# Towards a data-driven adaptive anomaly detection system for human activity

Salisu Wada Yahaya, Ahmad Lotfi, Mufti Mahmud*

*Department of Computer Science, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK*

## ARTICLE INFO

## ABSTRACT

Research in the field of ambient intelligence allows for the utilisation of different computational models for human activity recognition and abnormality detection to promote independent living and to improve the quality of life for the increasing ageing population. The existing monitoring systems are not adaptive to the overly changing human behavioural routine leading to a high rate of false predictions. An adaptive system pipeline is proposed in this paper for adapting to changes in human behaviour based on data ageing and data dissimilarity forgetting factors. The forgetting factor feature allows adaptation of the model to the current routines of an individual while forgetting outdated behavioural patterns. The data ageing forgetting factor discard old behavioural routine based on the age of the activity data while in the data dissimilarity approach, this is achieved by measuring the similarity of the activity data. Behaviour modelling is achieved using an ensemble of novelty detection models termed as Consensus Novelty Detection Ensemble consisting of One-Class Support Vector Machine, Local Outlier Factor, Robust Covariance Estimation and Isolation Forest. The proposed approach is data-driven and environment-invariant, making it feasible for deployment in heterogeneous environments. A comparative analysis carried out with other abnormality detection models for human activities across two datasets shows that the proposed approach achieved better results.

## 1. Introduction

Machine learning models create a mapping of input data to the desired output using complex mathematical functions [1,2]. For a classification problem, the input data are mapped to the class labels that are known beforehand [2–4]. Similarly, unsupervised models for outlier detection undergo similar training routine to identify a cluster boundary separating the normal data from the anomalous data (outliers) [2,5,6]. This requires a sufficient amount of training data to achieve an acceptable result. In a scenario with limited training data, or where the data is overly dynamic and changes over time, these conventional models can be inefficient. Hence, a model capable of learning continuously and adapting to new data is desirable.

A continuous learning system incorporating models capable of adapting to novel data that can be deployed in a dynamic environment can play a significant role in Human Activity Recognition (HAR), behaviour modelling and abnormality detection. Research progress in the field of Ambient Intelligence (AmI) allows

for the realisation of an in-home monitoring system for Activities of Daily Living (ADL) of older adults [7–11] with important applications such as fall detection [12–14]. Studies have shown that the increase in the global ageing population leads to the increase in the cost of care for the older adults, demand for human carers and financial burden on government and relatives [15–17]. Moreover, results obtained from published surveys show that the older adults prefer to stay in their own homes instead of a care facility [18], prompting for the need for an in-home system for monitoring their well-being. These systems have the potential to promote independent living and improve quality of life of the ageing population [9,18].

The existing systems for human behaviour monitoring and abnormality detection are not often adaptive to changes in behavioural routine. This generates high false alarm rate leading to lack of acceptability by the clients [19,20]. The abnormalities in human behaviour are instances of activities that differ from the usual behavioural routine of an individual [21]. Due to the variability of behaviour among individuals, anomalous instances are identified through modelling of the exiting routines of an individual and classifying deviating data instances as outliers [22,23]. As such, changes in individual routines are always predicted as abnormalities due to lack of adaptability. A self-supervised approach,

---

* Corresponding author.
*E-mail addresses:* ahmad.lotfi@ntu.ac.uk (A. Lotfi), mufti.mahmud@ntu.ac.uk (M. Mahmud).
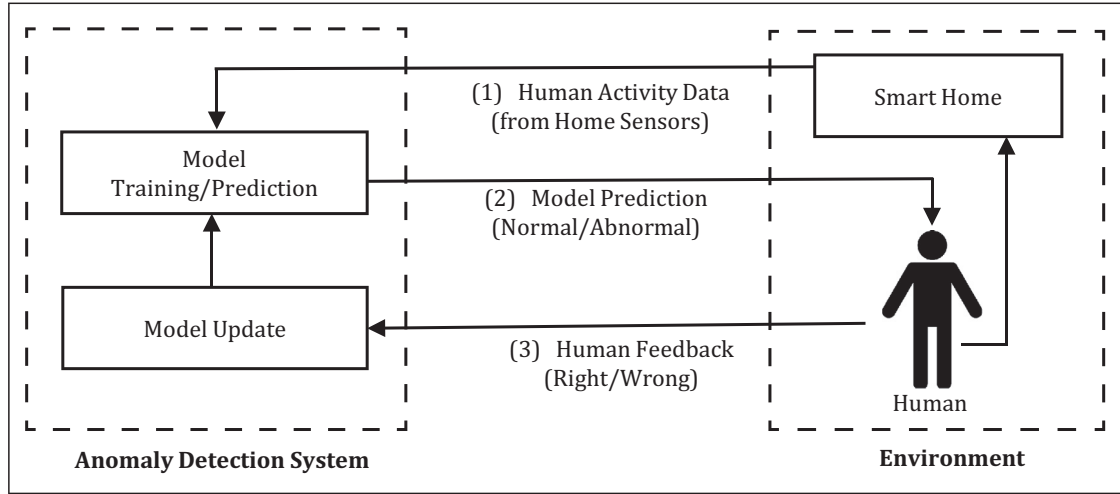
**Fig. 1.** An overview of the adaptive anomaly detection system.

