methods for power time-series segmentation. In this framework, we have proposed a two-stage method based on Self-Organizing Map (SOM) clustering for creating initial clusters, followed by extracting temporal features from clusters for merging similar ones. In the first step, SOM is applied to all the load shapes (M) to obtain K' clusters, in which $K' > K$. The estimation of K' is achieved by measuring the rate of change in Within Cluster Sum of Squared (WCSS) error as a cluster validation index over different numbers of clusters and identifying the point where the error decline rate for subsequent numbers is negligible (1% in this study). In the second step, clusters with similar behavior are automatically merged into K clusters to reduce redundancy. In doing so, by using centroids of clusters, a number of statistical features (see Figure 2) are extracted to provide quantitative metrics for describing the temporal shape or magnitude of power demand (such as the timing of peak demand) and to characterize the pattern of use in each cluster. Specifically, this feature set describes the energy consumption level (total consumption level), and the distribution pattern (number of relative peaks in each cluster), the timing of peak occurrence, and the intensity of the peak demand (magnitude of peak demand compared to adjacent points). A concatenated set of these features is used as a label for each of K' clusters, and pairs of clusters with a similar feature set are merged until each one has a distinct label (denoted as K final clusters). Therefore, in this framework, no further steps are required to configure the number of clusters. It must be noted that the objective of the clustering step is to identify distinct patterns for load shapes. Therefore, comparable clustering approaches could be applied for this purpose as well [49].

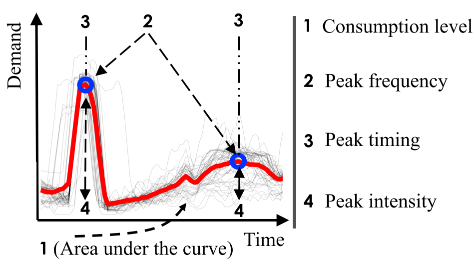


FIGURE 2. Example of a clustered load shape and its extracted features. Details about the feature extraction can be found in [48].

Figure 2 illustrates the extracted features from a cluster that characterizes the information about its load shapes. The red line indicates the cluster centroid averaged over all load shapes. In this figure, a data point is detected as a local peak if it is larger than its two adjacent neighboring points and its prominence is higher than a threshold. The prominence of a data point (the vertical line in Figure 2, feature 4) measures how it stands out due to its height and relative location with respect to adjacent peaks. We configured the thresholds according to three categories of energy consumption levels (Low: lower than 25th quantile, Medium: 25th-75th quantiles, and High: higher than 75th quantile), which are selected based on empirical observations on the power demand magnitude of building energy use. More details about the proposed clustering and feature extraction approach could be found in [48]. Upon creating the clusters, each load shape ($P_{i,n}$) will be associated with a cluster index $k \in \{1, 2, \ldots, K\}$.

2) ENERGY BEHAVIOR CHARACTERIZATION

The outcome of the segmentation for a given building includes a number of representative clusters that summarize different patterns of energy use across a historical period. To characterize the energy behavior of that building, one could consider the dominant cluster that includes the majority of observations over the historical days. However, residential buildings show high uncertainty in energy consumption patterns [12], and they are typically expected to be represented by different clusters and change their pattern over days. Therefore, it is more realistic to associate the behavior of each building with multiple clusters that are commonly observed. To this end, upon segmentation, each building will be characterized based on the frequency of clusters observed over historical data.

Considering the frequency of different clusters for each building, the energy behavior of each building i can be characterized in a feature vector π_i:

$$\pi_i = \{fr_{i,1}, fr_{i,2}, \ldots, fr_{i,K}\} \tag{1}$$

in which k is the index of the clusters and $fr_{i,k}$, $k \in [1 : K]$ is the frequency of observing $C_{i,k}$ across historical days for a building. $C_{i,k}$ is the k^{th} cluster in the load profile data set of building i. $fr_{i,k}$ is defined as:

$$fr_{i,k} = \frac{\|C_{i,k}\|}{N} \tag{2}$$

in which $\|C_{i,k}\|$ is the number of daily load shapes for building i in cluster k. The π_i represents the energy behavior of buildings according to their daily load shapes and the frequency of observations.

Figure 3 illustrates an example of clusters and the associated feature vector (π) that is derived according to the frequency of clusters over a period of historical data. Clustering is performed on the entire sample (both training and target buildings), and π (Eq. 1) is characterized for all the buildings.

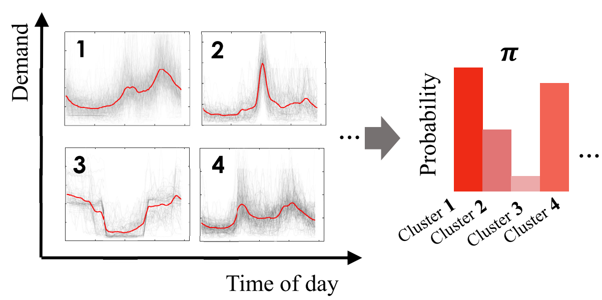


FIGURE 3. Representation of π according to the frequency of clusters.

3) IDENTIFYING NEAREST MATCHES TO THE TARGET FOR TRAINING

To identify the training sample buildings for ToU inference in the target building, we adopted the use of the KNN algorithm for finding the nearest training sample buildings that match the energy behavior of the target. Considering the training sample size of $I' \ll I$, the nearest matches in a community with respect to the target are the $[1, \ldots, I']$. Here, π_i vector is used as the feature vector for the KNN algorithm with Euclidean distance as the similarity measure. In this work, we primarily investigated 10 as the number of neighbors for the KNN, while comparing it against other values in the result section.

