An energy-cyber-physical system for personalized normative messaging interventions: Identification and classification of behavioral reference groups

Kwonsik Song[a], Kyle Anderson[a], SangHyun Lee*[a]

[a] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, United States of America

*Corresponding author at: 2350 Hayward St., 2340 G.G. Brown Building, Ann Arbor, MI, United States of America; PH (734) 764-9420; FAX (734) 764-4292; email: shdpm@umich.edu

emails: kwonsiks@umich.edu, kyleand@umich.edu, shdpm@umich.edu

Abstract

Within residences, normative messaging interventions have encouraged households to engage in various pro-environmental behaviors. In norm-based intervention campaigns, it is hypothesized that more personally relevant reference groups increase norm adherence, thus improving the effectiveness of normative messaging interventions. Advanced energy grid infrastructure, such as smart meters and cloud computing, enables the creation of highly personalized behavioral reference groups in a non-invasive manner by dynamically classifying households into highly similar user groups based on usage patterns. Unfortunately, it remains unclear how readily available data on household energy use and housing characteristics affect the classification performance of dynamic behavioral reference groups. Therefore, this research evaluates the classification performance of dynamic behavioral reference groups using readily available data. An energy-cyber-physical system for personalized normative messaging interventions is trained and tested using one-year of energy use data from 2,248 households in Holland, Michigan. Dynamic behavioral reference group classification proved very accurate, 94.7-95.9% for weekly feedback and 89.9-93.1% for monthly feedback using only readily available data. In addition, using more historical energy use data contributes to enhancing classification accuracy. Lastly, high classification performance for each behavioral reference group is achieved at 97.6% of precision, recall and F1-score. With the proposed system, it is possible to dynamically assign highly personalized behavioral reference groups to households every billing cycle even if behavioral patterns are subject to change. Thus, interveners will be able to deploy personalized normative feedback messages on a large scale.

Keywords

Household Energy Consumption, Behavior Change, Normative Feedback, Behavioral Reference Groups, Energy-Cyber-Physical System

1. Introduction

In the United States, residential buildings account for approximately 21% of all energy expenditure and 19% of all carbon dioxide (CO_2) emissions, making them a prime target for energy reduction strategies [1]. As occupant behaviors substantially influence household energy expenditures, behavioral interventions attempting to promote more pro-environmental household behaviors have become widespread [2,3]. One advantage of behavior interventions aimed at reducing household energy consumption, compared to technological methods (e.g., retrofitting), is that behavioral intervention methods are often cost efficient and applicable to most if not all of the residential population [4].

A wide variety of behavioral intervention methods have been designed and implemented to increase pro-environmental behaviors, including reducing home energy use [5]. One prominent intervention method for reducing home energy use is behavioral feedback, which is inexpensive to implement and has repeatedly been found to be effective at inducing occupants to reduce their energy consumption [6-10]. Energy use feedback informs residents of their energy consumption (i.e., individual feedback), may be presented in numerous different forms (e.g., power in watts and cost [11]), and can include descriptive and/or injunctive normative feedback elements [12,13]. Descriptive normative feedback compares a household's energy consumption to a reference group, providing the household with the social norm of the group for home energy consumption. Injunctive normative feedback indicates a level of social approval or disapproval of the household's behavior.

In several previous field experiments, including normative feedback elements on messages resulted in energy savings between 5.4% and 8.9% [14,15]. In these studies, normative messages were created based on the weekly energy use of nearby households and given to households via a smartphone application. Psychologists hypothesize that when individuals are given normative information from more personally relevant reference groups, the persuasiveness of the message and norm adherence increases [16,17]. Currently, most normative feedback reference groups are based on geographical proximity (e.g., street, city, and distance) or less commonly on housing characteristics (e.g., housing size and heating type). Unfortunately, creating more personalized reference groups with high degrees in similarity with regards to location, behavior, composition, and/or housing characteristics have traditionally required either households' participation to collect the data (e.g., surveys and home energy audits) or prohibitively costly data collection. Thus, it has been financially infeasible to use highly personalized normative messages on a large scale.

Recently, the increased deployment of advanced energy grid infrastructure (e.g., smart energy meters and cloud computing systems) offers new opportunities to overcome some of these limitations. With advanced energy metering technology, it is possible to collect highly granular energy consumption data in a non-invasive manner without requiring active participation from residents. Highly granular and readily available consumption data can be used to construct home energy use profiles for each household. Energy use profiles inherently include information about how occupants behave in their home. These profiles, in conjunction with geographic information and basic household information (e.g., housing size) offer new opportunities to generate more personally relevant behavioral reference groups for use in normative feedback messaging campaigns [18-22]. With the addition of advanced computation systems (e.g., cloud computing), interveners can store and process a large volume of energy use data making it possible to develop and deploy scalable personalized normative messaging interventions.

To date, readily available energy use data have been widely used to create representative behavioral patterns of electricity consumers for energy tariff structure modelling [23], building energy use prediction [24] and renewable electricity generation [25]. Also, behavioral patterns of residential and commercial energy customers have been found to shift energy consumption from peak hours to off-peak hours in response to changes in the price of electricity [26-29]. In order to isolate unique behavioral patterns, clustering algorithms have been successfully applied across numerous studies to categorize the energy use profiles of households into several meaningful groups without any prior knowledge of the groups [23,30]. Before conducting a clustering analysis, the dimensionality of energy use data is typically reduced to improve clustering performance [23,31]. The most widely used data reduction techniques include changing the resolution of energy use data (i.e., time interval) [32,33] and projecting the original data into a lower dimensional subspace (e.g., principal component analysis [25,34]). Several research efforts have transformed the measured time-domain data to a different frequency domain data (e.g., Fourier coefficients [35,36] and wavelet coefficients [37,38]) during the data preprocessing step to increase the similarity in behavioral patterns of electricity consumers within each group. Once representative behavioral pattern groups have been identified from historical energy use data, classifiers assign new and existing energy consumers into the identified groups using newly collected consumption data and housing characteristics. This classification approach has been applied in personalized energy service marketing [39,40] and demand response programs [28].

While it has been demonstrated that household can be classified into meaningful groups at any given time, the performance of behavioral reference group classification over time remains unclear. It is important to understand how classification performs over time as households can exhibit different behavioral patterns in different billing cycles. If behavioral reference groups are fixed when there are changes in household behaviors over time, norm adherence would be expected to degrade, reducing the effectiveness of the normative messaging interventions. This makes it necessary to periodically reassess and reassign household behavioral reference groups.

Until recently, many studies have created representative behavioral patterns of households using readily available energy use data but have not dynamically classified households; in other words, they used fixed groups. Clustering households every billing cycle enables interveners to provide more personally relevant behavioral reference groups, but also requires the interveners' participation to examine the appropriateness of the identified groups (e.g., difference in behavioral patterns between groups). This repetitive clustering processes limits the scalability of the intervention.

To the best of the authors' knowledge, no studies to date have investigated the performance of dynamic behavioral reference group classification using readily available data over time, but instead have only considered using fixed groups [41,42]. In the most related attempt, Figueiredo et al. [43] proposed an electric energy consumer classifier using real-time meter data and collected the data from residential, commercial and industrial customers. In addition, Martinez-Pabon et al. [28] trained four different machine learning-based classifiers using smart meter data to predict target groups of residential and commercial customers who participate in demand response programs. However, as different energy sectors have different data characteristics, which can significantly affect the classification performance, it remains unclear how readily available energy use data from households affect classification performance.

To address this gap in the literature, this paper evaluates the classification performance of dynamic behavioral reference groups using readily available data. To achieve this objective, we propose an energy-cyber-physical system (e-CPS) to identify representative behavioral reference groups from historical energy use data and repeatedly classify households into the identified groups as new energy use data are received. To improve the usability of the proposed e-CPS for weekly and monthly energy use feedback, its classification accuracy is evaluated for every messaging cycle for one year. The proposed e-CPS enables interveners to assign highly personalized behavioral reference groups and personalized normative feedback messages to households every billing cycle in a scalable and non-invasive manner.

This paper is organized as follows. First, an e-CPS is developed using data-mining techniques in conjunction with readily available energy use data. This is followed by an overview of data collection. Then, the clustering and classification results using the proposed e-CPS are presented. This is followed by a discussion of the results and conclusions.

2. Energy-Cyber-Physical System for Personalized Normative Messaging Interventions

This research proposes a non-invasive scalable energy-cyber-physical system (e-CPS) to provide households with personalized normative feedback messages every billing cycle. The proposed system consists of two physical modules (i.e., data entry and messaging modules) that interact with two cyber modules (i.e., identification and classification modules) (Fig. 1). The *data entry module* collects energy consumption and housing characteristics data. Energy data is collected using smart metering technology; housing characteristics are collected from public records. Using readily available data, the *identification module* conducts a clustering analysis using one-year of historical energy use data to identify representative behavioral reference groups for each season. Next the historical data is labeled using the identified groups. The energy profiles of each representative group are then stored in a database as daily energy use profiles (i.e., labeled data). The *classification module* trains classifiers using historical daily energy use profiles and then predicts the behavioral reference groups for new data (i.e., predict daily energy use profiles for the last billing cycle for each household). Lastly, in the *messaging module*, the predicted behavioral reference groups are used to construct the personalized normative messages. The messages are then sent to each household.

< ***Fig. 1.*** *Energy-Cyber-Physical System (e-CPS) for personalized normative messaging intervention framework* >

In the next section, we provide the details of how the two cyber modules identify representative behavioral reference groups and allocate the identified groups to households every billing cycle.

2.1 Identification Module

To find representative behavioral reference groups across all seasons, this module conducts a clustering analysis using the previous year's energy use data and housing characteristics as follows. First, historical energy use data and housing characteristics are preprocessed to improve the clustering performance. Second, clustering algorithms are applied to the preprocessed data to generate *k* historical daily energy use profile patterns. Third, all the clustering results are evaluated in terms of group similarity to identify the optimal number of behavioral reference groups for each season. Fourth, the historical data is labeled using the best clustering configuration and then stored in a database as daily energy use profiles.