called Abnormal Event Detection Network (AED-Net), for detecting anomaly in crowded scenes is proposed in [24]. Another approach based on Conditional Random Field (CRF) and Mixtures of Dynamic Textures (MDT) is proposed in [25]. However, these two approaches used video stream data, while this proposed approach utilised ambient sensors generated binary ADL data.

The main contribution of this paper is the proposal of an adaptive model based anomaly detection system for ADL. This system adapts to new data corresponding to changes in the human behavioural routines over time. We proposed a data-driven filtering approach termed "Forgetting Factor". The forgetting factor allows the system to identify outdated activity data to be discarded while incorporating newly identified data representing the changes in the human behavioural routine for adaptation. This is a practical approach since the goal is for the system to adapt to the current behavioural routine of an individual while discarding irrelevant characteristics feature of his/her old routines. Two forgetting factor approaches are proposed, namely; Forgetting Factor based on Data Ageing (FFDA) and Forgetting Factor Based on Data Dissimilarity (FFDD). An ensemble of novelty detection models (consisting of models such One-Class Support Vector Machine (OC-SVM) and Isolation Forest (iForest)) is then utilised for the behaviour modelling.

The rest of this paper is organised as follows: Section 2 presents the proposed adaptive approach. In Section 3, the datasets employed for the validation of the methodology, the obtained results and their evaluation along with a discussion is presented. A conclusion of this work is provided in Section 4.

## 2. Adaptive system for abnormality detection

The adaptive model pipeline allows the system to adapt to novel data representing changes in human activities. Novel activity instances may be a reflection of actual abnormality or a behavioural change requiring confirmation from an external agent (such as a human agent). Fig. 1 shows an overview of the proposed adaptive system with support for a human confirmation of detected abnormalities. To realise this, assumptions and constraints in line with research objectives are defined.

[-] *Assumption 1*: Anomalous activities of interest occur consecutively over a time period. Isolated outlier instances may be due to erroneous data, misprediction or a true anomaly.
[-] *Assumption 2*: A consecutive occurrence of anomalous activity could be an indication of behavioural change or abnormality due to health-related challenges. An external com-

munication intermediary can be incorporated to verify the status of the detected outliers, as shown in Fig. 1.
[-] *Assumption 3*: Anomalous instances verified as behavioural changes can be incorporated into the system for adaptation. A forgetting factor is introduced to discard outdated behavioural pattern while incorporating the currently verified routines (data).

### 2.1. Forgetting factor

The forgetting factor is introduced to enable the filtering of activity data to be used for updating the model. The characteristic features of the newly verified data are incorporated into the system by updating the anomaly detection model, while the outdated features are discarded. This is achieved by filtering both the existing and new activity dataset to remove entries that do not conform to the current behavioural routine of the individual. Two forgetting factor approaches are proposed. They are Forgetting Factor based on Data Ageing (FFDA) and Forgetting Factor Based on Data Dissimilarity (FFDD).

#### 2.1.1. Forgetting factor based on data ageing

In this approach, the existing training data entries are discarded based on the age of the data. This is based on the hypothesis that the oldest activity data represents an obsolete behavioural routine, while the recent data represents the current routine of an individual. Since the activity data contains a timestamp, the entries are sorted using the timestamp and the oldest entries are replaced by the newly verified data. A practical limitation of this approach is that the oldest activity data may have more similarity to the current routine than some of the recent entries. Algorithm 1 contains the FFDA implementation procedures.