C. APPLIANCE TIME OF USE INFERENCE

Given the power time series of the daily load shape for the target building i on day n, $P_{i,n}(t)$, the objective is to identify the time of use of a flexible appliance across a collection of time bins during the day – $\Omega_{i,n} = \{\omega_1, \omega_2, \ldots, \omega_\Gamma\}$ in which Γ is the total number of time bins ($\tau \in [1 : \Gamma]$). A time bin is a window of multiple data points with the length l, which represents a timeframe that could be selected for DER management or a DR event. The elements in the vector $\Omega_{i,n}$ have the following binary form:

$$\omega_\tau = \begin{Bmatrix} 1 & \textit{if appliance was used} \\ 0 & \textit{if appliance was not used} \end{Bmatrix} \tag{3}$$

Therefore, to infer a flexible load status we have:

$$\gamma_{i,n,\tau} = \begin{Bmatrix} 1 \;\; if\; P(\omega_\tau = 1|P_{i,n,\tau}(t)) > P(\omega_\tau = 0|P_{i,n,\tau}(t)) \\ 0 \qquad\qquad\qquad else \end{Bmatrix} \tag{4}$$

in which $\gamma_{i,n,\tau}$ is the appliance ToU prediction for building i, on day n, and at the time bin ω_τ. $P_{i,n,\tau}(t)$ denotes the $P_{i,n}(t)$ power time-series associated with the time frame τ, in which $t \in [\tau l - l + 1 : \tau l]$. $\gamma_{i,n,\tau}$ is used as the appliance ToU predictor for presenting the results in this paper.

In order to prepare the ground-truth in the form of $\Omega_{i,n}$ (binary series across different time bins per day) for the training sample buildings, the continuous power time-series need to be converted into the binary form at each time bin. Considering $P^a(t) = \{P^a_1, P^a_2, P^a_3, \ldots P^a_T\}$ as the appliance power time-series ground truth, we employed the following equation for preparing the ground-truth $\Omega_{i,n}$ in our training sample:

$$\omega_\tau = \begin{Bmatrix} 1 \;\; if \quad \max(P^a(t)|_{t=l(\tau-1)+1:\tau l} > P^n) \\ 0 \qquad\qquad\qquad else \end{Bmatrix} \tag{5}$$

in which P^n is the nominal power draw threshold of a load while being on.

Figure 4 shows the visualization for the power time-series from the smart meter at 15-minute intervals, the appliance-level power time-series, and extraction of the binary operational mode at each time bin. Figure 4(a) shows the smart meter data of one building over twenty days, and Figure 4(b) is the magnified view of the first day. Figure 4(c) is the power time-series of a dryer for the same building at the same timeframe as Figure 4(a), while Figure 4(d) shows the extracted binary operational mode at each time bin (with a duration of 2 hours) based on Eq. (5).

1) CLASSIFIER COMPONENT

To classify the appliance ToU at each time bin in the target building, we investigated the application of several classifiers. The processed historical data of daily load profiles $P(t)$ and Ω from the training sample buildings are used for training, and $P(t)$ from the target building is used for inferring the appliance ToU. Therefore, the classifier infers Ω from $P(t)$ by solely relying on the information from the training sample with known ToU [$f : P(t) \rightarrow \Omega$]. Using the nearest matches from the training sample building (described in section III.B), predictions for each load is carried out separately. To this end, we investigated Dense Neural Network (Dense NN), KNN, Support Vector Machine (SVM), and Random Forest (RF) as four classification algorithms.

Considering that building load shapes are presented in relatively low dimensions and the dataset that is moderate in size (compared to the common vision and text datasets), we opted for a two dense-layer NN. The architecture consists of 300 hidden units in the first layer with ReLU activation function, a sigmoid activation function in the second layer to map two states of 'On' and 'Off', and RMSprop as the

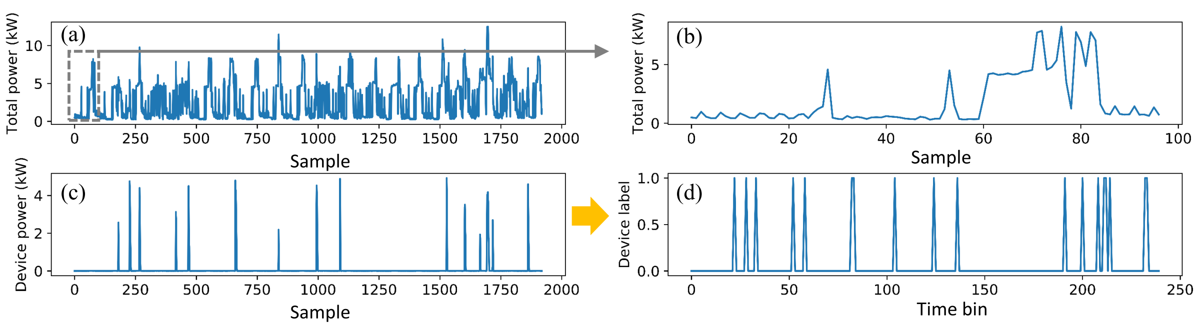


FIGURE 4. Example of daily load shapes over twenty days for one building and its corresponding dryer power time series and label for each time bin: (a) power consumption collected by using a smart meter at 15-minute interval, (b) magnified view of the first day from part (a), (c) power consumption of a dryer for the same time span in (a), and (d) extracted labels of dryer from part (c) at two-hour intervals.

optimizer. The second classifier, KNN, uses the class output of k nearest neighbors based on Euclidean distance for classification. 5 nearest neighbors for load shapes were used in the analysis upon empirical observation. The third classifier, SVM, defines decision boundaries based on hyperplanes to separate the feature space into classes. The fourth classifier, RF, is an ensemble learning method that leverages multiple decision trees and selects the mode of the classes of individual trees as the output.

2) OVERSAMPLING CONSIDERATION FOR TRAINING

The appliance ToU data sets are imbalanced by nature. Intuitively, for most buildings, it is expected that the "Off" class is dominant compared to the "On" class. However, for our problem, the minority "On" class has higher importance. Therefore, it is desired to reduce the number of false negatives (FN) compared to false positives (FP). Imbalanced datasets can adversely affect the performance of classifiers and lead to bias in favor of the majority class [50]. Solutions to mitigate this problem include oversampling or creating synthetic data for the minority class [51]. Here, we used the Synthetic Minority Over-sampling Technique (SMOTE) for mitigating the imbalanced problem. SMOTE [52] creates artificial instances of the minority class that are close to the existing ones in the dataset. Synthetic samples are created by calculating the difference between observations from the minority class and their nearest neighbors (Δ). These differences are then multiplied by a random coefficient between 0 and 1 and added along each feature to create a synthetic observation. Using this approach, the extra synthetic data for 'On' examples is only added to the training sample of buildings to ensure it does not affect the dataset for the target building.

