

Research Article

A Unified Framework for Behaviour Monitoring and Abnormality Detection for Smart Home

Anupam Grewal ¹, Maninder Kaur ¹, and Jong Hyuk Park ²

¹Computer Science & Engineering Department, Thapar Institute of Engineering and Technology (Deemed University), Patiala (Punjab), India

²Department of Computer Science and Engineering, Seoul National University of Science and Technology (SeoulTech), Seoul 01811, Republic of Korea

Correspondence should be addressed to Jong Hyuk Park; jhpark1@seoultech.ac.kr

Received 18 December 2018; Revised 18 April 2019; Accepted 7 May 2019; Published 20 May 2019

Academic Editor: Davide Mattera

Copyright © 2019 Anupam Grewal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the revolution in smart infrastructure in the recent past, the smart healthcare system has been paid more considerable attention. The continuous upgradation of electricity meters to smart electricity devices has probed into a new market of intelligent data analysis services, providing aid to the health care systems. This paper presents a unified framework for extracting user behaviour patterns from home-based smart electricity meter data. The structure allows exploration and integration of frequent pattern growth algorithm for pattern mining and application of a variety of machine learning algorithms for categorizing the activities into manually labelled classes along with the implementation of Local Outlier Factor method for detection of an abnormal pattern of the inhabitant of smart homes. To evaluate the proposed framework, the work is implemented on the smart electricity dataset from the United Kingdom by separating the data into four distinct data files meant for the morning, afternoon, evening, and night energy utilization records. The results show a remarkable performance of Support Vector Machine (SVM) and Multilayer Perceptron (MLP) classifiers with kappa statics greater than 0.95 for all time slots data. The resultant frequent device utilization patterns with anomaly score more than the threshold value, reflecting abnormal activity patterns, are found more in evening time data in comparison to other time slots, requiring the immediate attention of concerned healthcare authorities.

1. Introduction

With the emergence of new technologies in the last century, the world has changed drastically. In the last decade, smart devices have been deployed all around the world. Several digital technologies enhance the productivity and well-being of cities by optimizing the use of various resources like electricity, water, governance, healthcare, etc. [1–5]. The smart meters are one of the main aspects of infrastructure modernization of all the time in the field of energy distribution. Many countries throughout the world are rolling out smart electricity meters for recording the electricity consumption patterns. These meters with associated devices are the main elements of intelligent digital infrastructure that record more detailed electricity consumption information in contrast to classic electric meters [6]. As smart meter monitors electricity consumption in real time, these devices provide short time

and long-term benefits to consumers and utilities. Billing is no longer the only utility of smart meters. The meter also offers high-resolution data about energy behaviour and consumer lifestyle. Recently, smart meters are rolled out in the whole world on a large scale.

In the smart society, every house outfitted with several intelligent meters, communicate their readings to the main portal, which sequentially transmit data to the community center through the community network in a secure and privacy-friendly way. The society center keeps up an account, storing comprehensive statistics on the use of facilities for each household. The data is periodically transferred to the call center, which directs it to the utility enterprise for the computer bills. As there is no direct individual involvement in the facility management process, the operational cost is decreased by the service company. On the consumer side, the privacy and security of the home have been improved because

no measurement personnel have entered their homes. Smart meters act as a two-way interface with client-side devices via the Home Area Network. These provide gateway or point of contact to the home for important information such as price changes or emergency notification.

The potential of smart meter energy data can be utilized for one of the most important domains of basic living infrastructure for states as well as for the individual's healthcare system. As the population is increasing exponentially in recent years, there is a dire need of some digital solution to offer better healthcare services to people in a more responsive, personalized and cost-effective manner. Smart meters record the energy usage information in coordination with time. The data from a smart meter can be remodeled in numerous ways to utilize it in different fields such as accurate billing, optimization of energy consumption, and optimized network management [7]. These readings can be used to study energy consumption by each device as well as the user activity patterns.

As the individual's behaviour is narrowly associated with the health status of the human on a daily basis, the recurrent habits of a person can be inferred by investigating the actions of everyday livelihood in terms of commencement time, time period, and recurrence rate. The routine energy utilization patterns can provide insights on how the inhabitants stick to their usual behaviour of energy utilization and induce more information regarding their normal day-to-day activities. If any change in human behaviour is spotted, actions can be taken for health assessment.

1.1. Motivation. Detection of human activities in smart homes can be performed using sensors technology and smart meters.

The following points that motivated us to conduct this study are as follows:

(i) Presently, the activities of daily living (ADL) are recognized using sensors, wearable bands, and CCTV cameras, etc. Sensors have a high price associated with them, due to which it is not possible to extend the use of sensors for activity recognition on a large scale.

(ii) Wearable bands and CCTV's affect users' privacy, due to which many people deny to adopt them in their day-to-day life. In such a scenario, the use of some passive approach like the smart meter data for activity recognition appears to be a decent plan.

(iii) No work has been done so far in abnormal patterns detection using anomaly detection techniques.

(iv) Energy consumption data is not utilized in the form of time slots, which is an effective approach as compared to 24-hour time frames.

The aim of the study is to analyse the inhabitant's activity patterns to understand normal behaviour and to detect anomalous activities, which directly indicate health problems.

1.2. Research Contribution. This paper proposes a model, which is able to analyse and observe the readings of a smart meter to recognize activity patterns and reveal the changes in electricity consumption patterns to indicate the

diversion of normal behaviour of an individual to abnormal. Electricity data used here is disaggregated data in separate files that provide the energy consumption data of each device separately [8]. The present work deals with energy consumption data of various devices from a single house. Electricity usage data is analysed to find the patterns of switched on devices, which are further correlated to the day-to-day activities such as preparing food, listening to music, etc. For example, if the coffee maker is on it means, activity performed is "preparing food." Further, the phase of the day whether the operation is active in morning or evening further tells about meal type whether it is breakfast or dinner. Usually, an individual performs different tasks simultaneously such as "Cooking meal" and "Listening to the Music", which signifies multiple devices, are operational at the same time.

To gain further insights, frequent patterns are mined to trace different sequences of devices operating simultaneously. The data from smart meters is mined in four-time slots. These time slots represent different parts of day, i.e., morning afternoon, evening, and night. The results of these mining are stored in separate databases. The proposed techniques are implemented individually on each database. These patterns are used to train the model to detect activities.

The contribution of this paper is as follows:

(i) The model for human activity mining which is conferred that can analyse the correlation between electricity usage data and patterns of the device usage. The work employs the FP-growth algorithm for pattern mining and multiclass classification for segmenting the activities into normal daily life activities.

(ii) Implementation of multiclass classification in association with frequent patterns to correlate individual's activities and the temporary status of the individual's mind.

(iii) Implementation of Anomaly Recognition model, for trailing out those activities that do not match with normal activity patterns.

1.3. Organization. The rest of the paper is organized as follows: Section 2 reflects the related work; Section 3 discusses the details and theoretical background of the proposed model; Section 4 sums up the results of the proposed model against various performance metrics, and Section 5 presents the conclusion of the research work.