2.1.1 Data Preprocessing

First, households are divided into five groups based on their footprint (m^2). This makes for more meaningful and relevant normative comparisons since housing size is the most significant determinant of household energy use [44]. Square meter data is automatically scraped from public records. Households are grouped by size: HS1 (less than 92.9 m^2), HS2 (92.9 to 139.2 m^2), HS3 (139.3 to 185.7 m^2), HS4 (185.8 to 232.1 m^2), and HS5 (232.2 m^2 or more). Housing units are generally divided into seven housing size categories; however, using seven groups tends to have an unequally distributed number of households in each group, leaving some groups with few members, so five groups are used [45]. Behavioral reference groups with an insufficient number of members undermine the validity of normative comparisons within the group. For this reason, housing units with less than 92.9 m^2 and more than 232.2 m^2 are respectively combined into HS1 and HS5.

Next, daily energy use profiles are represented using six-hour intervals corresponding to morning (6 AM – 12 PM), afternoon (12 PM – 6 PM), evening (6 PM – 12 AM) and night (12 AM – 6 AM). Behavioral reference groups have been found to be the most distinguishable using slightly less granular data that represent larger periods (e.g., morning, afternoon, evening, and night) [30,46]. Using six-hour intervals for energy use data has the additional benefit of reducing data sizes and faster computing times.

Third, load shapes are extracted from the daily energy use profiles. In order for normative messages to be able to induce meaningful reductions in energy consumption, there must be non-trivial variation in household consumption within the peer reference group. In other words, when there is little difference between the norm and high users, there is little possibility for reductions in energy use as normative messaging campaigns attempt to get individuals to conform to social norms (i.e., reduce from high use to norm use). Therefore, it is important to be able to classify individual households by load shapes when they may vary in net consumption (Fig. 2-a). To enable this, three load shape (LS) extraction methods are applied:

< ***Fig. 2.*** *Daily energy use profile transformation using three load shape extraction methods* >

- LS1 (Gradient method): The rate of change in energy consumption (e_{rate}) is calculated by subtracting the original values of energy consumption (e) between two consecutive time points (Fig. 2-b). The original amount of energy $e_j(i)$ spent at time i for household j is transformed as follows.

$$e_{rate*j}(i) = e_j(i+1) - e_j(i) \tag{1}$$

where, $e_{rate*j}(i)$: the gradient in energy consumption at time i for household j, and $e_j(i+1)$: the original amount of energy consumption at time $i+1$ for household j.

- LS2 (Normalization method): Normalized energy consumption (e_{norm}) is calculated by transforming the original value of energy consumption (e) within a range of zero to one (Fig. 2-c). The mathematical terms of $e_{norm*j}(i)$ can be described by Eq 2.

$$e_{norm*j}(i) = \frac{e_j(i) - e_{min*j}}{e_{max*j} - e_{min*j}} \tag{2}$$

where, $e_j(i)$: the original amount of energy consumed at time i for household j, e_{min*j}: the minimum amount of energy consumption for household j, e_{max*j}: the maximum amount of energy consumption for household j.

- LS3 (Cumulative method): The cumulative percentage of energy consumption (e_{cper}) is calculated by dividing the cumulative energy consumption at each time by the total amount of daily energy consumption (Fig. 2-d). The mathematical terms of $e_{cper*j}(i)$ can be defined as follows.

$$e_{cper*j}(i) = \frac{e_{cum*j}(i)}{e_{d*j}} \times 100 \tag{3}$$

where, $e_{cum \cdot j}(i)$: the cumulative amount of energy consumption at time i for household j and e_{d*j}: the amount of daily energy consumption for household j.

2.1.2 Application of Clustering Algorithms

To create representative behavioral reference groups, a hierarchical clustering algorithm (HC) is applied to the preprocessed data set. A HC is used instead of the *k*-means algorithm (KC), a commonly used clustering algorithm for categorizing objects into several meaningful groups, for two main reasons. First, KC results are dependent on initial cluster centroid position for each iteration [47]. Second, several studies have shown that HC performs better than KC at categorizing households into several meaningful groups based on similarity in daily energy use patterns [23].

The HC clusters data objects by generating a hierarchy of nested partitions, called a dendrogram. The HC is implemented in the following four steps. First, a similarity matrix is constructed by calculating the distances among all the data objects. Second, each object is considered as a cluster. Third, the two clusters that are the closest in distance are merged until the total number of clusters reaches to one. Lastly, the dendrogram is cut at a certain level (i.e., k clusters) to determine the number of clusters.

2.1.3 Clustering Performance Evaluation

After applying the HC to the preprocessed historical data set, the clustering results are evaluated in terms of group similarity to find the best number of behavioral reference groups across all seasons. Various clustering evaluation criteria such as Silhouette Index (SI), Davies-Bouldin Index (DBI), Cluster Dispersion Indicator (CDI) have been used to determine the best number of groups in given datasets. The DBI is selected here due to its robust performance regardless of data properties (e.g., monotonicity, noise, density and skewed distributions) [48] and its accepted use in the field of residential energy consumer segmentation [23]. The DBI is based on a ratio of sum of within-cluster scatter to between-cluster separation [49] and is defined as follows:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j\neq i}\left\{\frac{\overline{d_i}+\overline{d_j}}{d_{ij}}\right\} \tag{4}$$

where k: the number of clusters; d_i: the average distance between all objects in the i_{th} cluster and the centroid of the i_{th} cluster; d_j: the average distance between all objects in the j_{th} cluster; and d_{ij}: the distance between the centroids of the i_{th} and j_{th} clusters. A lower DBI value indicates better clustering performance.

2.2 Classification Module

To classify households into one of the identified behavioral reference groups each billing cycle, this module performs the following two tasks. First, new and historic energy use data and housing characteristics (i.e., readily available data) are preprocessed in the same way as for the identification module. Next, we extract load indices as well as load shapes from daily energy use profiles for training the classifier. The details of load indices are described in the following section.

Second, this module trains and tests classifiers using the preprocessed historical data and then uses the trained classifiers every billing cycle to predict which group a household belongs to given their new consumption data.

2.2.1 Data Preprocessing

For each housing size category, new and historic daily energy use profiles are generated as described in the identification module. Then, two types of features (i.e., load shapes and load indices) are extracted from the daily energy use profiles since they influence classification performance. In general, features are used as input variables during the classifier training process. Previous studies to date have extracted load shapes [39] or load indices [39,40,50,51] from daily energy use profiles for classifying new electricity customers. Load indices here correspond to energy consumption characteristics for the four different time periods of the day and include: load factor (i_1), off-peak factor (i_2), night impact coefficient (i_3), lunch impact coefficient (i_4), and modulation coefficient for off peak hours (i_5). The mathematical terms of the load indices can be described by the following equations.

$$i_1 = {P_{av.day}}/{P_{max.day}} \tag{5}$$

$$i_2 = {P_{min.day}}/{P_{max.day}} \tag{6}$$

$$i_3 = {1}/{(3 \times P_{av.night} \times P_{av.day})} \tag{7}$$

$$i_4 = {}^{1}\!/\!_{(8 \times P_{av.lunch} \times P_{av.day})} \tag{8}$$

$$i_5 = {}^{P_{min.day}}\!/\!_{P_{av.day}} \tag{9}$$

where, $P_{av.day}$: average hourly energy consumption, $P_{max.day}$: maximum hourly energy consumption, $P_{min.day}$: minimum hourly energy consumption, $P_{av.night}$: average hourly energy consumption from 11 PM to 6 AM, $P_{av.lunch}$: average hourly energy consumption from 12 PM to 3 PM. The values for each load index are normalized using Min-Max normalization between zero and one since features with larger values can dominate other features.

2.2.2 Classifier Development

Classifiers are trained and tested using the preprocessed historical data. Since a classifier's performance varies depending on data characteristics, this study applies four standard classifiers to assign the identified behavioral reference groups to households. These classifiers have been widely used in the field of human activity recognition and are applicable for this objective [52,53]. Although state-of-the-art classifiers (i.e., deep learning-based classifiers) are able to solve complex classification problems (e.g., image recognition and natural language processing), they require significantly large datasets (i.e., a large number of data objects in a high dimensional space) to achieve a high classification performance [54]. Considering that households are represented in a very low dimensional space (i.e., four load shape features or five load index features), the following four traditional classifiers are sufficient to be able to classify new daily energy use profiles into the identified behavioral reference groups.

The first classifier, decision tree (DT), is a hierarchical model that recursively splits landslide conditioning factors into two classes (landslide and non-landslide) in terms of probability [55]. The second classifier, discriminant analysis (DA), is a statistical method for discriminating between categories of historical data based on the observed characteristics of a certain number of samples and discriminant criterion, according to their distance from the categories [56]. The third classifier, *k*-nearest neighbor (KNN), searches for the *k* most similar data points from the historical data to new data [57]. The similarity is typically calculated using distance measures such as Euclidean, Cosine, or Minkowski distances. The fourth classifier, support vector machine (SVM), uses kernel functions and hyperplanes as a decision boundary to separate historical data into two or more classes in a n-dimensional space where n corresponds to the total number of input variables [58].

During classifier training ten-fold cross validation is used which randomly divides the preprocessed historical data into ten folds for training and validation [59]. More specifically, the preprocessed historical data are divided into ten parts containing the same number of samples. In the first fold, 90% of the preprocessed historical data is used to train the classifiers and the remaining 10% is used for testing. In the second fold, the next 10% of the preprocessed historical data is used as a test dataset with the remaining 90% used as the training dataset. This process is repeated for the remaining eight folds and enhances the robustness of the trained classifiers.

The trained classifiers are evaluated using test data to identify the best classifier. Classification accuracy is the primary measure of classification performance. Classification accuracy is the percentage of correctly predicted observations (e.g., predicted vs actual group) with respect to the total number of observations (Eq. 10). This metric has been widely applied for classifying electricity customers [28,39,40]. In addition to raw accuracy, recall (Eq. 11), precision

(Eq. 12), and F1-score (Eq. 13) are also used to evaluate the classification performance for each class (i.e., behavioral reference group). These metrics are useful when classifiers are trained using imbalanced class data (i.e., different percentage of behavioral reference groups) [60,61]. The classification performance metrics are calculated using the following equations:

$$Accuracy = \frac{1}{total\ observations}\sum_{c=1}^{k} tp_c \tag{10}$$

$$Precision_c = \frac{tp_c}{tp_c + fp_c} \tag{11}$$

$$Recall_c = \frac{tp_c}{tp_c + fn_c} \tag{12}$$

$$F1 - score_c = 2 \times \frac{Precision_c \times Recall_c}{Precision_c + Recall_c} \tag{13}$$

where, tp_c: the number of correctly identified observations for class c, fp_c: the number of incorrectly identified observations for class c, fn_c: the number of incorrectly rejected observations for class c. After the classifier training and testing are complete, the newly collected energy use data is used as an input data of the trained classifier every billing cycle to classify households into one of the previously identified behavioral reference groups.