---

**Algorithm 1** Forgetting Factor based on Data Ageing.

---

1: $D_{\text{Old}}$ is the existing data
2: $D_{\text{New}}$ is the newly verified data      ▷ ($n$ = dataset size)
3: $D_{\text{Final}}$ is the filtered data
4: **procedure** FORGETOLDESTDATA($D_{\text{Old}}, D_{\text{New}}$)
5:      Sort ($D_{\text{Old}}$)      ▷ sort by data age
6:      RemoveOldest($D_{\text{Old}}, n$)      ▷ discard $n$-top entries
7:      $D_{\text{Final}} \leftarrow$ Append($D_{\text{Old}}, D_{\text{New}}$)
8:      **return** $D_{\text{Final}}$
9: **end procedure**

---

### 2.1.2. Forgetting factor based on data dissimilarity

In this approach, the discarded activity data are identified based on a similarity measure. Data entries in the existing data are measured for similarity against the newly verified data. A ranking of the data entries is performed based on linear regression to identify the most dissimilar entries. Unlike in the FFDA approach, where the oldest data are assumed to be less representative of the current behavioural routine, the FFDD is more generic since it relies on the data similarity measure. The FFDD implementation procedures are given in Algorithm 2.

---

**Algorithm 2** Forgetting Factor based on Data Dissimilarity.

1: $D_{Old}$ is the existing data
2: $D_{New}$ is the newly verified data $\qquad \triangleright$ ($n$ = dataset size)
3: $D_{Final}$ is the filtered data
4: **procedure** FORGETDISSIMILARDATA($D_{Old}, D_{New}$)
5: $\quad$ SimMatrix $\leftarrow$ EstimateSimilarity($D_{Old}, D_{New}$)
6: $\quad$ Rank(SimMatrix) $\qquad \triangleright$ using Linear Regression
7: $\quad$ RemoveDissimilarData($D_{Old}, n$)
8: $\quad$ $D_{Final} \leftarrow$ Append($D_{Old}, D_{New}$)
9: $\quad$ **return** $D_{Final}$
10: **end procedure**

---

### 2.2. Data similarity measure

The proposed FFDD performs a similarity measure between the existing and the newly verified activity data entries. The similarity is measured using a Euclidean distance function. Computational models for classification and outlier detection utilises distance as a measure of similarity [26]. For example, K-Mean clustering uses distance measure to estimate the proximity between data entries and the centroids while K-Nearest Neighbour (KNN) uses distance measure such as Euclidean distance to identify the neighbours of data entries for classification [26,27]. The choice of Euclidean distance measure over other distance functions is supported by our recent work in [28] in which Euclidean distance achieved a better result compared to Chebyshev and Canberra distance.

Let $A = \{a_1, a_2, \ldots, a_n\}$ be a $d$-dimensional set of $n$ activity data used for the initial training of the anomaly detection model and $B = \{b_1, b_2, \ldots, b_m\}$ be a $d$-dimensional set of $m$ verified data for adaptation. The similarity $\mathbb{S}(A, B)$ of $A$ and $B$ is a matrix estimated from the Euclidean distance between the pairwise features of the data. Expressing the data $A$ and $B$ in a matrix form:

$$A = \begin{bmatrix} a_1^1 & - & - & a_1^d \\ - & - & - & - \\ - & - & - & - \\ a_n^1 & - & - & a_n^d \end{bmatrix}, B = \begin{bmatrix} b_1^1 & - & - & b_1^d \\ - & - & - & - \\ - & - & - & - \\ b_m^1 & - & - & b_m^d \end{bmatrix} \qquad (1)$$

Given the similarity measure matrix $\mathbb{S}(A, B)$ for $A$ and $B$:

$$\mathbb{S}(A, B) = \begin{bmatrix} \overrightarrow{\mathbb{S}}(a_1, B) \\ - \\ - \\ \overrightarrow{\mathbb{S}}(a_n, B) \end{bmatrix}, \mathbb{S}(A, B) \in \mathbb{R}^{nxd} \qquad (2)$$

The similarity $\mathbb{S}(a_i, B)$ for the $i$th entry of $A$ is calculated as the per feature distance after normalisation of the entry as:

$$\mathbb{S}(a_i, B) = \begin{bmatrix} a_i^1 & - & - & a_i^d \end{bmatrix} \begin{bmatrix} b_1^1 & - & - & b_1^d \\ - & - & - & - \\ - & - & - & - \\ b_m^1 & - & - & b_m^d \end{bmatrix} \qquad (3)$$

$$\mathbb{S}(a_i, B) = \begin{bmatrix} \partial(a_i^1, b_1^1) & - & - & \partial(a_i^d, b_1^d) \\ - & - & - & - \\ - & - & - & - \\ \partial(a_i^1, b_m^1) & - & - & \partial(a_i^d, b_m^d) \end{bmatrix}, \mathbb{S}(a_i, B) \in \mathbb{R}^{mxd} \qquad (4)$$

where $\partial(a_i^k, b_j^k)$ is a 1-dimensional pairwise Euclidean distance expressed as:

$$\partial(a_i^k, b_j^k) = \sqrt{(a_i^k - b_j^k)^2} = |a_i^k - b_j^k| \qquad (5)$$