2. Related Works

With the emergence of smart technologies, every field is being digitized. As a result, monitoring healthcare using smart infrastructure data conjointly has become prevalent. This section presents the review of the literature, in context with the existing work done in this field till date. Wu et al. [9] presented a system for device-free activity monitoring, TW-SEE, that recognizes human activities based on Wi-Fi signals. It uses two techniques named robust PCA and sliding window algorithm to segment activities. The study by Passerini et al. [10] focused on finding abnormalities in electrical systems like network fault, cable degradation, etc. In this work, the authors presented different techniques to harvest anomalies. The study by Wang et al. [11] is based on

activity recognition based on sensors. They utilized sensor-based low-level readings to find patterns. The work presented the survey of deep learning based unsupervised and incremental learning based techniques for activity recognition from three different aspects: the application of study, deep model, and sensory modality. Malasinghe et al. [12] presented a study based on remote health monitoring of patients. The study provided the review of advancements occurred in the field of remote health monitoring with the use of contact-based and contactless devices. The study by Li et al. [13] presented the technique to read digital fundus images to find symptoms of diabetic retinopathy using deep convolutional neural network. Miguez et al. [14] specified, implemented, and validated the smart environment based on artificial intelligence. In this work, the system identified the activities using temporal logic. Donso et al. [15] presented a study that focused on ambient assisted living. The study targets on detecting those events in which elder residents of the home are at the risk of entering dangerous areas, fall detection, etc. Here robots are employed that detect potentially dangerous and unnoticed areas. The captured information is then sent to the AAL system to improve living conditions. Zhang et al. [16] presented an energy disaggregation algorithm based on hourly smart meter readings. They used clustering and optimization techniques to break down the energy into different load categories based on components having different power factors. Nweke et al. [17] proposed the combined functioning of different types of living activities of daily life. They applied deep learning based classifiers for human activity recognition. Chalmers et al. [18] recognized sudden changes in the behaviour of patients suffering from clinical depression, Parkinson disease, and Alzheimer disease. The work utilized a neural network-based approach to detect abnormalities in the behaviour of people. M.S. Hossain [19] proposed a recognition system based on tasks performed by patients of some particular disease for healthcare framework using Fourier Transformation, grey level conversion, and log-likelihood score, etc. The approach is based on speech and video inputs with real-time data wherein 100 people were recruited for the collection of facial expression data. Chatterjee et al. [20] presented a methodology to utilize dictionary learning algorithms for activities of daily living monitoring. They used sparse learning based classification techniques in association with clustering approaches. Chen, Hung et al. [21] worked on activity detection of elderly citizens by performing the categorization at the sensor level. The categorization of the sensors is performed according to daily activities. Ye et al. [22] proposed a novel technique known as CLEAN wherein the sensor data is checked for abnormalities using statistical techniques. Skocir et al. [23] proposed an approach in which data from two different sensors are used to detect opening and closing. The node to which all other nodes are connected is battery powered. In this work two different approaches, one sliding window and other based on artificial neural networks for activity detection, were proposed. Lu et al. [24] proposed a technique to extract features from sensor data using the Beta Process Hidden Markov Model (BP-HMM). Latent features are extracted using BP-HMM. Further Support Vector Machine

is used for activity recognition in a supervised way. Subasi et al. [25] employed wearable sensors and mobile devices for proposing a mobile healthcare model for monitoring the health and well-being of a person. The work utilized different data mining techniques, Mobile technologies like 4G systems, GPS, and Bluetooth Technologies. The well-being of a person is monitored using various activities like standing, sitting, relaxing, laying down, walking, climbing stairs, jogging, and running, etc. Yao et al. [26] presented an end-to-end healthcare monitoring model named Web-Based Internet of Things Smart Home (WITS). This web-based system employed data and knowledge-driven techniques for monitoring activities in real time. Hassan et al. [27] proposed a technique for human activity recognition using wearable smartphone inertial sensors using mean, median, and mode techniques. Sendra et al. [28] proposed a smart architecture for extracting different types of activities. The main features are extracted based on day-to-day activities. The work employed Linear Discriminant Analysis, Kernel Principal Component analysis for further analysing the features. The authors designed a wearable device, which measures body temperature and heart rate. The designed device is capable of making a fusion of data from different sources in real time for detecting strange conditions to send warnings to concerned people. Pham et al. [29] presented a Cloud-based Smart Home Environment for healthcare (CoSHE) facilities. The model collects motion, audio signals by using invasive wearable sensors and providing contextual information. Tran et al. [30] proposed a model, which is based on activity monitoring in a smart home environment for multiple residents. Different methods for multiple resident activity detection are evaluated on the same data. Ghaywat et al. [31] proposed a real-time approach to healthcare monitoring in real time where the activities of daily life are predicted based on past data. The work employed time series analysis techniques to detect normal and abnormal activities. Alberdi et al. [32] proposed a method that detects symptoms of Alzheimer's disease (AD). The dataset is collected from the records of 29 adults for a time span of more than 2 years. Various regression models are used to predict the symptoms of absolute changes in behaviour. The algorithm Smote BOOST is used to reduce the class imbalance. The results presented can significantly contribute to the early detection of Alzheimer's disease using smart home technology. Aramendi et al. [33] focused on detection of a decline in healthcare of older people, one of the prevailing problems in today's ageing society. The data is collected for 29 adults for a time span of 2 years. In this work, 10 different behaviour features are extracted. The health functionality of person is accessed every six months using the scale of instrumental activities of daily living compensation. Mora et al. [34] proposed a monitoring architecture based on home sensors and cloud-based back-end services. In this case, deep learning techniques are used for detecting changes in sensors activities for extracting patterns from sensor data. Chen et al. [35] proposed deep learning based techniques to automatically detect high-level activities by using binary sensor data. The work involved the use of an autoencoder to extract high-level features in three public smart home datasets. Verma et al. [36] proposed a methodology to

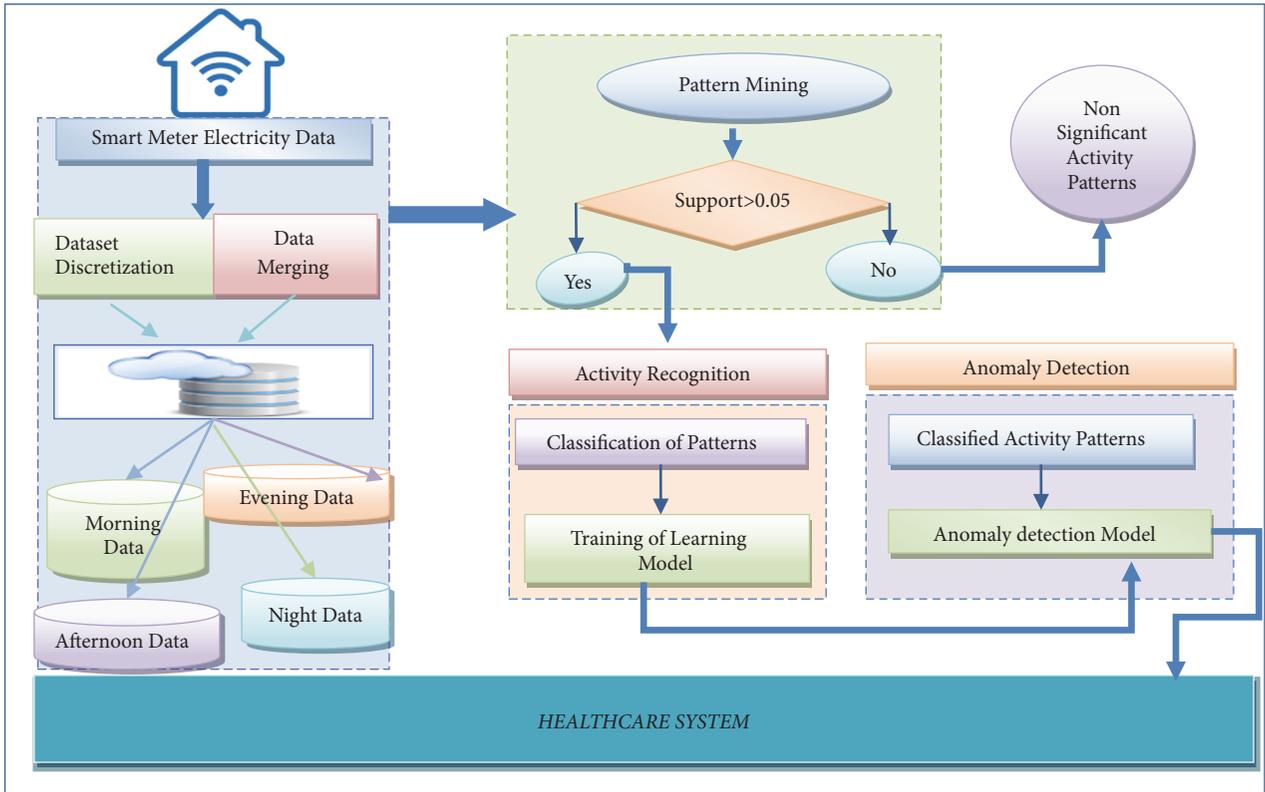


FIGURE 1: Conceptual diagram of anomaly detection for smart home inhabitants.

remotely monitor healthcare using fog computing. The model used the techniques like embedded data mining, notification service at the edge, distributed storage, etc. Iqbal et al. [37] proposed a web-based interpretable IOT platform for smart homes. The cloud-based platform is capable of handling the control of home appliances from any location. In this work, Raspberry Pi based platform is employed for interpretability and smart home appliances are migrated to the cloud using the Restful platform. Chellappan et al. [38] proposed a system that identifies and categorizes composite human activities at home by using wearable technology. It extracts a feature set from data using multimodal sensor suites. The accuracy of the model is increased using a two-level structured classification algorithm. Oliver et al. [39] proposed an approach for monitoring Activities of daily living (ADL) using Log Gaussian Cox Process. Wooi-Nee et al. [40] detected energy stealing and faulty meters by analysing consumer's energy utilization behaviour.