IV. RESULTS AND DISCUSSION

A. DATASET AND METRICS

A sample of 467 buildings from Pecan Street Dataport [47], primarily located in Austin, TX, in July and August 2015 was used as the case study for segmentation and creating clustered load shapes. After pre-processing and performing moving average filtering on 15-minute data to reduce noise, 26870 daily load shapes were obtained. Moving average with a 1-hour window was considered to reduce the impact of noise in energy load shapes.

To evaluate the framework, we considered EV and dryer as two instances of flexible loads that are suitable for DER management and DR applications. In our dataset, 80 and 180 buildings from the entire community of 467 buildings had EV and dryer, respectively. The process of identifying the nearest matches for each load was carried out with respect to its corresponding subset. For each load, we considered 10 buildings as the target buildings (test set) that owned the same appliance (using ownership data) with 20 days of data for evaluating the performance to provide adequate instances for evaluation. For each target building, 10 nearest matches were selected as the training sample with 60 days of data. The aggregate data and appliance data had a sampling rate at every 15 minutes. A value of $l = 8$, corresponding to time bins with a duration of $\tau = 2$ hours was selected to represent the common timeframe for running DR events [53].

For training the classifiers, we used two scenarios of training for the target buildings. In the first scenario, labeled as "RBL" (resident behavior learning), we used the data from 10 nearest matches, selected by using the method discussed in Section III.B. In the second scenario, we used the data from 30 buildings randomly sampled in the community without considering the nearest matches. We deliberately selected a higher number of buildings for the second scenario to use more data for training and compare it with the alternative approach of using similar matches for training. Nonetheless, a sensitivity analysis for the training sample size by evaluating different combinations of 10, 20, or 30 buildings is presented later in Section IV-C.

Standard metrics including precision, recall, and F-score were used for evaluation considering our imbalanced datasets and the need for assessing the trade-off among TP, FP, and FN.

B. BUILDING ENERGY CHARACTERIZATION

The two-stage clustering technique described in section III.B.1 was employed on the smart meter data. An initial 60 clusters were generated on the entire dataset by examining the WCSS error over the range of K' and selecting the point where the error decline for subsequent numbers of clusters is negligible (1%). In the second stage, features from clusters were extracted and those with similar features were merged. This was done to merge the highly correlated ones so that clusters reflect distinct load shape patterns. This process further reduced the cluster library size to 39 clusters, and they were considered as the final clusters for the rest of the analysis. Although other common cluster validation indices (e.g., silhouette coefficient) can be used in the process of selecting the number of clusters, these generic metrics may not work as expected for this domain-specific problem [54] and lead to underestimation of the number of clusters.

Figure 5 shows samples of clusters in the library and their frequency of occurrence. After this process, each daily load shape in the library (for buildings in both groups of training samples and target) was annotated with its cluster index and building ID.

Using the results in the previous section, the energy behavior of the building, π_i (described in Section III.B.2), was calculated. The nearest ten matches for each target were identified using KNN. Figure 6 demonstrates an example histogram of clusters in one of the target buildings and its nearest neighbor. Figure 6(a) compares the histogram of clusters, and Figure 6(b) shows the top three clusters with higher occurrence for two buildings. As shown, there is a high similarity in terms of energy use behavior across historical days for these two buildings.

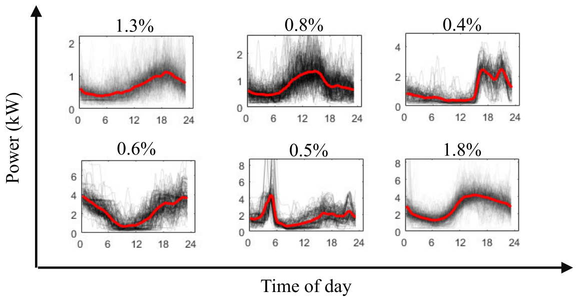


FIGURE 5. Examples of clustered energy load shapes and their frequency of occurrence in the community.

C. APPLIANCE TIME-OF-USE INFERENCE

Through using the statistical approach and different classifiers described in section III.C, ToU inference for individual loads based on smart meter data was investigated. Figure 7 shows a visual demonstration of the EV charging prediction for one building over 20 days. The upper subplot is the smart meter data, collected at 15-minute intervals. The lower subplot compares the ground-truth and prediction of EV charging status at each associated time bin using Dense NN+RBL. As can be seen, the detected charging instances are highly in agreement with the ground-truth, and most of the charging instances are correctly identified at its right time by load shape analysis.

To provide a comparison between different methods, Figure 8 shows the F-score values, averaged over all target buildings. As noted, 'RBL' represents the 'resident behavior learning' approach for selecting the training sample buildings. Based on the results shown in Figure 8, for both EV and dryer, the 'Dense NN+RBL' approach demonstrates the best performance, with an average F-score of 83% and 71% respectively. For EV, for all four classifiers, integrating the 'RBL' component improved the results (5% on average). For the dryer, integrating the 'RBL' component improved the results for two classifiers, while the impact was negligible (1% improvement). It must be noted that three times more data was used for training based on random sampling compared to the RBL approach.

Table 2 provides drill-down performance metrics for individual buildings. F-score, precision, and recall metrics were reported. For EV, out of 10 examples, 9 of them had improved F-score when classifiers were trained using the 'RBL' approach. For the dryer, 6 out of 10 examples had improved F-score when the classifier integrates the 'RBL' approach for training. It was also observed, that evaluation for EV achieved average recall and precision of 87% and 83%,

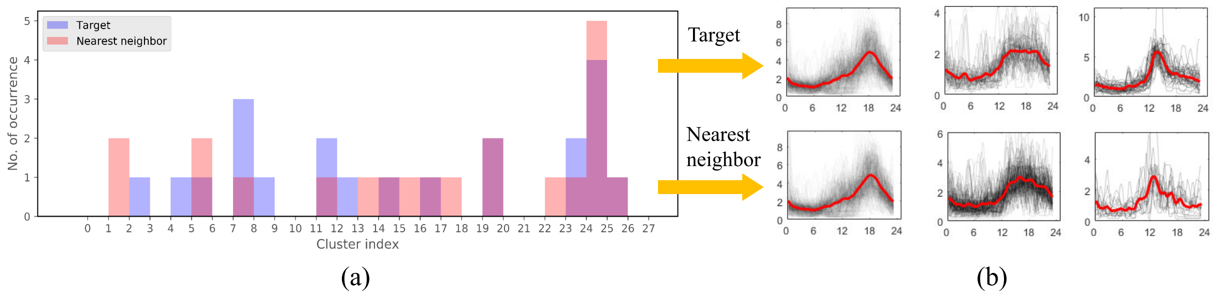


FIGURE 6. Building energy characterization and similarity search for one target building: (a) histograms of clusters for two buildings, and (b) top three daily load shape clusters.