3. Data Collection

To evaluate the performance of the proposed e-CPS, energy consumption and housing size data are collected from 3,000 households in Holland, Michigan. Housing size data is gathered from the Holland Board of Public Works (HBPW). Electrical energy consumption data is obtained from smart meters at a 15-minute interval from January 1, 2016 through December 31, 2016. Heating degree days (HDD) and cooling degree days (CDD) vary substantially by season of the year in Holland, MI (Fig. 3). Corresponding to the changes in HDD and CDD during the year, individuals exhibited substantially different energy use patterns by season.

*< **Fig. 3**. Heating degree days, cooling degree days and the mean monthly energy use in kWh per household >*

Of the 3,000 households 503 exhibited abnormal values or had missing data. In addition, no information could be collected on the square meter of 319 households. This resulted in 752 households being excluded from the analysis. In total, 2,248 households are included in the analysis (Table 1).

*< **Table 1**. Number of Households by Housing Size >*

4. Results

Using the proposed e-CPS, the performance of behavioral reference group classification is evaluated in the MATLAB environment using the Classification Learner Toolbox as follows. First, a clustering analysis is performed using one-year of energy use data to find representative behavioral reference groups by housing size category. The number of clusters, *k,* is tested from two to ten across all housing size categories. The upper bound on the number of clusters is limited to ten; with more clusters it increases the chances that each cluster has an insufficient number of observations which will lead to invalid normative comparison. The hierarchical clustering algorithm is run using Euclidean distance as a metric and Ward linkage.

Second, the parameterization values of the four different classifiers (i.e., DT, DA, SVM, and KNN) can be seen in Table 2. The three DT-based classifiers are trained using a different maximum number of branch nodes. For the two DA-based classifiers, different types of discriminant functions are considered during its development. The six SVM-based classifiers are developed using different combinations of kernel functions and kernel scale. The six KNN-based classifiers are trained by adjusting the distance metric, the number of nearest neighbors, and the distance weighting function. Using these parameters, a sensitivity analysis is conducted using 4,000 randomly selected daily energy use profiles from each housing size category to select the best parameter values for each classifier. Across all the housing size categories, the best classifiers are: KNN1 (76.1%), DT1 (74.3%), DA2 (70.6%), and SVM4 (70.4%) (Fig. 4).

*< **Table 2**. Parameter values for decision tree (DT), discriminant analysis (DA), support vector machine (SVM) and k-nearest neighbor (KNN) >*

*< **Fig. 4.** Classification accuracy across different parameter values of classifiers (DT: decision tree, DA: discriminant analysis, SVM: support vector machine, KNN: k-nearest neighbor) >*

Next, each classifier is trained and tested 20 times using all the collected data to improve the reliability of the proposed e-CPS. The data is randomly split between training and testing with a 60/40 split. Once the best classifier is identified by accuracy percentage for daily energy use profile classification, it is further evaluated for recall, precision, and F1-score. Then, the performance of the best classifier is evaluated with changes in the number and type of training data because they may significantly affect the classification performance [62-64]. The training/testing data split proportions is then evaluated by increasing training to testing ratio from 60/40 to 90/10 by altering it five percent an iteration. These are respectively identified as TD60, TD65, TD70, TD75, TD80, TD85, TD90. The training data is broken down into two varieties: all inclusive (i.e., overall training data from the whole year) and seasonal only (i.e., seasonal training data which groups data by season). Further, to validate the usability of the trained classifiers for weekly and monthly energy use feedback, the predicted values are only considered correct if the predicted values are equal to the target values for every day during the entire period.

4.1 Clustering Results

4.1.1 Clustering Performance by Load Shape Extraction Method

LS3 produces the lowest average DBI values for the HS1, HS3, HS4 and HS5 of the three load shape extraction methods (Fig. 5). However, for the housing group HS2, there is no difference in the average values of DBI by load shape extraction method.

< ***Fig. 5.*** *Average values of Davies-Bouldin Index by load shape extraction method across different housing sizes* >

4.1.2 Behavioral Reference Group Identification

Fig. 6 shows a dendrogram of hierarchical clustering algorithms of the collected daily energy use profiles. Five behavioral reference groups are found to be optimal, producing the lowest DBI values, for HS1, HS2, HS3 and HS5 (Fig. 7). In housing group HS4, using six reference groups performs the best. The daily energy use profiles for the identified behavioral reference groups are unique and easily distinguishable (Fig. 8). While households in the Group 1 tend to consume the most energy in the afternoon and evening, households in the Group 5 tend to consume the most energy in the morning. These energy use profiles also vary by housing size category.

< ***Fig. 6***. *A dendrogram of hierarchical clustering algorithm of the collected daily energy use profiles* >

< *Fig. 7.* *Davies-Bouldin Index by number of clusters across different housing size categories* >

< *Fig. 8.* *Typical daily energy use profiles for identified behavioral reference groups across different housing size categories* >

The frequency of households grouped into each reference group fluctuates with the seasons. Group 1 is the most commonly assigned reference group throughout the year (HS1: 32.6-45.1%, HS2: 36.9-49.3%, HS3: 38.6-56.0%, HS4: 25.7-37.3%, HS5: 33.2-47.9%) (Fig. 9). Alternatively, some reference groups are assigned less than 10% of all households at some points during the year (HS1: Group 6, HS2: Group 5, HS3: Group 3, HS4: Group 4, HS5: Group 3).

< *Fig. 9.* *Number of daily energy use profiles across different housing size categories* >

4.2 Classification results

The KNN1 with load shapes (i.e., KNN1-LS) predicts the most accurate behavioral reference groups every day (average classification accuracy: 97.7%) (Fig. 10). The difference between the maximum and minimum classification accuracy is relatively low in the KNN1-LS compared to the other classifiers. When investigating the average classification accuracy by feature type across the different classifiers, load shapes produce higher average classification accuracy than using load indices: 95.2% to 44.9% for DT1, 89.4% to 39.3% for DA2, 96.8% to 19.8% for SVM4, 97.7% to 38.2% for KNN1. The distribution of prediction accuracy is relatively low when the classifiers are trained using load shapes. Looking at the average classification accuracy by classifier across the different feature types, DT1 has on average the best classification accuracy when using load indices (44.9% for DT1-LI vs 39.3% for DA2-LI, 19.8% for SVM4-LI and 38.2% for KNN1-LI). When using load shapes, KNN1 produces the most accurate classification results (97.7% for KNN1-LS vs 95.2% for DT1-LS, 89.4% for DA2-LS and 96.8% for SVM4-LS).

< ***Fig. 10.*** *Average classification accuracy by classifier* >

Further, KNN1-LS has an average precision of 97.6% and average recall of 97.6% (Fig. 11). Across all the identified behavioral reference groups, KNN1-LS achieves 97.3 to 97.7% precision and 97.3 to 97.7% recall. The F1-score of KNN1-LS is on average 97.6% and ranges from 97.3 % to 97.7% depending on the behavioral reference group being scored (Table 3).

< ***Fig. 11.*** *Precision and Recall of KNN1 for behavioral reference group classification* >

< ***Table 3***. *F1-score of behavioral reference group classification using KNN1* >

Additionally, for KNN1-LS as the percentage of data attributed to the training data increases the testing accuracy increases from 97.7% to 98.4% (Fig. 12-a). Looking at the average classification accuracy of KNN1-LS by type of training data, higher classification accuracy is observed when the KNN1-LS is trained using the complete training dataset regardless of season (Winter: 97.6% to 94.1%, Spring: 97.5% to 94.0%, Summer: 97.7% to 94.1%, Fall: 97.6% to 93.9%) (Fig. 12-b). Also, across the seasons, using overall the complete training dataset has lower variance in classification accuracy.

< ***Fig. 12.*** *Classification accuracy of KNN1-LS by number and type of training data* >

The KNN1-LS produces an average prediction accuracy of 94.7 to 95.9% across all the billing cycles for weekly feedback (Fig. 13). For the monthly energy use feedback, the average classification accuracy of the KNN1-LS ranges from 89.9 to 93.1%.

< ***Fig. 13.*** *Classification accuracy of KNN1 for weekly and monthly energy use feedback* >

5. Discussion

Training the classifiers with load shapes of daily energy use profiles (i.e., DT1-LS, DA2-LS, SVM4-LS, and KNN1-LS) produces the highest levels of classification accuracy. This can be explained by the fact that the identified behavioral reference groups are created based on similarity in load shapes of daily energy use profiles. Therefore, using load shapes would enable classifiers to better learn about the characteristics of the identified behavioral reference groups during its training process and to produce more accurate classification results (i.e., assign more personally relevant behavioral reference groups to households). Additionally, classifier performance is conditional on feature type. For instance, while DT1 produces the highest classification accuracy when using load indices, KNN1 performs the best when using load shapes. These results are in line with observations from previous studies which found that classification performance significantly differs with data characteristics [39].

When considering both classifier and feature type, it is clear that training the KNN1 classifier with load shapes (i.e., KNN1-LS) produces the most accurate behavioral reference group predictions. Compared to the other classifiers, the KNN1-LS has higher accuracy as well as a less variance (Fig. 10). These results can be explained by the fact that the KNN1-LS classifier searches for the nearest neighbor (i.e., the most similar load shapes in historical dataset) that is close in distance to the new observations (i.e., new load shapes) and then determines the neighbor's group as a group of new data [57]. As previous studies have achieved between a 74.0-88.3% [39,43] classification accuracy using smart meter data and/or survey data, this marks a significant improvement in behavioral reference group classification performance (Table 4).

*< **Table 4**. Comparison of classification accuracy between KNN1-LS and other classifiers for energy consumer classification >*

Moreover, KNN1-LS's performance is slightly correlated with the percentage of data used in the training set (Fig. 12-a). These results align with the previous studies that found increasing the proportion of training data accounts for the diversity of data objects (i.e., households) during the classifier training and thus improve classification accuracy [47,62]. Interestingly, training the KNN1-LS using the complete dataset produces better classification accuracy than the KNN1-LS using the seasonal training dataset (Fig. 12-b). These results do not concur with the previous findings that classifiers using a large data do not improve the prediction accuracy due to the usage of unnecessary data during its training process [63,64]. This discrepancy may be caused by the fact that the seasonal training data has almost the same proportion of behavioral reference groups across different seasons (Fig. 9). In other words, the overall training data doesn't have unnecessary data to train the KNN1-LS classifier even though it includes data for all the seasons.