From the literature survey, it is concluded that the majority of work is done using sensor technology for activity detection, activity prediction, and behaviour analysis. Sensor-based data is used in most of the studies. The advanced metering based infrastructure is not utilized in either of these studies. The work, that finds abnormalities in the day-to-day routine, is mostly detecting outliers in data but anomalies detection techniques are not applied in this field to provide a general framework for an ambient assistant.

3. The Proposed Model

This section discusses various phases of the proposed model and particulars of respective methods besides related conceptual framework. The conceptual model of the proposed methodology is presented in Figure 1 with its different phases like preparation of data, pattern extraction, classification of patterns into manually annotated classes, and anomaly recognition.

In the initial phase, the raw data consisting of a collection of files storing the time series electricity usage pattern is processed and prepared for subsequent investigation. In the initial phase, data preparation and preprocessing techniques are applied to each data file. Data discretization and data merging techniques are used for preprocessing. The pre-processed files are combined to make an intermediate database. The intermediate database is further divided into 4 distinct databases meant for the morning, afternoon, evening, and night energy utilization records. In the following phase, frequent pattern mining, multiclass classification is carried out. The patterns point out repeated patterns of devices which frequently appear together in the dataset. If the sequence of devices (for example, bread maker and kettle) often appears together, then they are considered as frequent patterns. For making the system more efficient, progressive incremental data mining approaches are applied. If the support value of the extracted patterns is more than the threshold value

TABLE 1: Raw data.

Unix Timestamp	Energy Consumption
1388534500	589
1388534506	565
1388534512	566
1388534519	600
1388534524	565
1388534531	558

then these patterns are considered as normal. Similarly, all 4 distinct databases are mined to extract frequent patterns associated with time. The multiclass classification is performed to categorize the frequent patterns into manually labelled multiple categories. Different machine learning prediction models are trained on data to predict classes.

The final phase of the proposed work is anomaly recognition. The aim of this phase is to find out those abnormal patterns, which are deviating from normal activity patterns. In the current work, X-means clustering algorithm is used for anomaly detection. The model is based on X-means clustering which is an extension of the K-means clustering technique. It automatically makes the number of clusters, without prior knowledge of the number of clusters. In the following step, the Local Outlier Factor (LOF) detects outliers by calculating the deviation of given data from neighbors [41]. Anomaly Score is assigned to each transaction by the LOF method. This step is followed by the application of the prediction model for predicting the anomalies for the whole database. Thereafter filtration method is applied to filter out those transactions whose anomaly score is more than 1.6. The patterns having an anomaly score higher than 1.6 are considered as anomalies. They are filtered out and sent to the healthcare system for further actions.

The following section provides a brief outline of the methods, models, and evaluation parameters used in this proposed work.

3.1. Data Preparation. The data utilized in this work is time series energy consumption dataset from UK-DALE [8]. The data includes meter readings from smart meters installed at five different houses in the United Kingdom with a time gap of 6 seconds. The original dataset stores the readings from 53 different devices installed in a single house. The energy consumption records of each device are stored in separate files. Table 1 gives the format of raw data files. Various techniques of data preprocessing are implied to convert the raw data into the ideal format.

Conversion of UNIX Timestamp into Date and Time. The raw data is in the form of $\langle UNIX\ timestamp, Energy\ consumption \rangle$ as shown in Table 1. The UNIX timestamp is not easily understandable; therefore; it is converted into the date and time format.

Data Discretization. Data discretization is data preprocessing technique [42]. In data discretization, large numbers of values are converted in smaller ones for easiness in data

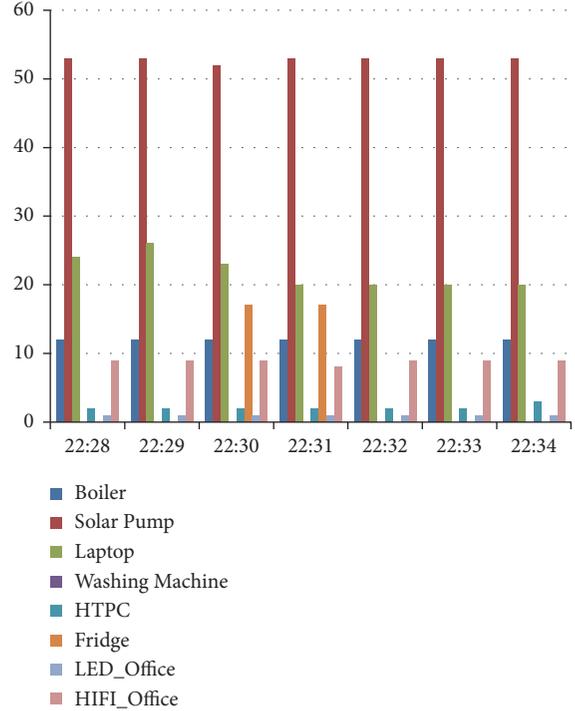


FIGURE 2: Association of energy consumption data with time in initial dataset.

evaluation and processing. In the current work, the equal width portioning technique is used for data discretization. As initial electricity consumption data is in redundant form, it is converted into a 1-minute time gap with the aid of this technique.

Data Merging. Data merging implies the integration of different data files into a single combined database. In this case, the many-to-one merge approach is used. This step is significant as single data files do not have any importance in the proposed model. After application of date and time conversion, data discretization and data merging an intermediate database are obtained as given in Table 2.

The table presents a unified view of the intermediate database. The table provides information on energy consumption by each device at a particular period of time.

Division of Data into Different Time Slots. To make the pattern discovery process more efficient and making the proposed model rapid at tracking abnormalities, the intermediate data file is further divided into four separate files based on different phases of the day, i.e., morning, afternoon, evening, and night. Table 3 gives a better preview of how the data is portioned according to different phases of the day.

Figure 2 is a visualization of energy consumption in association with time. Herein the x-axis represents the time of day when a device is operating (kWh) and the y-axis represents the amount of energy consumed by each device.

Conversion to Binary Values. The next step involves further processing of four (time slot based) data files to convert them

TABLE 2: Intermediate database.

Date	B	SP	LP	WM	DW	TV	KL
11/9/2012 22:28	12	53	24	0	1	1	0
11/9/2012 22:29	12	53	26	0	1	1	0
11/9/2012 22:30	12	52	23	0	1	1	0
11/9/2012 22:31	12	53	20	0	1	1	0
11/9/2012 22:32	12	53	20	0	1	1	0
11/9/2012 22:33	12	53	20	0	1	1	0
11/9/2012 22:34	12	53	20	0	1	1	0
11/9/2012 22:35	12	53	20	0	1	1	0
11/9/2012 22:36	12	53	20	0	1	1	0

*B: boiler, SP: solar pump, LP: laptop, WM: washing machine, DW: dish washer, and KL: kitchen lights.

TABLE 3: Data division based on different time slots.

Time of day	Database Name
5:01am-12:00pm	Morning
12:1-5:0pm	Afternoon
5:1pm-8:00pm	Evening
8:00pm-5:00am	Night

into the format that records the data telling the operating status of various appliances with respect to the time frame. Table 4 shows the final data file which is ready to mine, with 1 value signifying device being ON and 0 representing the nonoperational device in that time slot.

3.2. Extracting Frequent Patterns of Activities of Daily Living. One of the significant steps of the proposed work is the discovery of hidden patterns of human activities from smart meter electricity utilization data. Frequent patterns are sequences that are present in data more than the threshold value [43]. The sequences whose frequency is more than or equal to the minimum threshold value are coined as frequent patterns.