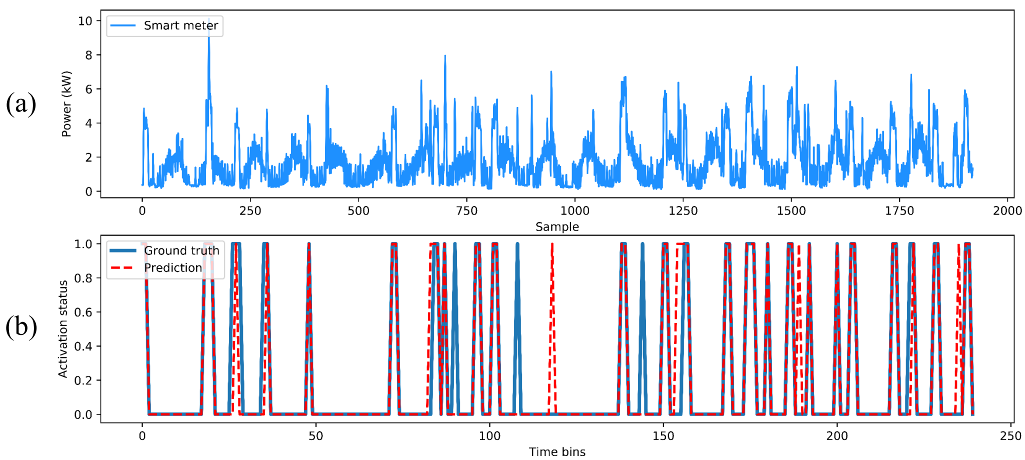


FIGURE 7. Visual demonstration for EV charging of one building for 20 days: (a) smart meter data, (b) ground-truth and prediction for EV charging.

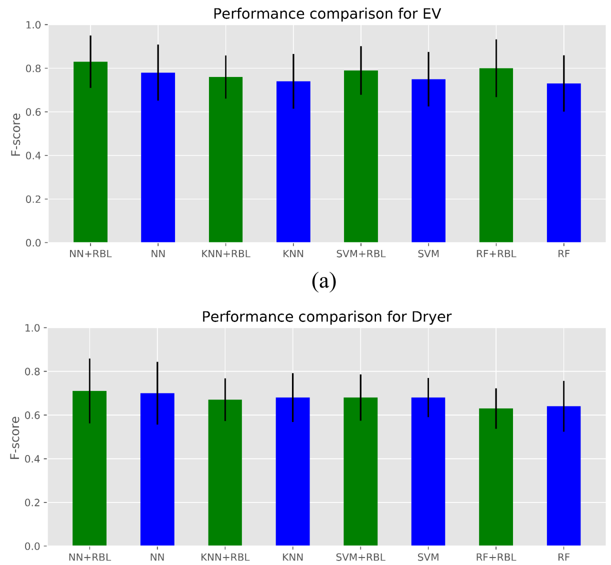


FIGURE 8. **Comparison of different algorithms for (a) EV and (b) dryer.**

respectively in the best case, while these values for dryer were 76% and 73%, respectively. In general, prediction on EV showed better performance compared to the dryer. Furthermore, the results in Table 2 show that in most cases, the ToU can be predicted with high accuracy with Dense NN+RBL, while there were a few instances that turned out to be challenging in prediction (further discussed in Section IV.D).

Figure 9 shows the aggregate confusion matrix for all ten buildings over twenty days (200 days of observations) with two-hour time bins. For EV, there were a total of 2061 'Off' and 339 'On' instances, and for dryer, there were a total of 2070 'Off' and 330 'On' instances. As could be inferred from Figure 9, for EV, out of 339 charging instances ('On' class), 273 of them were correctly identified. For the dryer, out of 330 'On' instance, 216 of them were correctly identified. Regarding the 'Off' instances, 1863 out of 2061 for EV, and 1740 out of 2070 instances for dryer were correctly classified. As the results show, the proposed approach has the potential to add a new layer of information to detect the operational details for flexible loads by using daily load shape information from smart meters.

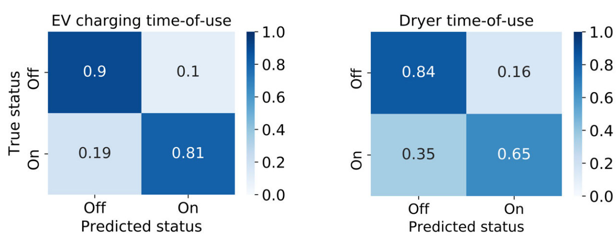


FIGURE 9. **Appliance ToU inference for (a) EV (including 2061 'Off' and 339 'On' observations) and (b) Dryer (including 2070 'Off' and 330 'On' observations).**

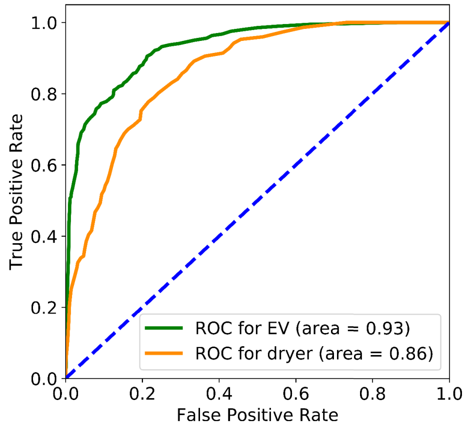


FIGURE 10. **ROC curve across all target buildings.**

To demonstrate the trade-off between true positive rate and true negative rate, Figure 10 shows the Receiver Operating Characteristic (ROC) curves across all buildings. For EV and dryer, the Area Under the Curve (AUC) shows a high value of 0.93 and 0.86 (1 is the perfect classification), corresponding to the probability of ranking a randomly chosen 'On' observation higher than a randomly chosen 'Off' observation. To adjust different levels of True Positive Rate (TPR) versus False Positive Rate (FPR), varying levels of thresholds for the classifier can be selected. For example, for EV, achieving TPR values of 0.7, 0.8, and 0.9 results in FPR values of 0.02, 0.10, and 0.22, respectively. For the dryer, the same TPR values of 0.7, 0.8, and 0.9 correspond to the FPR values of 0.18, 0.25, and 0.36, respectively.