The matrix $\mathbb{S}(a_i, B)$ is converted into a row vector $\overrightarrow{\mathbb{S}}(a_i, B)$ by calculating its per-feature (column) mean.

$$\overrightarrow{\mathbb{S}}(a_i, B) = mean\left(\mathbb{S}(a_i, B)\right), \quad \overrightarrow{\mathbb{S}}(a_i, B) \in \mathbb{R}^{1xd} \qquad (6)$$

$$\overrightarrow{\mathbb{S}}(a_i, B) = \begin{bmatrix} mean\left(\partial_{i,1}^1, \partial_{i,m}^1\right) & - & - & mean\left(\partial_{i,1}^d, \partial_{i,m}^d\right) \end{bmatrix} \qquad (7)$$

Since $\mathbb{S}(A, B)$ consist of row vectors, the vectors are converted into a scalar (ranked) for easy sorting as mentioned in Algorithm 2. The ranking $\mathbb{G}(A, B)$ of $\mathbb{S}(A, B)$ is determined using a linear regression model. Linear regression is expressed as the linear combination of the attributes(features) with their corresponding weights as expressed below:

$$\mathbb{G} = w^0 + w^1\mathbb{S}^1 + w^2\mathbb{S}^2 + \ldots + w^d\mathbb{S}^d = \sum_{k=1}^{d} w^k\mathbb{S}^k \qquad (8)$$

where $[w^1, \ldots, w^d]$ are the feature weights, $w^0$ is the error term set to 0, $\mathbb{S}^k$ is a scalar representing the value of the $k$th column for the given row, and $d$ is the number of features in the dataset.

Principal Component Analysis (PCA) is applied during the training of the anomaly detection model to determine the most discriminating features of the dataset. The weights of the features (ranging from $[0, 1]$) is taken as the percentage fraction of the ranking of the features based on PCA using the expression $(1 - \frac{\text{Feature Rank (r)}}{\text{Number of Features (d)}})$. The weight of the most discriminating feature is taken as $(1 - \frac{1}{d})$, while that of the second most discriminating feature is given as $(1 - \frac{2}{d})$ and vice versa. Algorithm 3 outlines the procedure for computing the similarity measure for sample datasets $A$ and $B$.

### 2.3. Adaptive system pipeline

The approaches described earlier are combined to form the adaptive system pipeline as shown in the schematic diagram in Fig. 2. This proposed adaptive system is activity-dependent and data-driven. Its goal is to adapt to the newly verified data while forgetting the characteristic features of the outdated data identified using the proposed forgetting factor approaches (i.e. FFDA and FFDD). The forgetting feature makes the system environment invariant and allows for easy adaptation to a different environment and datasets.

Considering that the proposed approach may be similar to a brute-force method of appending (merging) the newly verified data to the existing one, a significant difference is that this approach allows the system to forget the old behavioural routine data completely without retaining its characteristic features while a brute-force approach still retains features of the old behavioural data.