Our daily living activities include “*preparing food, studying, using a computer, watching TV, listening to radio*”, etc. Pattern discovery is more related to the discovery of an association between multiple devices operating together like “*listening to the radio while preparing food*”, “*washing clothes while watching television*”. The objective is to trace sequences of activities so that sudden change or imbalance in these activities can be monitored as earlier as possible and appropriate actions may be taken thereof. For frequent pattern mining, all the devices which are shown to be active in the preprocessed data file are considered to mine the activity patterns.

For pattern mining, different techniques are applied to find candidates; thereafter those candidates are utilized to generate frequent patterns. The frequent pattern (FP) growth is a popular algorithm for association rule mining that uses divide and conquers method to divide the large dataset into smaller one, decreasing the size of search space [44]. In this algorithm, databases are converted into smaller data

structures called FP trees that avoid the repeated scans of the database [45].

The algorithm consists of two subprocesses: Step 1 is FP-tree construction and Step 2 generation of repeated patterns based on FP-tree. The steps of construction of the FP-tree are as follows.

The Method

- (i) For every item, build a conditional base of patterns followed by construction of conditional FP-tree.
- (ii) Repeat the method on every recently formed conditional FP-tree till the resultant FP-tree collapses or it has just a single pathway left.

Phase 1 (construction of conditional pattern base)

- (a) Start from the header table of frequent items within the FP-tree.
- (b) Using the linkage of every recurrent item traverses the whole frequent pattern tree.
- (c) All of the prefix ways which are remodeled are joined to create a pattern base (conditional).

Phase 2 (construction of conditional FP-tree)

- (a) Begin from the tail of the header list
- (b) For every base of patterns
- (iii) The items which are frequent in the pattern base construct an FP-tree.

To find frequent item, all the frequent item sets containing item are mined by following all the node links of the item, starting from head of the item in FP-tree. The algorithm starts with least frequent item present in header table. It finds all the paths to that item and increases the counter according to support count. In the subsequent step, the algorithm finds conditional pattern base of each item. Conditional pattern base of specific item consists of all the patterns leading to that item. The conditional frequent pattern tree is constructed from conditional pattern base. The algorithm for frequent item set mining is again called for new conditional FP-tree. This process carries on until tree is empty or only one path is left. If tree contains only single path, the algorithm finds

TABLE 4: intermediate database.

Date	B	SP	LP	WM	DW	TV	KL
1/1/2014 0:10	1	0	0	0	1	1	1
1/1/2014 0:11	1	0	0	0	1	1	0
1/1/2014 0:12	1	0	0	0	1	1	1
1/1/2014 0:13	1	0	0	0	1	1	0
1/1/2014 0:14	1	0	0	0	1	1	1
1/1/2014 0:15	1	0	0	0	1	1	0
1/1/2014 0:16	1	0	0	0	1	1	1
1/1/2014 0:17	1	0	0	0	1	1	1
1/1/2014 0:18	1	0	0	0	1	1	1

*B: boiler, SP: solar pump, LP: laptop, WM: washing machine, DW: dish washer, and KL: kitchen lights.

all the possible combinations of items that support minimum support values.

In the current work, pattern mining is done in different time slots on distinct databases. The patterns are stored in different databases to accurately detect the patterns of activities performed in morning, afternoon, evening, and night.

3.3. Multiclass Classification for Activity Representation. In this technique, frequent patterns are classified into manually labelled categories with the aid of machine learning prediction models. As delineated in previous sections, information is distributed into different parts in line with time frames (morning, afternoon, evening, and night). Each database is mined individually; therefore the patterns in every time slot are distinct.

Multiclass classifications algorithms are applied to frequent patterns so as to map the device operative status with manually annotated classes. An example of manual classes is given as if kitchen lights, bread maker, and coffee maker are ON; it signifies that person is preparing food. Similarly, different patterns are manually labelled for classification. Different classes are allotted to patterns on the basis of normal activities of daily living (ADL) like bathing, showering, cleaning, maintenance of the house, preparing meals, etc.

Mostly binary category classification algorithms are tuned to perform the multiclass classification techniques. Existing strategies to handle multiclass classification problems are Transformation to binary, extension from binary, and hierarchical classification [46].

In the current work, the binary classification techniques are extended to resolve the multiclass classification problems. These techniques are also known as algorithm adaptation techniques. The main classifiers used in the work are as follow.

Support Vector Machine. It is a computational algorithm that builds a hyperplane or a group of hyperplanes in a high or infinite dimensional space [47]. It is based on the principle of increasing the least distance from the hyperplane, which separates them to the nearest example set. Basic SVM only deals with binary classification problems, but they are extended to handle the multiclass classification problems. In this, additional constraints and parameters are added to deal with multiple categories in multiclass datasets.

Naïve Bayes. Naïve Bayes is based on the Maximum A Posteriori (MAP) technique [48]. It uses the concept of conditional probability. It is naturally extensible to the multiclass problems. The classifier performs very well in multiclass conditions in spite of its simplifying approach of solving the problem using conditional independence.

Reptree. It belongs to decision tree category [49]. It uses regression tree logic and creates multiple trees in incremental iterations. It is a powerful classification technique. The training data split infers a good general decision based on available features. It is automatically compatible to handle binary class as well as multiclass classification problems.

MLP. Multilayer Perceptron is a class of artificial neural network [50]. It consists of a minimum of three layers of nodes where each node is a neuron in MLP except input nodes. It uses backpropagation technique of supervised learning. Multiple layers and nonlinear activation differentiate MLP from other linear preceptors. It can distinguish the data, which is difficult to separate linearly.

J48. It is a decision tree, which is an implementation of the Iterative Dichotomiser 3 (ID3) algorithm [51]. For classifying new data it needs to make the decision tree of training data. It keeps on comparing the testing data. When features are matched with any tree more than other, then instances are classified to that class.

KNN. K nearest neighbor approach is based on similarity calculation between instances. It uses local average calculation for classification of data into different classes.

3.4. Anomaly Recognition Model. The next step of the proposed model is to discover those activities, which do not seem to be expected or are missing out in line with the time slot. Any uncommon pattern in data that does not tally to normal behaviour is considered as an anomaly. When frequent patterns are mapped onto different activities of everyday life, then there may be some patterns that do not match with day-to-day activities patterns. Those activities need to be detected and evaluated to find whether those are anomalies or the training model is incapable of recognizing them. The task is performed using an anomaly recognition

- (1) Set the maximum number of clusters to be k_{\max} .
 - (2) Repeat the steps (3) to (6) for $k_0 = 2$ to k_{\max}
 - (3) For $K=k_0$ apply K-means clustering
 - (4) Label the divided clusters as $C_1, C_2, C_3 \dots C_{k_0}$
 - (5) Repeat the steps (4.a) to (4.c) for $i \in [1, k_0]$
 - (5.a) For every cluster C_i , generate two new centroids from original centroid (by transforming the initial value of centroid in two different directions along a randomly chosen value of vector by an amount equals to their cluster size.)
 - (5.b) Apply K means with $K=2$. Label the divided clusters $C_i^{(1)}, C_i^{(2)}$
 - (5.c) Apply the model selection test BIC to check if two clusters better than original single cluster in each case. Replace each centroid based on the criteria of model selection.

$$\text{BIC}(k) = 2 * \log p + k \log q$$
 Where p - no. of observations, k - clusters and $\log q$ - log probability.
 - (5.d) Let BIC_C represents -score of children clusters and BIC_P -score of parent cluster
 - (5.e) If $\text{BIC}_C > \text{BIC}_P$, the two-divided clusters are accepted, and the division is continued;
 - (5.f) If $\text{BIC}_C < \text{BIC}_P$ the two-divided clusters are discarded, and the original cluster is kept;
- (6) If the condition of convergence is not fulfilled, go to Step (2). Otherwise, Stop.