- ***Sensitivity to the training size:*** To evaluate the impact of training sample size in our comparative analyses, we performed an experiment by changing the ratio of training sample for the scenario without "RBL" and compared it with the scenario that included "RBL". Specifically, in addition to the presented results in the previous section in which the ratio of training size in the scenario without "RBL" to the scenario with "RBL" was set to 3, we also performed the same experiments with the training size ratio of 2 and 1.
  Figure 11 presents the results for improvements in the F-score by the inclusion of the RBL step in the framework. The F-score was averaged over the entire test set for different ratios of random sampling versus RBL for smart sampling. The first three subplots show F-score and improvement in F-score for different training ratios of 1, 2, 3. The x-axis represents the F-score using "RBL" component, and the y-axis shows the difference of F-score compared to the baseline of random sampling. The last subplot represents the variation of F-score improvement across the training ratios. As can be seen, in most cases, the inclusion of the "RBL" component improves the performance of algorithms since data points are placed above the baseline (the gray dashed line indicates no change). Besides, it confirms our assumption that setting the same number

TABLE 2. Performance comparison of different algorithms for individual buildings (best results are shown in bold).

						EV						
Method	**Dense NN+RBL**			**KNN+ RBL**			**SVM+RBL**			**RF+RBL**		
Building ID#	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**
6072	**0.86**	**0.87**	0.86	0.82	0.84	0.80	0.83	0.84	0.83	**0.86**	0.85	**0.89**
9776	0.59	0.63	0.59	0.61	0.63	0.60	0.58	**0.64**	0.59	**0.62**	**0.64**	**0.61**
545	**0.91**	**0.88**	0.96	0.80	0.78	0.83	0.86	0.80	**0.97**	0.84	0.80	0.91
3036	**0.89**	**0.93**	0.86	0.79	0.87	0.75	0.85	0.90	0.81	0.87	0.91	0.84
1169	**0.91**	**0.90**	0.94	0.78	0.86	0.73	0.89	0.87	0.91	0.84	0.84	0.85
5749	**0.72**	**0.93**	**0.67**	0.67	0.86	0.63	0.65	0.80	0.62	0.67	0.84	0.63
9935	**0.66**	**0.78**	**0.68**	0.61	0.72	0.64	0.59	0.69	0.62	0.64	0.75	0.59
1629	**0.92**	0.90	**0.94**	0.87	0.88	0.86	0.88	0.86	0.90	0.89	**0.91**	0.89
4352	**0.96**	**0.97**	0.95	0.91	0.92	0.90	0.95	**0.97**	0.93	**0.96**	**0.97**	0.95
6990	0.84	0.87	0.82	0.78	0.83	0.75	0.78	0.81	0.77	0.84	**0.88**	0.81
Average	**0.83**	**0.87**	**0.83**	0.76	0.82	0.75	0.79	0.82	0.79	0.80	0.84	0.80
	Dense NN			**KNN**			**SVM**			**RF**		
6072	0.76	0.79	0.73	0.76	0.82	0.74	0.73	0.75	0.71	0.74	0.76	0.73
9776	0.58	0.61	0.58	0.54	0.61	0.57	0.55	0.63	0.58	0.54	0.60	0.54
545	0.87	0.83	0.93	0.87	0.84	0.90	0.83	0.78	0.92	0.82	0.76	0.96
3036	0.87	0.87	**0.88**	0.77	0.74	0.81	0.78	0.76	0.80	0.75	0.71	0.80
1169	**0.91**	**0.90**	0.94	0.86	0.84	0.88	0.88	0.89	0.87	0.89	0.85	**0.95**
5749	0.71	**0.93**	0.66	**0.72**	**0.93**	**0.67**	0.71	0.88	0.66	0.67	0.76	0.62
9935	0.52	0.68	0.63	0.50	0.66	0.61	0.49	0.57	0.54	0.50	0.63	0.46
1629	0.86	0.83	0.91	0.70	0.68	0.73	0.77	0.73	0.88	0.69	0.67	0.80
4352	0.84	0.78	0.94	0.89	0.87	0.91	0.93	0.90	**0.96**	0.91	0.88	0.95
6990	**0.85**	0.85	**0.85**	0.80	0.81	0.78	0.81	0.81	0.82	0.80	0.78	0.83
Average	0.78	0.81	0.80	0.74	0.78	0.76	0.75	0.77	0.78	0.73	0.74	0.76
						Dryer						
	Dense NN+RBL			**KNN+ RBL**			**SVM+RBL**			**RF+RBL**		
Building ID#	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**
6990	0.65	0.79	0.63	**0.70**	**0.81**	**0.67**	0.69	0.77	0.65	0.64	0.70	0.62
8282	**0.78**	**0.87**	**0.74**	0.65	0.69	0.64	0.72	0.77	0.70	0.70	0.74	0.67
3044	0.64	0.64	0.63	0.68	0.69	**0.68**	0.63	0.63	0.63	0.59	0.59	0.62
6139	0.71	0.70	**0.73**	0.66	0.67	0.65	**0.73**	0.75	0.71	0.64	0.63	**0.73**
7901	**0.91**	**0.86**	**0.98**	0.79	0.77	0.82	0.77	0.72	0.85	0.73	0.69	0.83
9356	**0.76**	**0.78**	**0.75**	0.64	0.65	0.63	0.70	0.72	0.69	0.59	0.61	0.65
4874	**0.85**	**0.83**	0.88	0.82	0.78	0.89	0.79	0.74	0.89	0.80	0.75	0.92
4505	0.74	0.75	0.73	0.67	0.69	0.66	0.75	0.74	0.76	0.62	0.62	0.63
5288	0.34	0.59	0.52	0.44	0.49	0.50	0.40	0.50	0.50	0.45	0.52	0.44
3367	0.71	0.74	0.70	0.68	0.69	0.68	0.67	0.66	0.68	0.55	0.56	0.61
Average	**0.71**	0.75	**0.73**	0.67	0.69	0.68	0.68	0.70	0.71	0.63	0.64	0.67
	Dense NN			**KNN**			**SVM**			**RF**		
6990	0.64	0.77	0.62	0.64	0.73	0.62	0.65	0.70	0.62	0.68	0.72	0.64
8282	0.69	0.76	0.67	0.65	0.69	0.64	0.62	0.70	0.62	0.59	0.60	0.60
3044	0.68	0.72	0.66	**0.69**	**0.75**	0.67	0.67	**0.75**	0.65	0.57	0.58	0.56
6139	0.68	0.71	0.67	0.63	0.65	0.62	0.71	**0.76**	0.68	0.57	0.57	0.63
7901	0.86	0.79	0.97	0.80	0.83	0.78	0.80	0.82	0.79	0.82	0.80	0.85
9356	0.68	0.76	0.67	0.61	0.69	0.62	0.60	0.67	0.61	0.51	0.55	0.47
4874	0.83	0.78	**0.95**	0.80	0.76	0.85	0.74	0.70	0.82	0.68	0.65	0.85
4505	**0.85**	**0.88**	**0.82**	0.79	0.83	0.77	0.76	0.83	0.73	0.83	0.87	0.80
5288	0.33	**0.64**	**0.54**	0.42	0.50	0.50	**0.48**	0.54	0.51	0.46	0.60	0.44
3367	0.74	**0.80**	0.72	**0.75**	0.79	0.73	**0.75**	0.79	**0.73**	0.66	0.66	0.67
Average	0.70	**0.76**	**0.73**	0.68	0.72	0.68	0.68	0.73	0.68	0.64	0.66	0.65