From the illustration in Fig. 2, the training of the anomaly detection models (such as OC-SVM) is performed with the initial set of collected data. A copy of the data and the trained model are saved on a storage component of the system for future use. Subsequent data collected after system deployment is validated by the model to detect outliers (anomalies). The result of the model prediction can either be normal or anomaly for any given activity. In
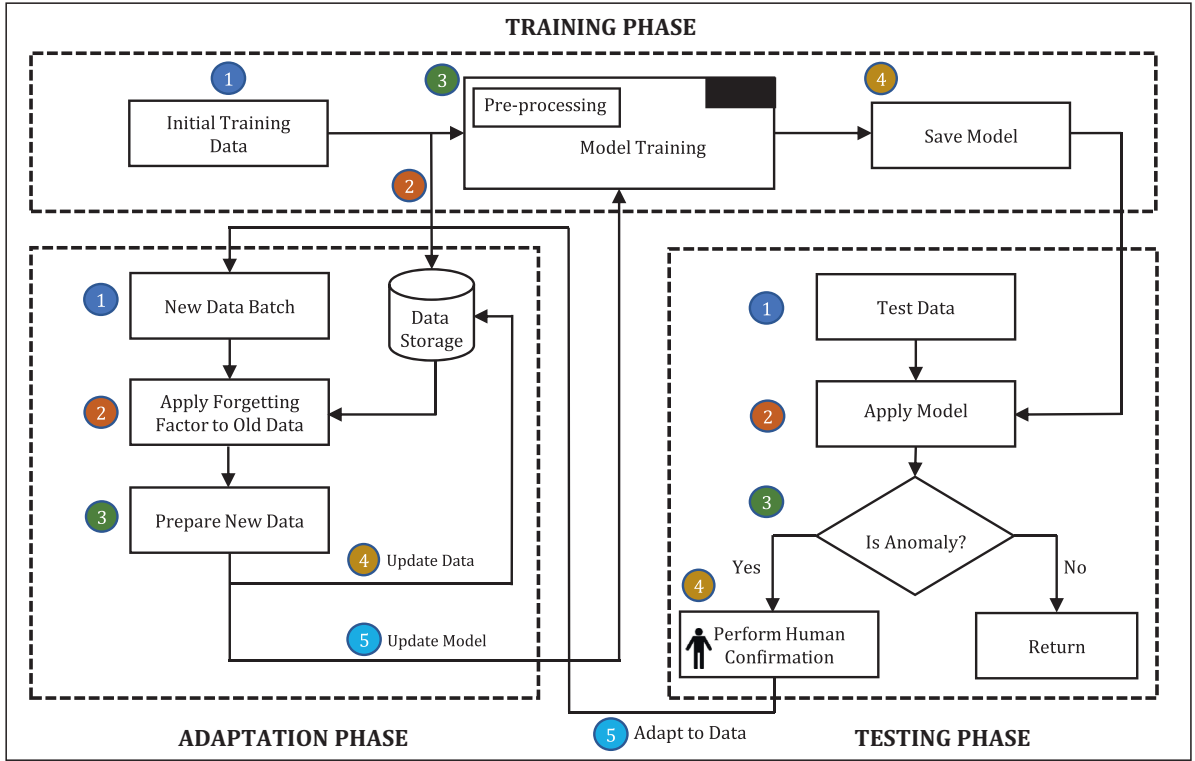
**Fig. 2.** A schematic diagram of the data-driven adaptive system pipeline.

**Algorithm 3** Procedure for computing Similarity Measure.

1: $A = \{a_i\}, i = 1, \ldots, n$ is the initial training data
2: $B = \{b_j\}, j = 1, \ldots, m$ is newly verified data
3: $\mathbb{S}(A, B)$ is an $[n, d]$ similarity matrix for $A$ & $B$
4: **procedure** ESTIMATESIMILARITY($A, B$)
5:     **for** $a_i \in A$ **do**
6:         $\mathbb{S}(a_i, B)$ is an $[m, d]$ similarity matrix
7:         **for** $b_j \in B$ **do**
8:             $\vec{V_r}$ is a $[1, d]$ row vector
9:             **for** $[k := 1 : d]$ **do**       ▷ for-each feature
10:                $\partial := distance(a_i^k, b_j^k)$    ▷ euclidean dist.
11:                $\vec{V_r} \leftarrow \partial$       ▷ insert column to $\vec{V_r}$
12:             **end for**
13:             $\mathbb{S}(a_i, B) \leftarrow \vec{V_r}$    ▷ insert row to matrix
14:         **end for**
15:         $\vec{\mathbb{S}}(a_i, B)$ is a $[1, d]$ row vector
16:         **for** $[k := 1 : d]$ **do**      ▷ for-each feature
17:             $\alpha^k := mean\left[\mathbb{S}\left(a_i^k, B_{\{b=1,\ldots,m\}}^k\right)\right]$
18:             $\vec{\mathbb{S}}(a_i, B) \leftarrow \alpha^k$   ▷ insert as column
19:         **end for**
20:         $\mathbb{S}(A, B) \leftarrow \vec{\mathbb{S}}(a_i, B)$   ▷ insert row to matrix
21:     **end for**
22:     **return** $\mathbb{S}(A, B)$
23: **end procedure**

an ideal scenario, a single abnormality prediction may not constitute a major concern as this may be due to erroneous sensor reading or model misprediction while consecutive repeated anomalies will undergo human verification as highlighted in the defined constraints. In a situation where the prediction is confirmed to be wrong and the activities are identified as normal behavioural changes, the stored data is retrieved from the storage component.