ALGORITHM 1: Algorithm for X-means clustering.

technique. The technique discovers odd patterns that do not match with other items and patterns in the dataset. The model takes its input from the activity classification model. First of all, it performs X-means clustering on categorical data to make different clusters of multiple activity classes. The X-means clustering is the improved version of K-means clustering. In K-means clustering number of clusters are needed to be specified while in X-means clustering there is no need of specifying the number of clusters. It automatically chooses the number of clusters. As a result, the performance of the anomaly detection algorithm improves. The next step discovers the outlier score using different algorithms. An interval, which has higher outlier score associated with it, is significant and requires investigation. Next, the prediction model is applied to filter out instances as normal or abnormal activity patterns. The abnormal patterns may be passed on to the healthcare system to evaluate those patterns for taking relevant actions according to the problem.

In the following sections, the theoretical background of all the subprocesses used in anomaly detection model is discussed.

X-Means Clustering. It is performed on multiclass classification dataset. The algorithm determines the exact number of centroids on the basis of a heuristic. It starts with a minimum number of centroids and iteratively increases the centroids according to data.

X-means clustering algorithm is an extension to the K-means that automatically determines the number of clusters

based on BIC scores. Starting with only one cluster, the X-means algorithm goes into action after each run of K-means, making local decisions about which subset of the current centroids should split themselves in order to better fit the data. Given a range for k , $[k_{\min}, k_{\max}]$, the X-means algorithm starts with $k = k_{\min}$ and continues to add centroids when they are needed until the upper bound is reached. New centroids are added by splitting some centroids. During the process, the centroid set with the best score is recorded as the final output (see Algorithm 1).

The splitting decision is done by computing the Bayesian Information Criterion (BIC) [52]. According to the outcome of the BIC test, either the original centroid or new two centroids are discarded. If the BIC score of the produced children clusters (new centroids) is less than the BIC score of their parent cluster (original centroid), the split is not accepted and parent cluster is kept in the pool. Otherwise, the split is accepted and the algorithm proceeds similarly to lower levels.

The output of the algorithm is the set of centroids along with the possible value of K within the given range which scores best by BIC model selection criteria.

Local Outlier Factor (LOF). It is technique for outlier scoring. LOF is based on the concept of local density. The points are termed local to each other based on density. Local Outlier Factor is a measure that calculates the density of a particular point compared to the density of its neighbors.

```

Input: Database D
Output: (normalized) anomaly score for all and one  $d \in D$ 
{Phase 1: Model Construction}
  for all  $d \in D$  do
    select condition (d)
    Model(d):=build model(d, condition(d))
  end for
{Phase 2: Model Comparison}
  for all  $d \in D$  do
    choose reference (d)
    Score(d):= compare (model(d), {model(n)}  $n \in$  reference d)
  end for
{Phase 3:Normalization}
  if Normalization needed then
    for all  $d \in D$  do
      Normalized Score(X):=normalize (score(d))
    end for
  end if

```

ALGORITHM 2: Algorithm for local outlier factor.

When an object is compared to local densities of neighbors, the points, whose density is different from neighbors, are termed as anomalies. Main components of local outlier detection algorithmic scheme are as follows:

- (i) Condition: it is a local condition of some object (x) for building model (condition(x))
- (ii) Model: model stands for the technique employed for construction of underlying model
- (iii) Reference: it indicates the reference condition of object for comparison of model (reference (x))
- (iv) Comparison: comparison is a method used to compare various models
- (v) Normalization: it signifies the global procedure of normalization
- (vi) Build model: it implies a local condition of some object (x) for defining the criteria to calculate its density as compared to its neighbors
- (vii) Compare model: it compares the densities of a particular object (x) with respect to its neighbor object (n)

See Algorithm 2.

Here input database “D” is resultant dataset formed in X-means clustering, while “d” denotes individual data points in resultant clusters.

Prediction Model. It is a technique that employs probability and data mining to predict the ending result. In this, every model is formed from a variety of variables that are likely to affect future outcomes. In the context of anomaly detection, the model is first trained on the example set using different learning algorithms. Thereafter the learnt classifier is applied on testing data, to predict the final result.

Filtration. This technique is employed to select favourable values and discard the unfavourable results. In the context of

this model, those examples that match the given condition, i.e., anomaly score greater than the threshold value, are filtered and coined as abnormal patterns using this method.

4. Experimental Analysis

The proposed technique is implemented on Standalone Machine using Python 3.7 version under Intel Core®i7 processor @ 3.2 GHz, 8.00 GB RAM, running on a 64-bit operating system. The model analysis and experiments are implemented on energy time series dataset recorded from a real house in the United Kingdom [8]. This energy utilization time series dataset includes energy consumption records of fifty-three domestic devices with a time gap of six seconds.

The simulation results of different phases of the experiment are summed up in the following sections. To carry out the implementation of the proposed model as discussed in Section 3, the data is converted into the format, which is suitable for proposed techniques. In the initial phase of data preparation, data is converted from UNIX timestamp into 1-minute gap dataset. Different techniques of data preprocessing are applied to convert the raw data into the ideal format. Data discretization and data merging techniques are enforced on the data files for creating a single data source of 53 devices. Further, the data is divided into 4 different time slots of the morning, afternoon, evening, and night. The data is further processed to convert it into a format which records the data in a table according to the operating status of various appliances.

The processed dataset is utilized to extract frequent patterns from the dataset. The resulting patterns are of different sizes and show different combinations. Minimum support value for FP-growth is 0.5. All the patterns having support more than 0.5 are considered as frequent patterns and are stored in the frequent itemset database.

Table 5 depicts some of sample frequent patterns. These patterns provide an idea of conjoint operation of various devices. After mining, frequent patterns of data is categorized

TABLE 5: Frequent activity patterns.

Support	D1	D2	D3	D4	D5	D6	D7
0.532	GO	Wi-Fi	KL	RD	KS	T	LRL
0.532	R	GO	Wi-Fi	PC	SRL	KS	
0.857	R	PC	TV	LRL	LRW		
0.997	R	PC	SRL				
0.501	R	DW	Iron	WM	IC	TV	
0.693	KS	KL					
0.537	T	LRW					
0.532	GO	Wi-Fi					

*GO: gas oven, R: router, KL: kitchen lights, DW: dishwasher, LRW: living room woofers, KR: kitchen radio, T: toaster, SRL: study room led, LRL: living room lights, RD: radio, WM: washing machine, and IC: iPod charger.

into different manually annotated classes. Activities are classified into multiple classes based on Instrumental Activities of Daily Living and Activities of daily living. The classes assigned to different patterns resemble daily life activities like if a person is working in a kitchen then activity performed is “preparing food”. If Wi-Fi, PC, and study room lights are on then the person is “studying”. If iron is on and in parallel TV is also on then activity performed is “pressing cloths while watching TV”. Here we are considering only those patterns whose support value is greater than 0.5. In this way classes are assigned to frequent patterns.

For multiclass classification, five different classifiers are used which classify data into different classes. Classifiers employed are SVM, MLP, J48, RepTree, and Naïve Bayes. The results of the classification models are represented in Table 7. The table gives detailed accuracy statistics for each time slot (morning, afternoon, evening, and night). The model performance is evaluated based on various evaluation criteria like FP rate, recall, precision, F measure, PRC, and ROC that points out the error rate in the accuracy results. In morning data, all devices, which are active in the morning from 5 a.m. to 12 p.m., are considered. The following text gives a brief description of the evaluation parameters.

Accuracy. It evaluates the correctness of classifier. Accuracy can be given as in (1)

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{TotalData}} * 100 \quad (1)$$

Precision. Precision tells how accurate the model is in finding the predictive positive values, with respect to the count of values actually positive. It is calculated as in (2)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

Recall. It tells how many accurate positives the model predicts by telling them positive. The formula of the same is given in (3)

$$\text{Recall} = \frac{\text{True positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

Kappa Statistics. It computes the conformity between two rates that classify all n items into m mutually exclusive groups. The value of kappa is defined as in (4).

$$\kappa\text{appa} = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

where P_o is the relative observed agreement among rates and P_e is the imaginary probability of chance agreement.