of buildings for both scenarios would further accentuate the improvement by including the 'RBL' component.

D. DISCUSSION AND LIMITATION

There are a number of limitations associated with this work as elaborated here:

(1) We only focused on two classes of loads including EV and dryer that have high power draws. The usage of these devices can induce some noticeable change in load shapes pattern which makes them favorable for our analysis. However, the identification of ToU for appliances with lower power draws will not make such changes in the load shape pattern, and therefore, might not be detected by the proposed approach. Nonetheless, we focused on two classes of flexible loads that had the highest importance to DR applications in the residential sector [55], [56].

(2) We considered that $P(t)$ time-series are available as historical data. Using this framework for predictive DR requires the information of $P(t)$ in the near future that in turn requires forecasting of the load shape and prediction of the appliance use. In recent years, forecasting at the individual building level using smart meter data has received increased attention, and recent deep learning methods have shown promising performance (e.g., [12], [57], with a mean absolute percentage error of ∼10-20%). Given the ongoing research efforts in energy forecasting, we expect the predictive ToU to be applicable by forecasting $P(t)$ with some levels of cascading errors. Nonetheless, by using historical load shape data and ToU information over a span of several days, predictive Markov models of future ToU for specific loads could be also developed by training over the historic patterns of energy use.

(3) As the results in Table 2 showed, although the framework showed a relatively high detection rate in most buildings, there are a few instances that turned out to be challenging. Figure 12 compares the ToU identification of

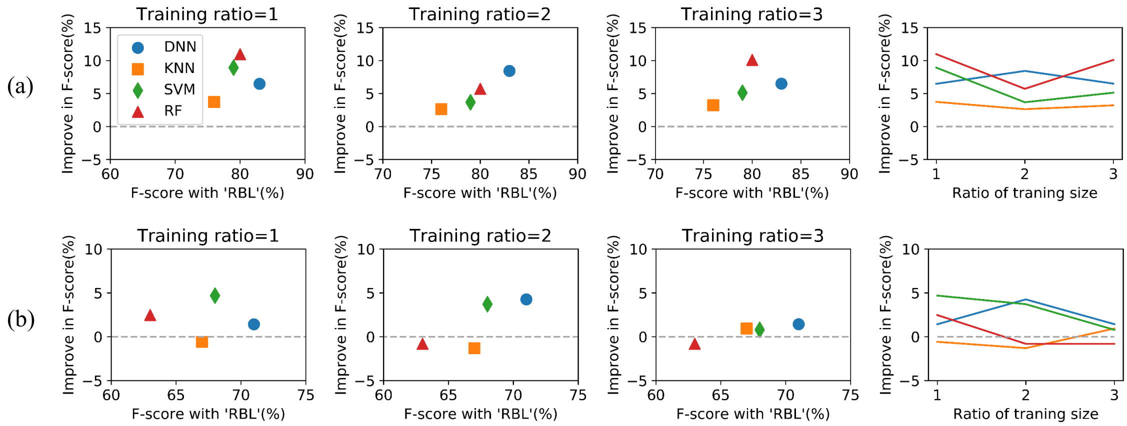


FIGURE 11. **Comparison of F-score by changing the ratio of training size for (a) EV and (b) dryer.**

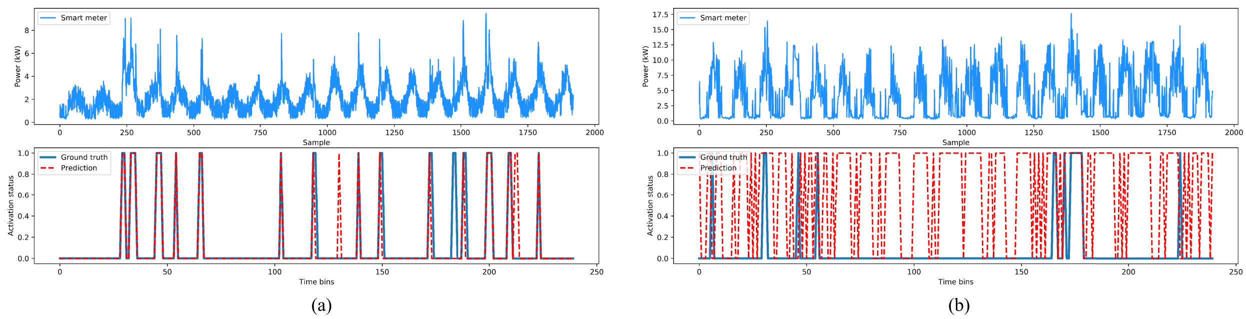


FIGURE 12. **Comparison of dryer identification for two buildings with (a) good performance and (b) worst performance.**

dryer for a building with the worst performance (Building #9: ID 5288; F-score = 0.34) with another building with good performance (Building #5: ID 7901; F-score = 0.91). As could be seen, for the building with the high error value (Figure 12(b)), the demand magnitude is considerably high at most times. Therefore, it could be considered an outlier in terms of energy consumption, with no possible similar match for the training. Furthermore, there are many considerable sharp peaks in the load shapes, and the associated time bins were classified with dryer operation as false positives. This could be caused by simultaneous interference from other appliances that have similar load behaviors such as a water heater or an AC. To alleviate problems for such cases, we could add heuristics to detect outliers by using the known power draw information from typical appliances and integrate contextual attributes such as occupancy information and outdoor temperature data for estimation of baseline loads. Additionally, outlier detection techniques can be applied to detect load profiles with discord in the building pool as a pre-processing or post-processing step [58], [59].