Throughout the year, using readily available data makes it possible to have high classification accuracy over seven-and 28-consecutive day periods (Fig. 13). These results indicate that after investigating the most common behavioral patterns for the week or month, interveners are able to provide households with dynamically updated highly accurate personalized behavioral reference groups every billing cycle. This ability offers a unique opportunity; Ozawa et al. [65] suggested that uncommon behavioral patterns for the week or month can be used by households to recognize the dates and time when they consume more energy compared to average. Consequently, this outlier behavior could also be presented with the normative feedback in an attempt to help further promote energy savings behavior.

The proposed e-CPS has three main limitations in the identification and classification modules. First, representative behavioral reference groups from previous years may not be the ideal basis for effective normative comparison. This is possibly a limitation as housing characteristics can change over time, although widespread changes would likely not occur year to year. Therefore, as representative behavioral reference groups are updated periodically (e.g., every year), the classifiers should be retrained using the updated historical energy use profiles and groups.

Second, the developed classifiers that were identified to be the best may be climate and region specific. Without testing additional regions and climates the generalizability of them remains limited. In order to develop a generic classifier for all regions additional data from other climate regions is required [30]. Alternatively, classifiers for each climate region could be created and would likely perform better than a generalized classifier for all regions. Lastly, it is possible that imbalanced data classification issues may arise if representative behavioral reference groups are unevenly distributed (e.g., 1:100 among two groups); this can significantly compromise the performance of most classifiers [66]. In practice, many real-life classifier applications suffer having imbalanced data during its training (e.g., insider threat detection in organizations [67], medical diagnosis [68], face recognition [69]). Fortunately, significant attention has been applied to this problem and many potential solutions exist to balance the distribution of the original data (e.g., oversampling) and to modify existing classifiers to alleviate bias towards majority groups (e.g., different parameter weights) [70].

Future research efforts should investigate how households perceive their behavioral reference group, since norm adherence is correlated with identification with the reference group [71]. In addition, future work should examine how the time scale of social norm presentation (e.g., daily norm, weekly norm and monthly norm) affects households' actual energy use behaviors. This will provide insights into how to better design personalized normative feedback messages. Lastly, future studies should investigate how households change once they receive information about other group members' energy consumption. The proposed e-CPS will provide households with highly personalized normative feedback messages every billing cycle (e.g., "You used less electricity this month compared with the average Night Owl! Night Owls are people who exhibits peak energy use behavior late at night."). Nevertheless, some households may exhibit undesirable behavior changes. For example, Schultz et al. [72] discovered that when residents recognize that they used less energy than neighbors (i.e., descriptive norms), they increase energy consumption up to their group norms. This adverse effect of descriptive norm messages has been found to be prevented by including injunctive norms to the messages (e.g., a smiley face indicating approval of their behavior). On the other hand, the benefit of using injunctive norms was not found when normative feedback messages were delivered via email [7]. Therefore, it remains is necessary to investigate how highly personalized normative feedback messages affect households' behavior change.

6. Conclusions

In recent years, normative feedback messaging interventions have been gaining increased interest as a cost-efficient means to promote pro-environmental behaviors in residences. In energy use normative messaging feedback campaigns, households are given data on their own consumption as well the norm consumption of a reference group. Psychologists believe that the more personally relevant and meaningful the behavioral reference groups are to the recipient, the more likely the recipient is to adhere to the described social norm within it. Integrating homes with sensing, communication, and computation technologies enable a means to collect home energy use profiles on a grand scale. Categorizing households into meaningful reference groups based on similarities in their energy use profiles enables interveners to create more personally relevant normative messages.

This research bridges a gap in the literature on behavioral intervention strategies through the development of an energy-cyber-physical system for behavioral reference group classification using readily available energy use data. Testing the proposed system, three key results are found. First, using readily available energy use data can substantially improve behavioral reference group classification accuracy for weekly and monthly feedback. By using only readily available data, it is inherently scalable. Second, using longer histories of energy use data also contributes to enhancing the accuracy of classifying behavioral reference groups. Third, using readily available data yields high classification performance for each behavioral reference group.

More accurate classification is expected to increase the personal relevance of normative messages and increase norm adherence. This will in turn promote energy saving behaviors among residences. Further, the proposed system paves a path for accurate reference group assignment dynamically every messaging cycle even when household behaviors are subject to changes. In other words, this system permits to deployment of highly personalized normative messaging interventions on a large scale in a non-invasive manner.

Acknowledgement

The work presented in this paper was supported financially with a National Science Foundation Award (No. CBET-1705273) and in collaboration with Holland Board of Public Works (HBPW) of Holland at Michigan, U.S. Also, we thank Ms. Anne Saliers, Community Energy Services Manager of HBPW, for her assistance with the data collection. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Energy Information Administration (EIA). Total Energy: Energy Consumption by Sector, https://www.eia.gov/totalenergy/data/monthly/index.php; 2017 [accessed 30 March 2019].

[2] Bahaj AS, Myers L, James PAB. Urban energy generation: Influence of micro-wind turbine output on electricity consumption in buildings. Energy Build 2007;39:154-65.

[3] Yu Z, Fung BCM, Haghighat F, Yoshino H, Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy Build 2011;43:1409-17.

[4] Allcott H, Rogers T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. Am Econ Rev 2014;104:3003-37.

[5] Osbaldiston R, Schott JP. Environmental sustainability and behavioral science: Meta-analysis of proenvironmental behavior experiments. Environ Behav 2011;44:254-99.

[6] Allcott H, Rogers T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. Am Econ Rev 2014;104:3003-37.

[7] Anderson K, Song K, Lee S, Krupka E, Lee H, Park M. Longitudinal analysis of normative energy use feedback on dormitory occupants. Appl Energy 2017;189:623-39.

[8] Peschiera G, Taylor JE. The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems. Energy Build 2012;49:584-90.

[9] Darby S. The effectiveness of feedback on energy consumption: A review for DEFRA of the literature on metering, billing and direct displays. Environmental Change Institute, University of Oxford; 2006.

[10] Khosrowpour A, Jain RK, Taylor JE, Peschiera G, Chen J, Gulbinas R. A review of occupant energy feedback research: Opportunities for methodological fusion at the intersection of experimentation, analytics, surveys and simulation. Appl Energy 2018;218:304-16.

[11] Lehrer DR, Vasudev J, Kaam S. A usability study of a social media prototype for building energy feedback and operations, In: Proceedings of 2014 ACEEE Summer Study on Energy Efficiency in Buildings; 2014. p. 173-86.
[12] Allcott H. Social norms and energy conservation, J Public Econ 2011;95:1082-95.
[13] Ayres I, Raseman S, Shih A. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. J Law Econ Organ 2013;29:992–1022.
[14] Wemyss D, Castri R, Cellina F, de Luca V, Lobsiger-Kägi E, Carabias V. Examining community-level collaborative vs. competitive approaches to enhance household electricity-saving behavior. Energy Effic 2018;11:2057-75.
[15] Wemyss D, Cellina F, Lobsiger-Kägi E, de Luca V, Castri R. Does it last? Long-term impacts of an app-based behavior change intervention on household electricity savings in Switzerland. Energy Res & Soc Sci 2019;47:16-27.
[16] Festinger L. A theory of social comparison processes. Hum Relat 1954;7,117-40.
[17] Goldstein NJ, Cialdini RB, Griskevicius V. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. J Consum Res 2008;35:472–82.
[18] Capasso A, Grattieri W, Lamedica R, Prudenzi A. A bottom-up approach to residential load modeling. IEEE Trans Power Syst 1994;9:957–64.
[19] Richardson I, Thomson M, Infield D, Clifford C. Domestic electricity use: a high-resolution energy demand model. Energy Build 2010;42:1878–87.
[20] Widén J, Wäckelgård E. A high-resolution stochastic model of domestic activity patterns and electricity demand. Appl Energy 2010;87:1880–92.
[21] Shimoda Y, Asahi T, Taniguchi A, Mizuno M. Evaluation of city-scale impact of residential energy conservation measures using the detailed end-use simulation model. Energy 2007;32:1617–33.
[22] Tanimoto J, Hagishima A, Sagara H. Validation of methodology for utility demand prediction considering actual variations in inhabitant behavior schedules. J Build Perform Simul 2008;1:31–42.
[23] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 2012;42:68-80.
[24] Song K, Kwon N, Anderson K, Park M, Lee H, Lee S. Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups. Energy Build 2017;156:121-33.
[25] Motlagh O, Paevere P, Hong TS, Grozev G. Analysis of household electricity consumption behaviours: Impact of domestic electricity generation. Appl Math Compu 2015;270:165-78.
[26] Afzalan M, Jazizadeh F. Residential loads flexibility potential for demand response using energy consumption patterns and user segments. Appl Energy 2019;254:113693.
[27] Auder B, Cugliari J, Goude Y, Poggi J. Scalable clustering of individual electrical curves for profiling and bottom-up forecasting. Energies 2018;11:1893.