FP Rate. The probability of rejecting null hypothesis falsely for a particular test is termed as false positive rate. FP rate is calculated by taking the ratio of the number of positive events termed wrongly as negative, to the actual number of total negative events. The value of the FP rate is defined as in (5)

$$\begin{aligned} \text{False Positive rate} \\ = \frac{\text{Number of False positives}}{\text{True negatives} + \text{Total number of negatives}} \end{aligned} \quad (5)$$

F Measure. It represents the harmonic mean of precision and recall. The value of F measure lies between zero and one. The higher value of F measure indicates high classification performance [53]. The value of FP measure is given as in (6)

$$\begin{aligned} \text{F measure} \\ = \frac{2 * \text{True Positive}}{2 * \text{True Positive} + \text{False Positive} + \text{False Negative}} \end{aligned} \quad (6)$$

Precision-Recall Curve (PRC). Precision-Recall curve shows the relationship between recall and precision. PRC curve is based on the same concept as the ROC curve. In PR curve x-axis represents precision and y-axis represents recall. It is widely used to evaluate the classification performance.

Receiver Operating Characteristics (ROC). ROC curve is a two-dimensional curve in which x-axis represents false positive rate while the y-axis represents True Positive Rate. It is used to make the balance between True Positives and False Positives [53].

Table 6 depicts that the best kappa value, i.e., 0.9723, is given by the MLP model for the morning data. Furthermore,

TABLE 6: Detailed accuracy statistics.

	Model	Accuracy	FP Rate	Precision	Recall	F-Measure	Kappa statistic	PRC	ROC
Morning Results (5am-12pm)	SVM	92.33%	0.003	0.984	0.983	0.981	0.9082	0.977	0.990
	MLP	97.67%	0.003	0.979	0.977	0.977	0.9723	0.986	0.998
	KNN	92.81%	0.011	0.930	0.928	0.927	0.9144	0.967	0.995
	J48	94.96%	0.008	0.950	0.950	0.949	0.9399	0.959	0.992
	RepTree	91.72%	0.014	0.918	0.917	0.915	0.9010	0.996	0.970
Afternoon Results (12pm-5pm)	Naïve Bayes	79.91%	0.029	0.811	0.799	0.791	0.7610	0.981	0.888
	SVM	95.32%	0.007	0.951	0.954	0.952	0.9438	0.945	0.987
	MLP	99.92%	0.001	0.999	0.999	0.999	0.9991	1.000	1.000
	KNN	88.59%	0.006	0.965	0.965	0.965	0.8620	0.989	0.998
	J48	93.99%	0.011	0.941	0.940	0.946	0.9278	0.994	0.968
Evening Results (5pm-8pm)	RepTree	97.42%	0.02	0.974	0.974	0.973	0.9690	0.999	0.987
	Naïve Bayes	84.30%	0.015	0.880	0.843	0.850	0.8152	0.934	0.986
	SVM	95.35%	0.003	0.954	0.953	0.949	0.9440	0.949	0.992
	MLP	98.44%	0.000	0.998	0.999	0.999	0.9810	1.000	1.000
	KNN	96.51%	0.006	0.966	0.965	0.965	0.9580	0.988	0.998
Night Results (8pm-5am)	J48	80.62%	0.040	0.809	0.806	0.805	0.7660	0.806	0.994
	RepTree	82.62%	0.260	0.886	0.826	0.829	0.7920	0.973	0.862
	Naïve Bayes	93.30%	0.007	0.912	0.993	0.921	0.9190	0.999	0.978
	SVM	98.86%	0.001	0.990	0.989	0.989	0.9864	0.998	1.000
	MLP	99.97%	0.000	1.000	0.999	1.000	0.9997	1.000	1.000
Night Results (8pm-5am)	KNN	96.55%	0.006	0.966	0.966	0.960	0.9580	0.987	0.998
	J48	96.54%	0.002	0.959	0.969	0.905	0.9510	0.986	0.956
	RepTree	90.39%	0.013	0.912	0.904	0.905	0.8850	0.986	0.980
	Naïve Bayes	94.89%	0.006	0.935	0.949	0.941	0.9390	0.986	0.980
								0.999	0.988

TABLE 7: Anomaly score table for morning data.

ID	Cluster Id	Anomaly Score
581	cluster_0	1.73
582	cluster_0	1.51
587	cluster_0	1.75
589	cluster_0	1.52
601	cluster_1	1.50
602	cluster_1	1.56
604	cluster_1	1.52
605	cluster_1	1.59
606	cluster_1	1.56
610	cluster_2	1.50
611	cluster_2	1.53
613	cluster_2	1.52
614	cluster_2	1.59
615	cluster_2	1.56

the same model is able to be exceedingly precise in filtering out the classes from the dataset with a precision of 0.979. It also gives the best accuracy value of 97.67% followed by KNN and SVM classifier. The Naïve Bayes classifier has the lowest recall and precision value of 0.811, 0.799, respectively, in comparison to other classifiers.

For afternoon data (i.e., between 12 p.m. to 5 p.m.), the RepTree performed well with 0.9690 kappa value but not better than the MLP model. The MLP model provides the best kappa value of 0.9724, whereas SVM and J48 give 0.9438 and 0.9278 kappa values, respectively. In addition, the precision and recall value of the MLP model is best, in this case, i.e., 0.999. The MLP provided the best results in terms of accuracy, i.e., 99.92% followed by RepTree with 97.42% and SVM with 95.32%. The records of all the active devices between 5 p.m. to 8 p.m. are associated with evening data. Most of the activities in this time slot are related to preparing food and relaxing. In this case, MLP, KNN, and SVM are giving the best performance with a kappa value of 0.9810, 0.9580, and 0.9440, respectively. The model J48 gives the least value of kappa, i.e., 0.766. The MLP model gives the best values of recall and F measure. The table presents comparable values of the accuracy of more than 95% for MLP, KNN, and SVM classifiers.

The night time data files consider frequent patterns of devices, which are active from 8 p.m. in the night to 5 a.m. in the morning. This is the largest time slot but activities in this time slot are limited as most of the devices are OFF because occupants of the house are resting. The table depicts that both MLP and SVM models give the best kappa values of 0.9997 and 0.9864, respectively. In this case, RepTree turns out to be the worst performer with a kappa value of 0.8850. The SVM and MLP classifiers are the winners in terms of precision and recall values. Two models SVM and MLP performed best for the night time data for recognition of different types of activities of the inhabitants of the house.

From Table 6, it is concluded that the MLP classifier is the overall best performer in terms of kappa parameter with value greater than 0.95 for all the four-time slots whereas the SVM

TABLE 8: Anomaly score table for afternoon data.

ID	Cluster Id	Anomaly Score
47	cluster_0	1.86
208	cluster_3	1.57
213	cluster_0	1.52
1412	cluster_2	1.50
2357	cluster_2	1.57
2365	cluster_2	1.53
3211	cluster_2	1.59
3220	cluster_2	1.56
3761	cluster_2	1.70
3771	cluster_2	1.52
4009	cluster_2	1.56
4020	cluster_2	1.51

model turns out to be the second best model in terms of kappa statistics.

The most significant and final step of the proposed model is anomaly recognition. As already discussed in Section 3.4, anomalies or outliers are patterns, which deviate from normal behaviour. In the context of this paper, if imbalanced appliance operating status is detected then it means they are anomalies. This section shows that the anomalies in Tables 8–10 are the results of anomaly recognition where each table has an original id, cluster id, and anomaly score. Herein the table, only those patterns are presented which are have anomaly score more than the threshold value, that clearly signifies that patterns are abnormal. The Id column depicts the original id of classification pattern, which is anomalous to the normal behaviour. The tables are adjoined by graphs, which give a better overview of associated anomalies according to anomaly score.

Table 7 depicts the anomalies recognized from morning time data. This table shows those entries in morning database which are deviating from normal behaviour as they have anomaly score more than the threshold value (>1.6). The data in Figure 3(a) shows a large number of anomalous entries but the anomaly scores associated with them are not showing many extreme values.

Table 8 depicts the anomaly score table for afternoon database. From results, it can be seen that anomaly score for most of the patterns is not much extreme only one or two values have shown extreme behaviour of deviation from normal patterns.

Most of the anomalies belong to cluster number three. Figure 3(b) is the anomaly graph associated with afternoon time. In this case, only one pattern has an anomaly score of 1.86 (higher than threshold value of 1.6). It signifies that the behaviour of resident is not extreme in this timeslot, as very few patterns deviate from normal activity patterns.