V. CONCLUSION

In this study, we investigated a scalable framework for inferring time-of-use (ToU) for major flexible loads in support of DERs applications and DR operations. The framework draws on the use of low-resolution real power data, sampled at 15-minute intervals through ubiquitously available smart meter infrastructure, and a training scheme that excludes in-situ training – i.e., the inference is carried out for buildings that are not part of the training data set. The framework uses a number of known buildings with specialized appliance-level metering instrumentation to provide the basis for generalized training (i.e., the training pool), and searches in the training pool to identify the buildings (i.e., the training sample) in which their energy behavior matches to the target building.

The framework employs a segmentation component that uses the clustering of daily load shapes and compares the frequency of observed clusters across buildings to match them as buildings with similar energy behaviors. Upon identification of the training sample, a classification algorithm that uses the daily load shapes as the input parameter detects the use of flexible loads across different times of a day. Various classifiers were employed to evaluate their performance in identifying the ToU. A case study on Pecan Street Dataport for EV and dryer was conducted. Using only the information from a limited number of buildings in the training sample, the best average recall of 87% and 76%, and precision of 83% and 73% were achieved for EVs and dryers, respectively. The results show the feasibility and potential of the approach for adding a new level of information to the daily load shapes (i.e., the ToU) that reflects on the casualty of the observed energy consumption patterns.

Future directions of this research will include the use of machine learning algorithms for automated feature extraction, accounting for simultaneous operation of loads with high power draw through heuristic augmentation of the framework, incorporating contextual information such as occupancy or building area, in addition to the identification of other flexible load types.

REFERENCES

[1] M. Afzalan and F. Jazizadeh, "Residential loads flexibility potential for demand response using energy consumption patterns and user segments," *Appl. Energy*, vol. 254, Nov. 2019, Art. no. 113693.

[2] M. Afzalan and F. Jazizadeh, "Data-driven identification of consumers with deferrable loads for demand response programs," *IEEE Embedded Syst. Lett.*, vol. 12, no. 2, pp. 54–57, Jun. 2020.

[3] M. Afzalan and F. Jazizadeh, "Efficient integration of smart appliances for demand response programs," in *Proc. 5th Conf. Syst. Built Environ.*, Nov. 2018, pp. 29–32.

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

[5] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016.

[6] O. Alrumayh and K. Bhattacharya, "Flexibility of residential loads for demand response provisions in smart grid," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6284–6297, Nov. 2019.

[7] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "An algorithm for intelligent home energy management and demand response analysis," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2166–2173, Dec. 2012.

[8] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[9] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Appl. Energy*, vol. 141, pp. 190–199, Mar. 2015.

[10] J. Kwac, J. Flora, and R. Rajagopal, "Lifestyle segmentation based on energy consumption data," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2409–2418, Jul. 2018.

[11] M. S. Piscitelli, S. Brandi, and A. Capozzoli, "Recognition and classification of typical load profiles in buildings with non-intrusive learning approach," *Appl. Energy*, vol. 255, Dec. 2019, Art. no. 113727.

[12] 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.

[13] M. Afzalan and F. Jazizadeh, "Investigating the appliance use patterns on the households' electricity load shapes from smart meters," in *Proc. Comput. Civil Eng, Smart Cities, Sustainability, Resilience, Amer. Soc. Civil Eng.*, Reston, VA, USA, 2019, pp. 154–161.

[14] K. S. Cetin, P. C. Tabares-Velasco, and A. Novoselac, "Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use," *Energy Buildings*, vol. 84, pp. 716–726, Dec. 2014.

[15] Y. Ji and R. Rajagopal, "Demand and flexibility of residential appliances: An empirical analysis," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 1020–1024.

[16] S. Lin, F. Li, E. Tian, Y. Fu, and D. Li, "Clustering load profiles for demand response applications," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1599–1607, Mar. 2019.

[17] A. Malik, N. Haghdadi, I. MacGill, and J. Ravishankar, "Appliance level data analysis of summer demand reduction potential from residential air conditioner control," *Appl. Energy*, vol. 235, pp. 776–785, Feb. 2019.

[18] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage," *Energy Buildings*, vol. 67, pp. 668–679, Dec. 2013.

[19] A. Barbato, A. Capone, M. Rodolfi, and D. Tagliaferri, "Forecasting the usage of household appliances through power meter sensors for demand management in the smart grid," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2011, pp. 404–409.

[20] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ. (BuildSys)*. New York, NY, USA: ACM, 2015, pp. 55–64.

[21] K. Said Barsim and B. Yang, "On the feasibility of generic deep disaggregation for single-load extraction," 2018, *arXiv:1802.02139*. [Online]. Available: http://arxiv.org/abs/1802.02139

[22] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, "Denoising autoencoders for non-intrusive load monitoring: Improvements and comparative evaluation," *Energy Buildings*, vol. 158, pp. 1461–1474, Jan. 2018.

[23] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "Appliance classification using VI trajectories and convolutional neural networks," *Energy Buildings*, vol. 158, pp. 32–36, Jan. 2018.

[24] Y. Liu, X. Wang, and W. You, "Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5609–5619, Sep. 2019.

[25] F. Jazizadeh, M. Afzalan, B. Becerik-Gerber, and L. Soibelman, "EMBED: A dataset for energy monitoring through building electricity disaggregation," in *Proc. 9th Int. Conf. Future Energy Syst.* New York, NY, USA: ACM, Jun. 2018, pp. 230–235.