[28] Martinez-Pabon M, Eveleigh T, Tanju B. Smart meter data analytics for optimal customer selection in demand response programs. Energy Procedia 2017;109:49-59.
[29] Zhou D, Balandat M, Tomlin C. Residential demand response targeting using machine learning with observational data. Decision and control (CDC), In: 2016 IEEE 55th conference on. IEEE; 2016. p. 6663-8.
[30] Kwac J, Flora J, Rajagopal R. Households energy consumption segmentation using hourly data. IEEE Trans Smart Grid 2014;5:420-30.
[31] Rajabi A, Li L, Zhang J, Zhu J, Ghavidel S, Ghadi MJ. A review on clustering of residential electricity customers and its applications. In: 20th International conference on electrical machines and systems (ICEMS); 2017.p. 1–6.
[32] Granell R, Axon CJ, Wallom DCH. Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles. IEEE Trans Power Syst 2015;30:3217-24.
[33] Notaristefano A, Chicco G, Piglione F. Data size reduction with symbolic aggregate approximation for electrical load pattern grouping. IET Gener Transm Distrib 2013;7(2):108-17.
[34] Ferraro P, Crisostomi E, Tucci M, Raugi M. Comparison and clustering analysis of the daily electrical load in eight European countries. Electr Power Syst Res 2016;141:114-23.
[35] Al-Otaibi R, Jin N, Wilcox T, Flach P. Feature construction and calibration for clustering daily load curves from smart-meter data. IEEE Trans Ind Inform 2016;12(2):645-54.
[36] Kotzur L, Markewitz P, Robinius M, Stolten D. Impact of different time series aggregation methods on optimal energy system design. Renew Energy 2018;117:474–87.
[37] Mets K, Depuydt F, Develder C. Two-stage load pattern clustering using fast wavelet transformation. In: IEEE trans smart grid 2016;7(5):2250-9.
[38] Jiang Z, Lin R, Yang F, Wu B. A fused load curve clustering algorithm based on wavelet transform. IEEE Trans Ind Inf 2018;14(5):1856-65.
[39] Viegas J. Vieira SM, Melício R, Mendes VMF, Sousa JMC. Classification of new electricity customers based on surveys and smart metering data. Energy 2016;107:804-17.
[40] Vercamer D, Steurtewagen B, den Poel DV, Vermeulen F. Predicting consumer load profiles using commercial and open data, IEEE Trans power syst 2016;31:3693-701.
[41] Ozawa A, Furusato R, Yoshida Y. Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles. Energy Build 2016;119:200-10.
[42] Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans Smart Grid 2016;7:136-44.
[43] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst. 2005;20:596-602.

[44] Santin OG, Itard L, Visscher H. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. Energy Build 2009;41:1223-32.
[45] Energy Information Administration (EIA). Residential Energy Consumption Survey (RECS): 2015 RECS Survey Data, https://www.eia.gov/consumption/residential/index.php; 2015 [accessed 30 March 2019].
[46] Smith BA, Wong J, Rajagopal R. A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting. In: ACEEE Summer Study on Energy Efficiency in Buildings; 2012.
[47] Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques, third ed., Morgan Kaufmann, Waltham, 2012.
[48] Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation measures. In: 2010 IEEE International Conference on Data Mining. IEEE; 2010. p. 911–6.
[49] Davies DL, Bouldin DW. A cluster separation measure, IEEE Trans Pattern Anal Mach Intell 1979:2;224–7.
[50] Chicco G, Napoli R, Postolache P, Scutariu M, Toader C. Customer characterization options for improving the tariff offer. IEEE Trans Power Syst 2003;18(1):381e7.
[51] Ramos S, Duarte JM, Duarte FJ, Vale Z. A data-mining-based methodology to support MV electricity customers characterization. Energy Build 2015;91:16-25.
[52] Zebin T, Scully PJ, Ozanyan KB. Evaluation of supervised classification algorithms for human activity recognition with inertial sensors. In: Proceedings of 2017 IEEE Sensors; 2017. p. 1-3.
[53] Cha SH, Seo J, Baek SH, Koo C. Towards a well-planned, activity-based work environment: Automated recognition of office activities using accelerometers. Build Environ 2018;144:86-93.
[54] Chen Z, He N, Huang Y, Qin WT, Liu X, Li L. Integration of a deep learning classifier with a random forest approach for predicting malonylation sites. Genom Proteom Bioinform 2018;16(6):451-9.
[55] Ruggieri S. Efficient C4.5 [classification algorithm]. IEEE Trans Knowl Data Eng 2002;14:438-44.
[56] Zhang W, Li X, Gong F. Stability classification model of mine-lane surrounding rock based on distance discriminant analysis method. J Cent South Univ Technol 2008;15:117–20.
[57] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13: 21–7.
[58] Kecman V. Learning and soft computing. Cambridge, M.A.: MIT Press; 2001.
[59] Kohavi R. A study of cross validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence; 1995;2. p. 1137-45.
[60] Saito T, Rehmsmeier M. The Precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 2015;10.

[61] Jeni LA, Cohn JF, de La Torre. Facing imbalanced data-Recommendations for the use of performance metrics. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; 2013. p. 245-51.

[62] Li D, Menassa CC, Kamat VR. Personalized human comfort in indoor building environments under diverse conditioning modes. Build Environ 2017;126:304-17.

[63] Zhang S, Li X, Zong M, Zhu X, Cheng D. Learning k for kNN classification. ACM Trans Intell Syst Technol 2017;8:43.

[64] Aci M, Avci M. K nearest neighbor reinforced expectation maximization method. Expert Syst Appl 2011;38:12585-91.

[65] Ozawa A, Furusato R, Yoshida Y. Tailor-made feedback to reduce residential electricity consumption: The effect of information on household lifestyle in japan. Sustain 2017;9:528.

[66] He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng 2009;21:1263-84.

[67] Azaria A, Richardson A, Kraus S, Subrahmanian VS. Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data. IEEE Trans Comput Soc Syst 2014;1:135-55.

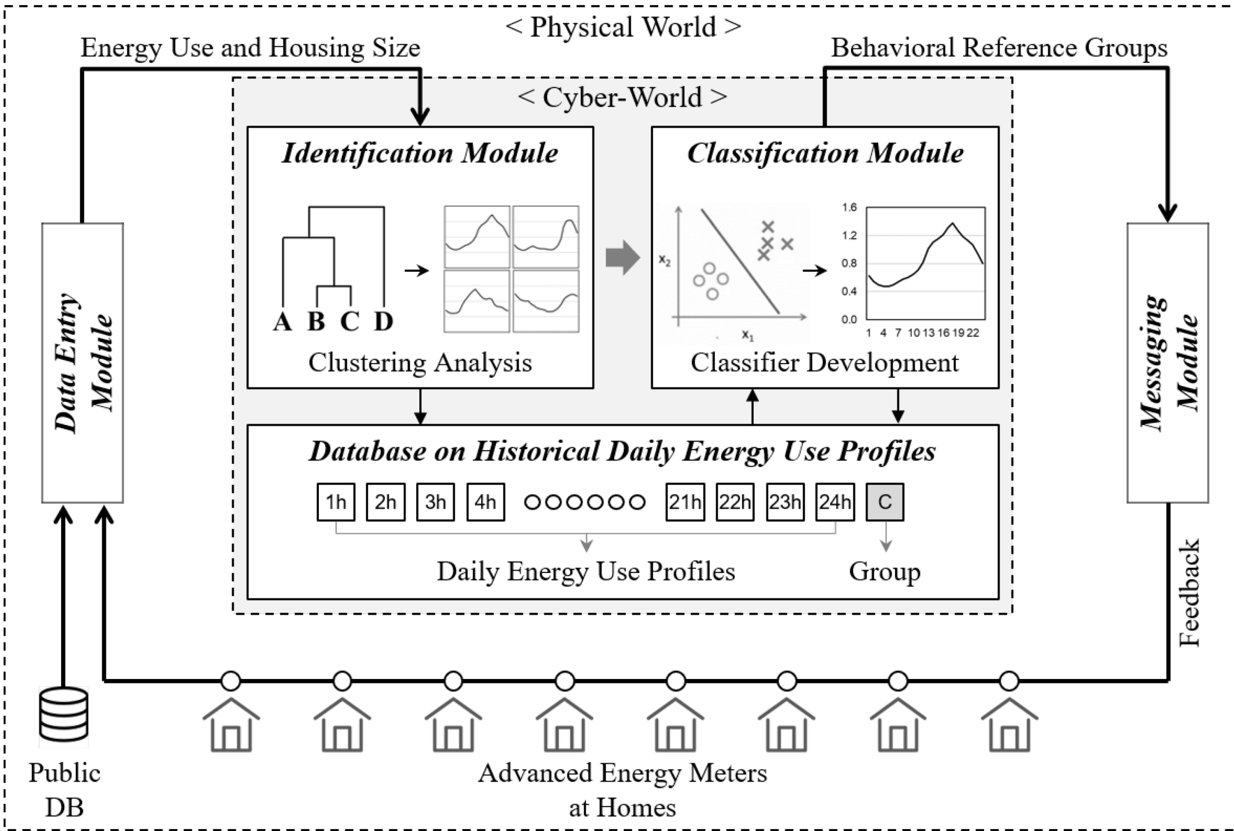
[68] Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. Neural Netw 2008;21:427-36.

[69] Liu YH, Chen YT. Total margin-based adaptive fuzzy support vector machines for multiview face recognition. In: Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics; 2005. p. 1704-11.

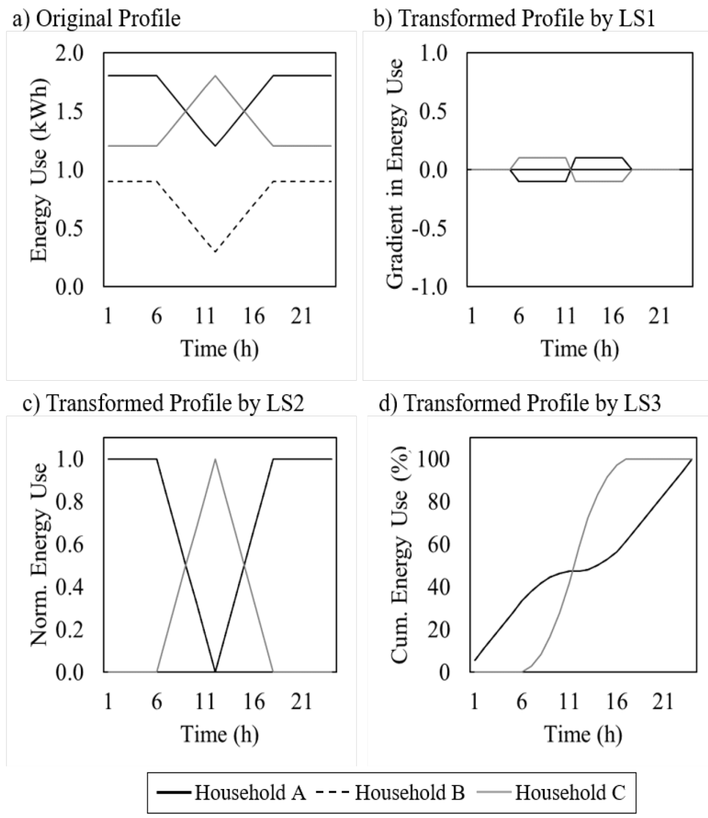
[70] Krawczyk B. Learning from imbalanced data: Open challenges and future directions. Prog Artif Intell 2016;5:221-32.