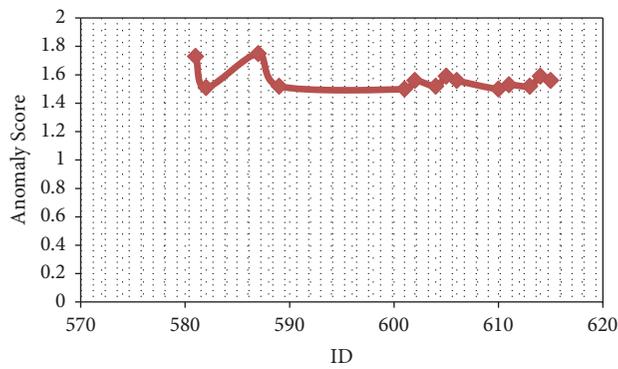
Table 9 shows the anomaly values for evening data, which is recorded between 5 pm. in the evening to 8 p.m. Figure 3(c) depicts that the number of anomalies in evening time is less as compared to morning time (as shown in Figure 3(a)), but anomaly score associated with evening entries expresses

TABLE 9: Anomaly score table for evening data.

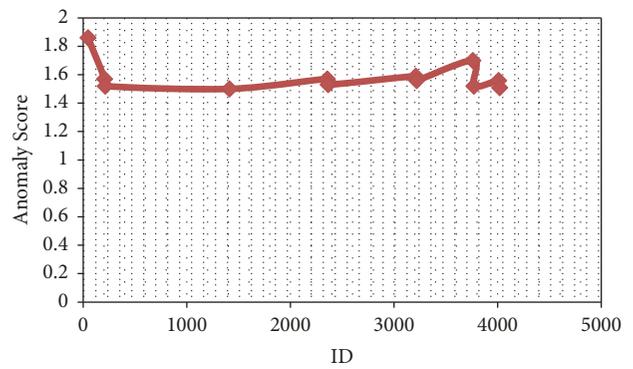
ID	Cluster Id	Anomaly Score
48	cluster_0	1.73
49	cluster_0	2.18
4043	cluster_3	1.73
4319	cluster_3	1.72

TABLE 10: Anomaly score table for night data.

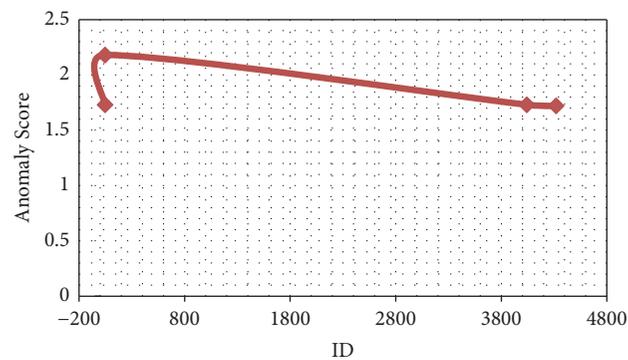
ID	Cluster Id	Anomaly Score
1	cluster_1	1.61
300	cluster_2	1.52
803	cluster_2	1.51
1624	cluster_2	1.60
2610	cluster_2	1.59
3487	cluster_2	1.63
4058	cluster_2	1.65
4321	cluster_2	1.61
4401	cluster_2	1.62



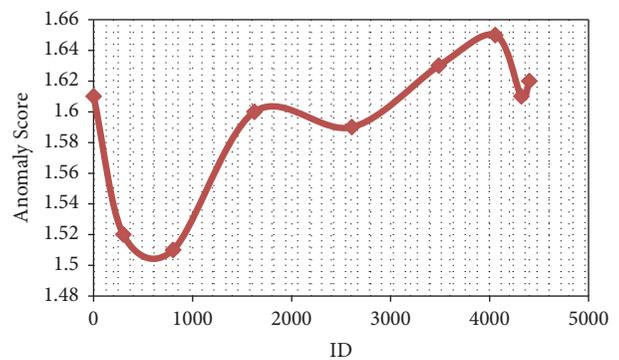
(a) Anomaly graph for morning



(b) Anomaly graph for afternoon



(c) Anomaly graph for evening



(d) Anomaly graph for night

FIGURE 3

extreme higher scores as compared to morning and afternoon data. Higher anomaly score signifies the extreme behaviour of residents of the house in a particular time slot. It indicates that the patterns have much difference from normal pattern and it needs the attention of concerned healthcare attendants as soon as possible

Table 10 and Figure 3(d) show the results for the night time slot. The night time data file is recorded from 8 p.m. in the night to 5 a.m. in the morning. It is the largest time slot but most of the devices are inactive in this time slot as the occupants sleep at this time. It is clear from Table 10 that the anomaly scores in this time slot are not extreme, which indicates a slight difference in activity patterns from normal patterns. From anomaly values, the probability of anomalous activities is less.

From the simulation results related to behaviour abnormality detection in human behaviour using anomaly detection techniques for different time slots, it is clear that the applied technique is successful in detecting considerable abnormal patterns which have different abnormally score values as compared to threshold values. The generalized conclusion of results is not possible as different time slots have shown different trends related to abnormalities. Morning time data has shown a large number of anomalous entries but the anomaly scores associated with them are not showing many extreme values. Similarly, for afternoon data, the number of entries is high but anomaly scores associated are not much extreme, which directly indicates that the activities at the particular house are slightly different from normal day to day, records. It is quite possible, as humans tend to change their work routine slightly according to their comfort and other reasons. Anomaly scores associated with evening time data are high as compared to the threshold value of 1.6. Therefore, it requires the immediate attention of concerned authorities as a high score is an indicator of the high difference in activities patterns from normal patterns. For night time data anomaly score value is not much extreme as night time data is related to inactivity at home.

5. Conclusion and Future Work

Most of the times a person perform similar habitual activities every day. However, sometimes the activities are not performed in a normal sequence by the person. The current work focuses on the development of anomaly detection model by extracting and classifying the regular activity patterns of the residents, thereby further utilizing those patterns for predicting the anomalous behaviour of a resident using LOF method. This anomaly recognition model is able to track individual's activities from low-resolution smart meter energy consumption data on time series data and that confronted anomalous pattern can be sent to the healthcare system for further action. The cost of implementation of this model is also negligible as compared to deploying sensors in each house. In the future, this system can be used for old age citizens or patients who are suffering from serious health issues for keeping an eye on their health status. The model can be improved to work on real-time data. Other smart meter

data's like gas consumption data can be collaborated with electricity data to predict the health status of an occupant in a much accurate way.

Data Availability

The UK-DALE data supporting this meta-analysis are from previously reported studies and datasets, which have been cited. The processed data are available at https://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl?GoButton=Detail&WantComp=41.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2014-1-00720) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References

- [1] I. Beretta, "The social effects of eco-innovations in Italian smart cities," *Cities*, vol. 72, pp. 115–121, 2018.
- [2] B. Kim, "A distributed coexistence mitigation scheme for IoT-based smart medical systems," *Journal of Information Processing Systems*, vol. 13, no. 6, pp. 1602–1612, 2017.
- [3] P. K. Sharma, S. Y. Moon, and J. H. Park, "Block-VN: a distributed blockchain based vehicular network architecture in smart city," *Journal of Information Processing Systems*, vol. 13, no. 1, pp. 184–195, 2017.
- [4] H. Jo and Y. I. Yoon, "Intelligent smart home energy efficiency model using artificial TensorFlow engine," *Human-Centric Computing and Information Sciences*, vol. 8, no. 1, pp. 1–18, 2018.
- [5] P. K. Sharma, J. H. Ryu, K. Y. J. H. Park, and J. H. Park, "Li-Fi based on security cloud framework for future IT environment," *Human-Centric Computing and Information Sciences*, vol. 8, no. 1, pp. 23–36, 2018.
- [6] Z. Ibrahim, F. A. Rahim, and S. Mokhtar, "Digital forensics issues in advanced metering infrastructure," *Journal of Fundamental and Applied Sciences*, vol. 10, no. 6S, pp. 2714–2726, 2018.
- [7] P. Siano, "Demand response and smart grids—a survey," *Renewable & Sustainable Energy Reviews*, vol. 30, no. 5, pp. 461–478, 2014.
- [8] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, no. 1, pp. 150007–150021, 2015.
- [9] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "TW-See: human activity recognition through the wall with commodity Wi-Fi devices," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 306–319, 2019.
- [10] F. Passerini and A. M. Tonello, "Smart grid monitoring using power line modems: Anomaly detection and localization," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 27302–27312, 2019.