[26] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[27] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, Dec. 2012.

[28] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.

[29] L. Pereira, "NILMPEds: A performance evaluation dataset for event detection algorithms in non-intrusive load monitoring," *Data*, vol. 4, no. 3, p. 127, Aug. 2019.

[30] F. Jazizadeh, B. Becerik-Gerber, M. Berges, and L. Soibelman, "An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems," *Adv. Eng. Informat.*, vol. 28, no. 4, pp. 311–326, Oct. 2014.

[31] F. Jazizadeh, B. Becerik-Gerber, M. Berges, and L. Soibelman, "Unsupervised clustering of residential electricity consumption measurements for facilitated user-centric non-intrusive load monitoring," *Comput. Civ. Build. Eng*, pp. 1869–1876, 2014.

[32] M. Afzalan and F. Jazizadeh, "An automated spectral clustering for multi-scale data," *Neurocomputing*, vol. 347, pp. 94–108, Jun. 2019.

[33] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1153–1161.

[34] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "On the Bayesian optimization and robustness of event detection methods in NILM," *Energy Buildings*, vol. 145, pp. 57–66, Jun. 2017.

[35] M. Afzalan, F. Jazizadeh, and J. Wang, "Self-configuring event detection in electricity monitoring for human-building interaction," *Energy Buildings*, vol. 187, pp. 95–109, Mar. 2019.

[36] M. Gaur and A. Majumdar, "Disaggregating transform learning for non-intrusive load monitoring," *IEEE Access*, vol. 6, pp. 46256–46265, 2018.

[37] N. Batra, A. Singh, and K. Whitehouse, "Gemello: Creating a detailed energy breakdown from just the monthly electricity bill," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: ACM, Aug. 2016, pp. 431–440.

[38] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, Feb. 2013.

[39] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2012, pp. 1472–1482.

[40] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, "A hierarchical hidden Markov model framework for home appliance modeling," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3079–3090, Jul. 2018.

[41] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234, Jan. 2013.

[42] A. A. Munshi and Y. A.-R.-I. Mohamed, "Unsupervised nonintrusive extraction of electrical vehicle charging load patterns," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 266–279, Jan. 2019.

[43] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. Y. Hwang, and J. K. Moon, "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate," 2014, *arXiv:1404.5020*. [Online]. Available: http://arxiv.org/abs/1404.5020

[44] N. Batra, H. Wang, A. Singh, and K. Whitehouse, "Matrix factorisation for scalable energy breakdown," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 4467–4473.

[45] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Appl. Energy*, vol. 185, pp. 331–344, Jan. 2017.

[46] *Assessment of Demand Response and Advanced Metering*, Federal Energy Regulatory Commission, Washington, DC, USA, 2008.

[47] *Dataport: The World's Largest Energy Data Resource*, Pecan. Street, Pecan Street Inc, Austin, TX, USA, 2015.

[48] M. Afzalan and F. Jazizadeh, "Semantic search in household energy consumption segmentation through descriptive characterization," in *Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Buildings, Cities, Transp.*, Nov. 2019, pp. 263–266.

[49] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5196–5206, Sep. 2018.

[50] T. Hasanin, T. M. Khoshgoftaar, J. L. Leevy, and N. Seliya, "Examining characteristics of predictive models with imbalanced big data," *J. Big Data*, vol. 6, no. 1, p. 69, Dec. 2019.

[51] T. Hasanin, T. M. Khoshgoftaar, J. L. Leevy, and R. A. Bauder, "Severely imbalanced big data challenges: Investigating data sampling approaches," *J. Big Data*, vol. 6, no. 1, p. 107, Dec. 2019.

[52] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[53] J. Kwac and R. Rajagopal, "Data-driven targeting of customers for demand response," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2199–2207, Sep. 2016.

[54] S. Iyengar, S. Lee, D. Irwin, and P. Shenoy, "Analyzing energy usage on a city-scale using utility smart meters," in *Proc. 3rd ACM Int. Conf. Syst. Energy-Efficient Built Environ.*, Nov. 2016, pp. 51–60.

[55] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load profiles of selected major household appliances and their demand response opportunities," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 742–750, Mar. 2014.

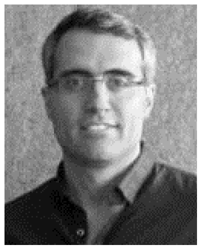
[56] K. Knezovic, M. Marinelli, P. Codani, and Y. Perez, "Distribution grid services and flexibility provision by electric vehicles: A review of options," in *Proc. 50th Int. Univ. Power Eng. Conf. (UPEC)*, Sep. 2015, pp. 1–6.

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

[58] J. Y. Park, E. Wilson, A. Parker, and Z. Nagy, "The good, the bad, and the ugly: Data-driven load profile discord identification in a large building portfolio," *Energy Buildings*, vol. 215, May 2020, Art. no. 109892.

[59] A. Capozzoli, M. S. Piscitelli, S. Brandi, D. Grassi, and G. Chicco, "Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings," *Energy*, vol. 157, pp. 336–352, Aug. 2018.

MILAD AFZALAN (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees from the University of Tehran, Iran, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Via Department of Civil and Environmental Engineering, Virginia Tech, USA. His research interests lay at the intersection of intelligent cyber-physical systems through data sensing, control of building systems, and infrastructure operational analytics.



FARROKH JAZIZADEH received the M.Sc. degrees in computer science and civil engineering and the Ph.D. degree in civil engineering with a focus on informatics for intelligent built environment from the University of Southern California (USC), Los Angeles, in 2013 and 2015, respectively. He is currently an Assistant Professor with the Via Department of Civil and Environmental Engineering, Virginia Tech. He is the author of more than 70 peer reviewed journal and conference publications in high impact journals and international ASCE and ACM conferences. He has also diverse experience in the design and construction of civil infrastructure systems as a Project Manager, a Design Manager, and a Design Engineer, from 2002 to 2009. His research looks at the intersection of cyber-physical systems, data-driven informatics, infrastructure operational analytics, and built environment sustainability and adaptability. Among his recent leadership efforts, he has Co-Chaired the Technical Program Committee of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys 2019) and organized and ran the first workshop on Building Blocks for Smarter Cities workshop, Arlington, VA, in 2019.