[71] Hogg MA. Social identity theory. In: McKeown S, Haji R, Ferguson N, editors. Understanding peace and conflict through social identity theory, Cham: Springer International Publishing; 2016, p. 3-17.

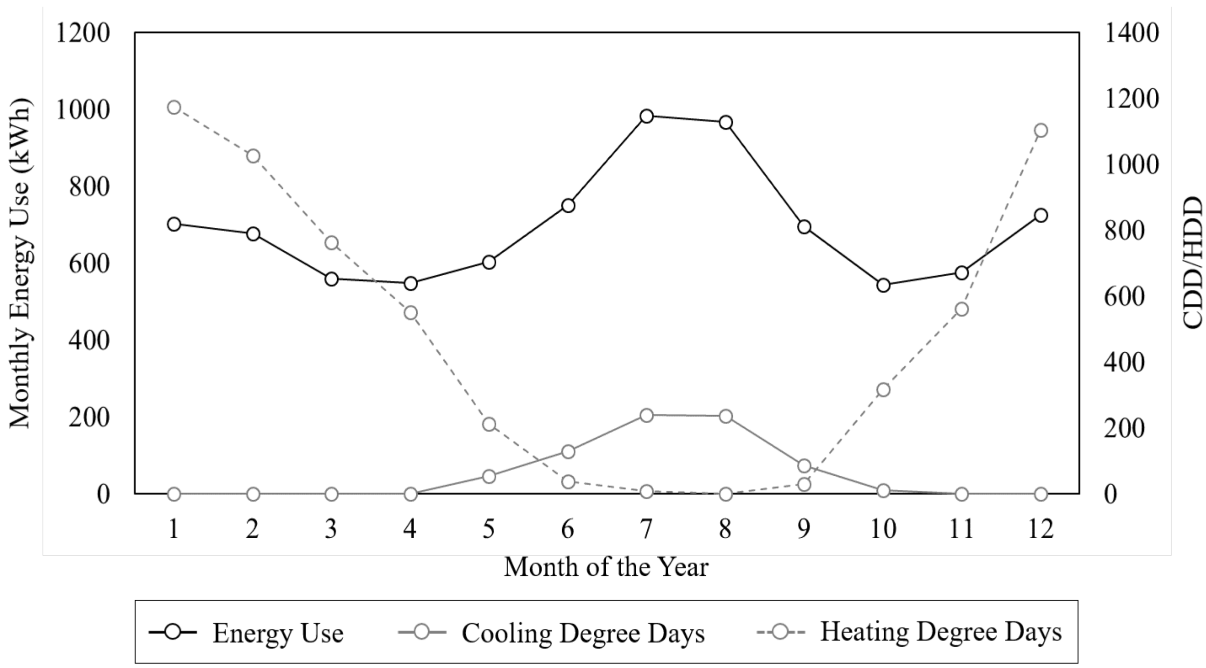
[72] Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychol Sci 2007;18(5):429–34.



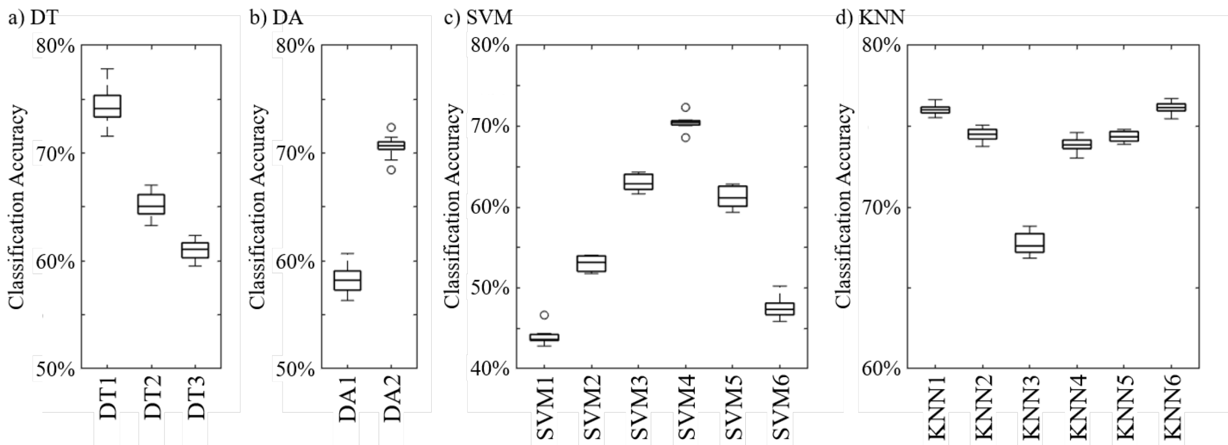
< **Fig. 1.** Energy-Cyber-Physical System (e-CPS) for personalized normative messaging intervention framework >



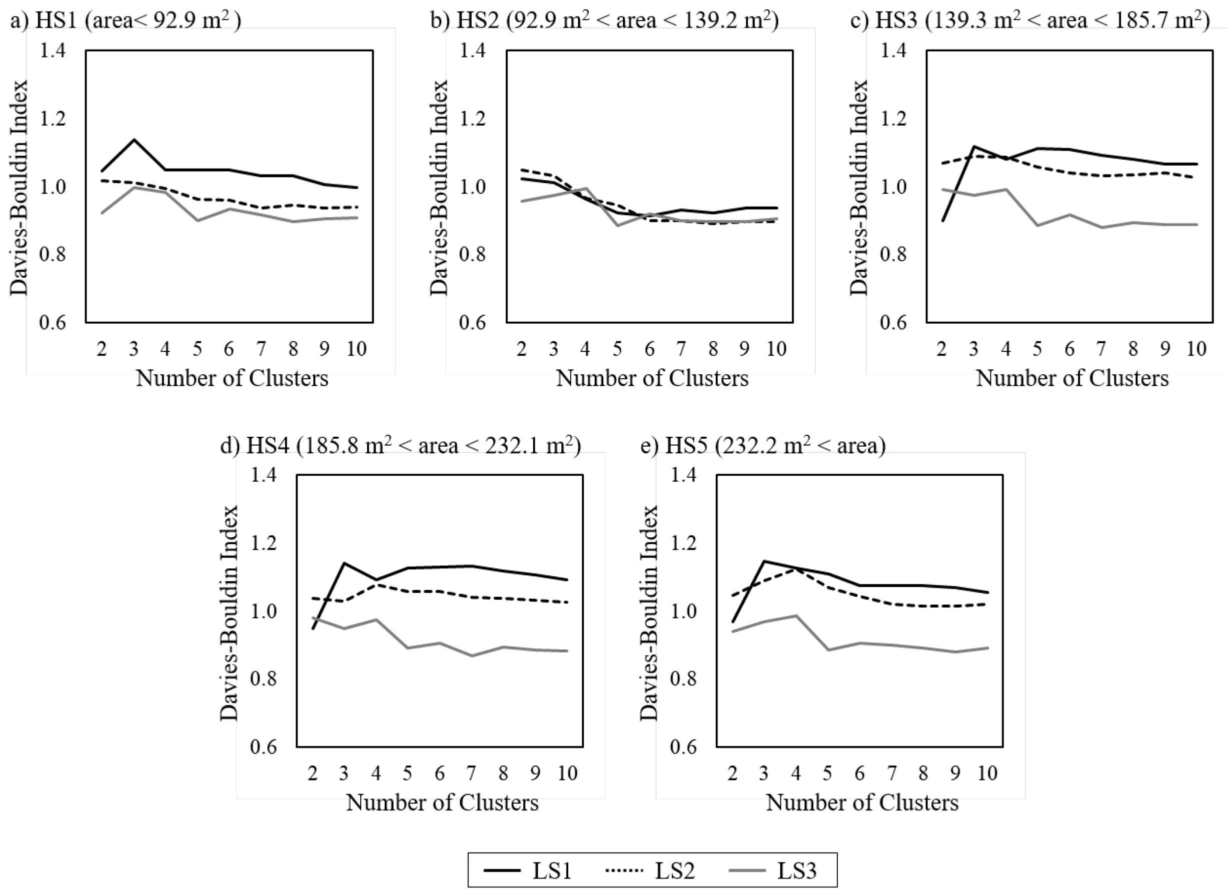
< **Fig. 2.** Daily energy use profile transformation using three load shape extraction methods (LS1: gradient method, LS2: normalization method and LS3: cumulative method) >



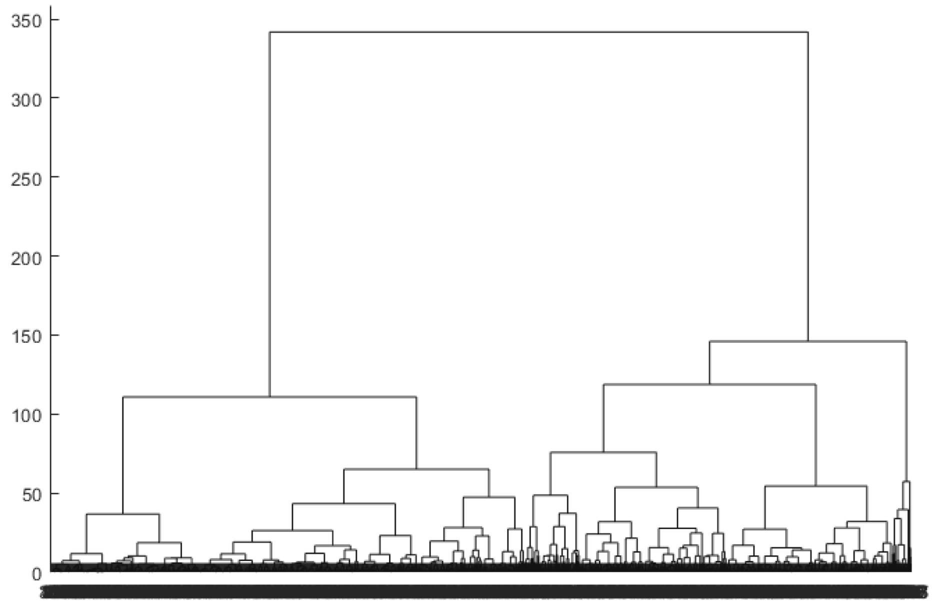
< **Fig. 3**. Heating degree days, cooling degree days and the mean monthly energy use in kWh per household >