- [11] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, no. 1, pp. 3–11, 2019.
- [12] L. P. Malasinghe, N. Ramzan, and K. Dahal, "Remote patient monitoring: a comprehensive study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 57–76, 2019.
- [13] Y.-H. Li, N.-N. Yeh, S.-J. Chen, and Y.-C. Chung, "Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network," *Mobile Information Systems*, vol. 2019, Article ID 6142839, 14 pages, 2019.
- [14] A. Miguez, C. Soares, J. M. Torres, P. Sobral, and R. S. Moreira, "Improving Ambient Assisted Living Through Artificial Intelligence," in *Proceedings of the World Conference on Information Systems and Technologies*, vol. 931, pp. 110–123, 2019.
- [15] F. Gomez-Donoso, F. Escalona, F. M. Rivas, J. M. Cañas, and M. Cazorla, "Enhancing the ambient assisted living capabilities with a mobile robot," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 9412384, p. 15, 2019.
- [16] G. Zhang, G. G. Wang, H. Farhangi, and A. Palizban, "Data mining of smart meters for load category based disaggregation of residential power consumption," *Sustainable Energy, Grids and Networks*, vol. 10, no. 1, pp. 92–103, 2017.
- [17] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-Garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: review and open research directions," *Information Fusion*, vol. 46, no. 1, pp. 147–170, 2019.
- [18] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, "Smart health monitoring using the advance metering infrastructure," in *Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, vol. 15, pp. 2297–2302, 2015.
- [19] M. S. Hossain, "Patient state recognition system for healthcare using speech and facial expressions," *Journal of Medical Systems*, vol. 40, no. 12, pp. 272–280, 2016.
- [20] P. De, A. Chatterjee, and A. Rakshit, "Recognition of human behavior for assisted living using dictionary learning approach," *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2434–2441, 2018.
- [21] H. Ghayvat, S. Mukhopadhyay, B. Shenjie, A. Chouhan, and W. Chen, "Smart home based ambient assisted living: Recognition of anomaly in the activity of daily living for an elderly living alone," in *Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2018*, pp. 1–5, MTC, May 2018.
- [22] J. Ye, G. Stevenson, and S. Dobson, "Detecting abnormal events on binary sensors in smart home environments," *Pervasive and Mobile Computing*, vol. 33, pp. 32–49, 2016.
- [23] P. Skocir, P. Krivic, M. Tomelj, M. Kusek, and G. Jezic, "Activity detection in smart home environment," *Procedia Computer Science*, vol. 96, no. 1, pp. 672–681, 2016.
- [24] L. Lu, C. Qing-ling, and Z. Yi-Ju, "Activity recognition in smart homes," *Multimedia Tools and Applications*, vol. 76, no. 22, pp. 24203–24220, 2017.
- [25] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," in *Proceedings of the 15th Learning and Technology Conference, L and T 2018*, pp. 29–34, Saudi Arabia, February 2018.
- [26] L. Yao, Q. Z. Sheng, B. Benatallah et al., "WITS: an IoT-endowed computational framework for activity recognition in personalized smart homes," *Computing*, vol. 100, no. 4, pp. 369–385, 2018.
- [27] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, no. 1, pp. 307–313, 2018.
- [28] S. Sendra, L. Parra, J. Lloret, and J. Tomás, "Smart system for children's chronic illness monitoring," *Information Fusion*, vol. 40, no. 1, pp. 76–86, 2018.
- [29] M. Pham, Y. Mengistu, H. Do, and W. Sheng, "Delivering home healthcare through a Cloud-based Smart Home Environment (CoSHE)," *Future Generation Computer Systems*, vol. 81, no. 1, pp. 129–140, 2018.
- [30] S. N. Tran, Q. Zhang, and M. Karunanithi, "On multi-resident activity recognition in ambient smart-homes," <https://arxiv.org/abs/1806.06611v1>.
- [31] H. Ghayvat, S. Mukhopadhyay, B. Shenjie, A. Chouhan, and W. Chen, "Smart home based ambient assisted living: Recognition of anomaly in the activity of daily living for an elderly living alone," in *Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2018*, pp. 1–5, USA, May 2018.
- [32] A. Alberdi, A. Weakley, M. Schmitter-Edgecombe et al., "Smart home-based prediction of multidomain symptoms related to Alzheimer's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1720–1731, 2018.
- [33] A. A. Aramendi, A. Weakley, A. Aztiria Goenaga, M. Schmitter-Edgecombe, and D. J. Cook, "Automatic assessment of functional health decline in older adults based on smart home data," *Journal of Biomedical Informatics*, vol. 81, pp. 119–130, 2018.
- [34] N. Mora, G. Matrella, and P. Ciampolini, "Cloud-based behavioral monitoring in smart homes," *Sensors*, vol. 18, no. 6, pp. 1951–1967, 2018.
- [35] G. Chen, A. Wang, S. Zhao, L. Liu, and C.-Y. Chang, "Latent feature learning for activity recognition using simple sensors in smart homes," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 15201–15219, 2018.
- [36] P. Verma and S. K. Sood, "Fog assisted-IoT enabled patient health monitoring in smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1789–1796, 2018.
- [37] A. Iqbal, F. Ullah, and K. S. Hafeez Anwar, "Interoperable Internet-of-Things Platform for Smart," *Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [38] P. Bharti, D. De, S. Chellappan, and S. K. Das, "HuMAN: complex activity recognition with multi-modal multi-positional body sensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 857–870, 2019.
- [39] J. Alcalá, O. Parson, and A. Rogers, "Detecting anomalies in Activities of Daily Living of elderly residents via energy disaggregation and cox processes," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built, BuildSys 2015*, vol. 2, pp. 225–234, November 2015.
- [40] S.-C. Yip, W.-N. Tan, C. Tan, M.-T. Gan, and K. Wong, "An anomaly detection framework for identifying energy theft and defective meters in smart grids," *International Journal of Electrical Power & Energy Systems*, vol. 101, no. 1, pp. 189–203, 2018.
- [41] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *ACM Sigmod Record*, vol. 29, no. 2, pp. 93–104, 2000.
- [42] H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: an enabling technique," *Data Mining and Knowledge Discovery*, vol. 6, no. 4, pp. 393–423, 2002.

- [43] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 55–86, 2007.
- [44] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," *SIGMOD Record*, vol. 29, no. 2, pp. 1–12, 2000.
- [45] C. Borgelt, "An Implementation of the FP-growth Algorithm," in *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*, vol. 5, pp. 1–5, 2005.
- [46] H.-Y. Lin, "Efficient classifiers for multi-class classification problems," *Decision Support Systems*, vol. 53, no. 3, pp. 473–481, 2012.
- [47] S. Winters-Hilt, A. Yelundur, C. McChesney, and M. Landry, "Support vector machine implementations for classification & clustering," *BMC Bioinformatics*, vol. 7, no. 2, pp. 1–18, 2006.
- [48] S. Wang, L. Jiang, and C. Li, "Adapting naive Bayes tree for text classification," *Knowledge and Information Systems*, vol. 44, no. 1, pp. 77–89, 2015.
- [49] S. Kalmegh, "Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news," *International Journal of Innovative Science, Engineering & Technology*, vol. 2, no. 2, pp. 438–446, 2015.
- [50] M. C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009.
- [51] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers—a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 35, no. 4, pp. 476–487, 2005.
- [52] Q. Zhao, V. Hautamaki, and P. Fränti, "Knee point detection in BIC for detecting the number of clusters," in *Proceedings of the International conference on advanced concepts for intelligent vision systems*, vol. 5259, pp. 664–673, 2008.
- [53] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 15, no. 1, pp. 1–13, 2018.