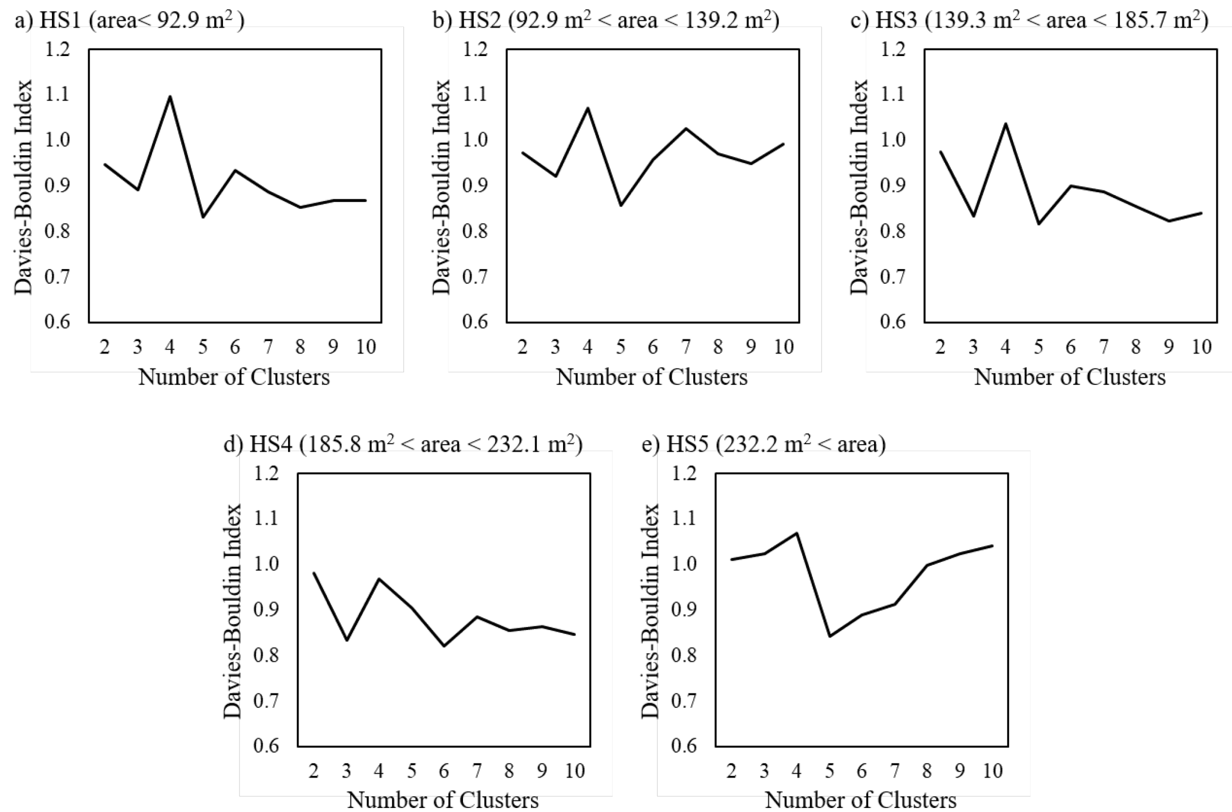
< Fig. 4. Classification accuracy across different parameter values of classifiers (DT: decision tree, DA: discriminant analysis, SVM: support vector machine, KNN: *k*-nearest neighbor) >



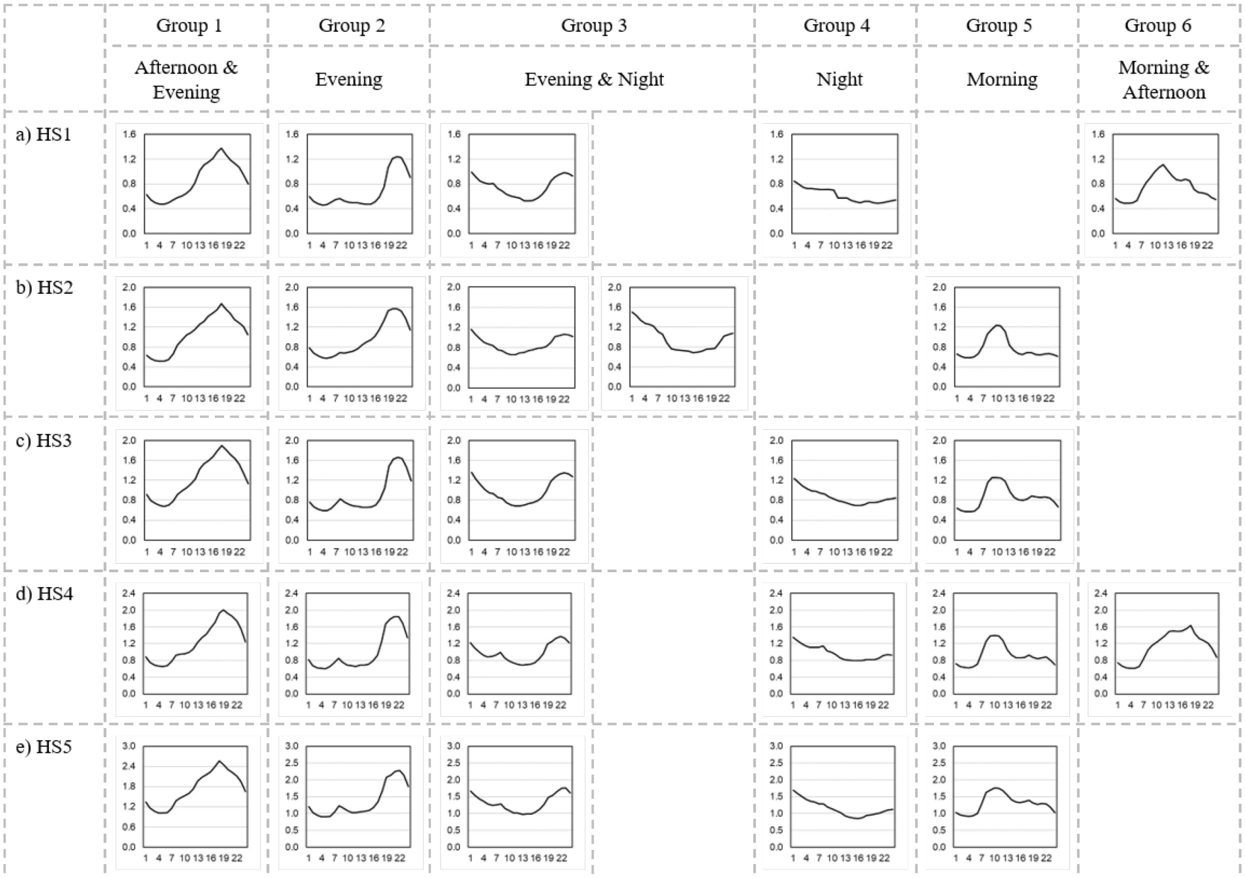
< Fig. 5. Average values of Davies-Bouldin Index by load shape extraction method across different housing sizes (LS1: gradient method, LS2: normalization method and LS3: cumulative method) >



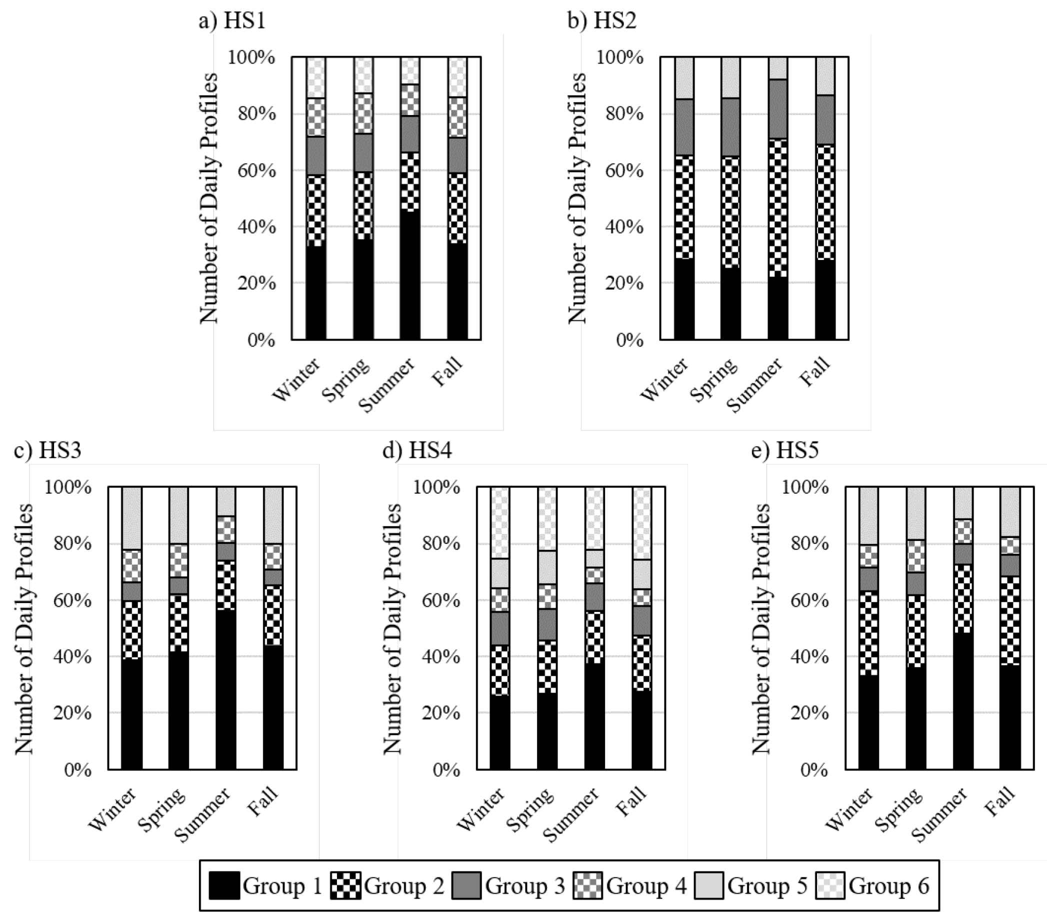
< Fig. 6. A dendrogram of hierarchical clustering algorithm of the collected daily energy use profiles >



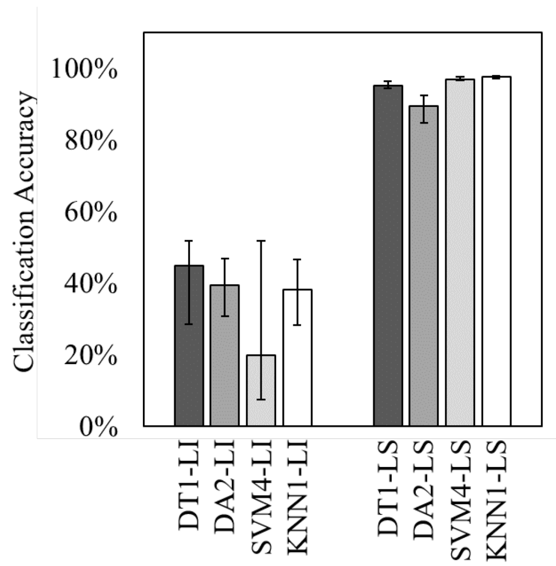
< **Fig. 7.** Davies-Bouldin Index by number of clusters across different housing size categories >



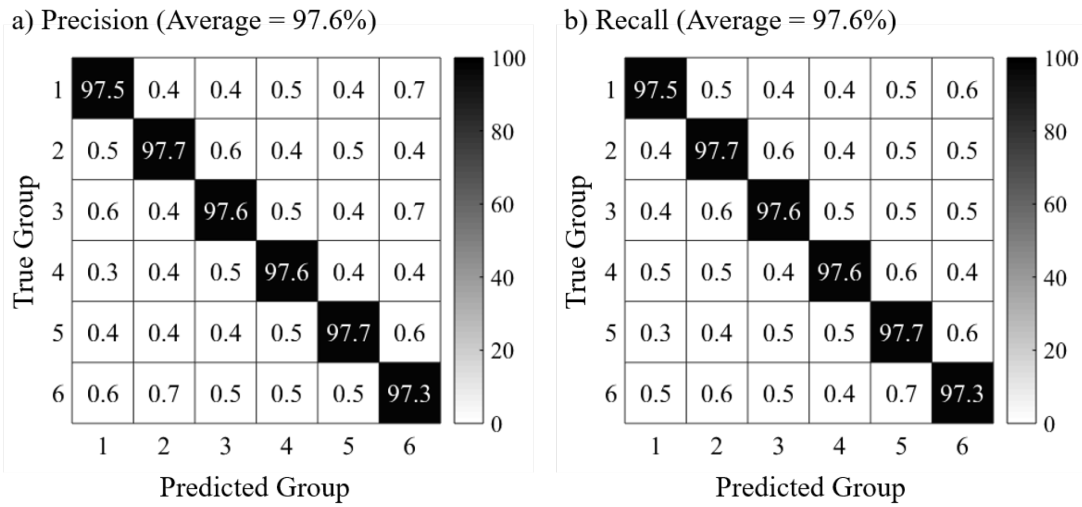
< **Fig. 8.** Typical daily energy use profiles for identified behavioral reference groups across different housing size categories >



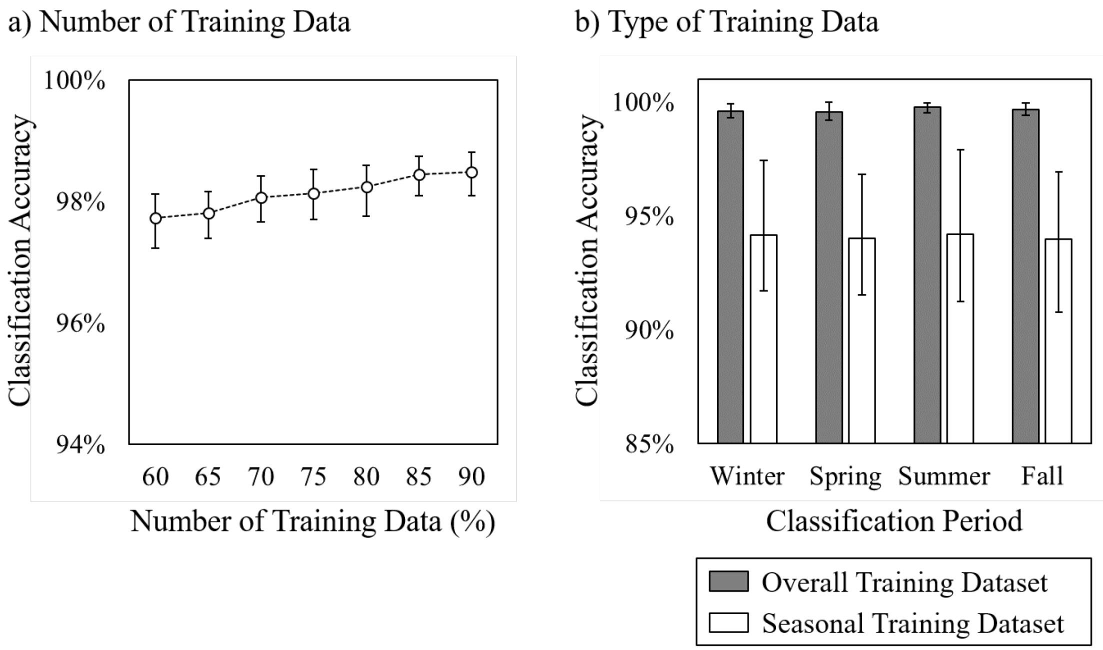
< **Fig. 9.** Number of daily energy use profiles for each behavioral reference group across different housing size categories and seasons >



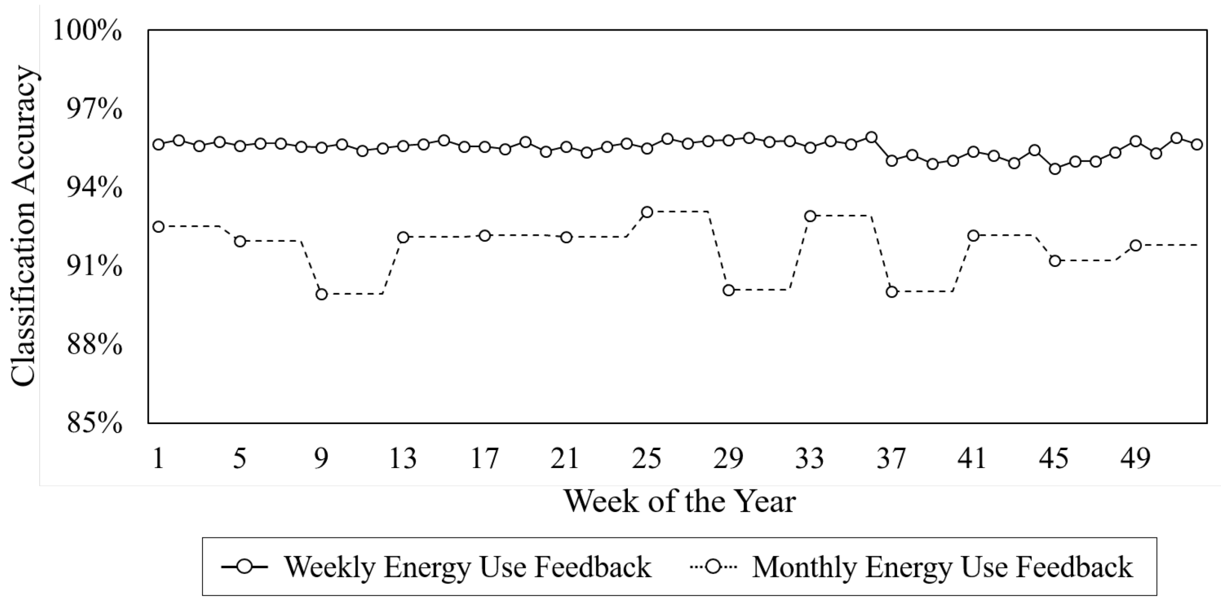
< **Fig. 10.** Average classification accuracy across different feature types by classifier >



< Fig. 11. Precision and Recall of KNN1-LS for behavioral reference group classification >



< Fig. 12. Classification accuracy of KNN1-LS by number and type of training data >



< Fig. 13. Classification accuracy of KNN1-LS for weekly and monthly energy use feedback >

< **Table 1**. Number of Households by Housing Size >

Housing Size Category	# of Households
HS1 (Less than 92.9 m^2)	525
HS2 (92.9 to 139.2 m^2)	1,083
HS3 (139.3 to 185.7 m^2)	435
HS4 (185.8 to 232.1 m^2)	132
HS5 (232.2 m^2 or more)	73
Total	2,248

< **Table 2**. Parameter values for decision tree (DT), discriminant analysis (DA), support vector machine (SVM) and k-nearest neighbor (KNN) >

Classifier		Parameter setting
DT	DT1	100 branch nodes
	DT2	20 branch nodes
	DT3	4 branch nodes
DA	DA1	Linear discriminant analysis
	DA2	Quadratic discriminant analysis
SVM	SVM1	Linear kernel function, kernel scale = auto
	SVM2	Polynomial kernel function of order 2, kernel scale = 0
	SVM3	Polynomial kernel function of order 3, kernel scale = 0
	SVM4	Gaussian kernel function, kernel scale = 0.5
	SVM5	Gaussian kernel function, kernel scale = 2
	SVM6	Gaussian kernel function, kernel scale = 8
KNN	KNN1	Euclidean distance, 1 nearest neighbor, equal distance weighting
	KNN2	Euclidean distance, 10 nearest neighbors, equal distance weighting
	KNN3	Euclidean distance, 100 nearest neighbors, equal distance weighting
	KNN4	Cosine distance, 10 nearest neighbors, equal distance weighting
	KNN5	Minkowski distance, 10 nearest neighbors, equal distance weighting
	KNN6	Euclidean distance, 10 nearest neighbors, square inverse distance weighting

< **Table 3**. F1-score of behavioral reference group classification using KNN1-LS >

Classifier	Behavioral Reference Group						Average
	1	2	3	4	5	6	
KNN1	97.5%	97.7%	97.6%	97.6%	97.7%	97.3%	97.6%

< **Table 4**. Comparison of classification accuracy between KNN1-LS and other classifiers for energy consumer classification >

Classifier	Sector	Meter Data	Survey Data				Accuracy
			Household	Housing	HVAC	Appliance	
KNN1-LS	Residential	O					97.8%
Viegas et al. [39]	Residential	O	O	O	O	O	80.8-88.3%
Figueiredo et al. [43]	Multiple	O					74~81%