The proposed forgetting factor approaches are applied to both the existing and newly verified activity data. A model update is performed and the storage component is updated with the consolidated data entries to enable future occurrences of similar activity pattern to be identified as normal behavioural routine.

While this paper puts more emphasis on adapting to changes in behavioural routine that are wrongly identified as outliers, the adaptive system can benefit from periodic model updates at regular intervals (e.g., weekly, monthly, quarterly, etc. based on deployment requirements and data size) with data of normal behavioural routines (data that are not identified as outliers) to ensure that the model reflects the current behaviour of the monitored individuals.

The reporting and confirmation aspect of the detected abnormalities are outside the scope of this paper. This paper only aims to show how the model is able to adapt to new data representing the changes in the behavioural routine of an individual. However, the confirmation of anomalies can be achieved through different modalities. We presented an approach for incorporating human in the loop in [29] highlighting the role of a human agent in the learning process of the system while utilising an assistive robot as intermediary. As shown in Fig. 1, the detected anomalies are confirmed from the human agent before the data entries are transferred to the computational model for adaptation. Therefore, the role of the human agent is to confirm whether detected activities are abnormal or not with the help of a communication intermediary. To facilitate the Human-Robot Interaction (HRI), different communication modalities are considered, such as touch, speech and gesture. A gesture recognition model for the HRI using a 2D camera, and using a wrist-worn triaxial accelerometer are presented in our works in [30] and [31] respectively.

## 3. Experimentation and discussion

In this section, the ADL datasets utilised for the validation of the proposed model pipeline, as well as the sensing devices and
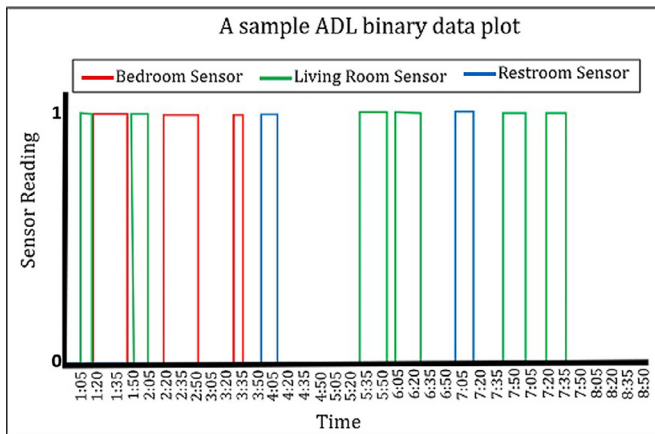
**Fig. 3.** A sample plot of raw binary data from ambient sensors.

**Table 1**
A sample of interpreted ADL data.

| Start Time | End Time | Activity | Location |
|---|---|---|---|
| 21:10:04 | 03:11:07 | Sleeping | Bedroom |
| 03:11:51 | 03:14:11 | Toileting | Toilet |
| 03:14:52 | 07:13:07 | Sleeping | Bedroom |
| 07:15:47 | 07:22:04 | Preparing Meal | Kitchen |
| 07:25:01 | 09:14:17 | Watching TV | Living Room |

data collection modalities are described. The preprocessing technique, experimental scenarios and the obtained results are also presented.

### 3.1. Datasets and preprocessing

The two datasets utilised for this evaluation include data collected as part of this research work from the smart home facility of Nottingham Trent University (NTU) termed as "SmartNTU" data and a publicly available data provided by the Center for Advanced Studies in Adaptive Systems (CASAS) termed as "CASAS HH111" [32]. For the data collection, non-invasive, low cost ambient sensors capable of generating binary data are utilised such as Passive Infrared sensors (PIR) for identify human presence in the smart home, pressure sensors installed on bed and sofas to identify sleeping and sitting activity, contact sensors installed on doors to detect opening and closing etc. These sensors are utilised because they are generally more acceptable than visual sensors (such as camera) due to privacy concerns [32].

The CASAS HH111 dataset contains the recorded daily activity for a volunteer older adult living alone for a duration of 50 days [32]. The activities recorded include but not limited to sleeping, toileting, eating, watching TV, kitchen-related activities etc. For the SmartNTU data, the duration of the recorded activity data is 72 days. The activities recorded from the individual is similar to that of the CASAS HH111 data. A sample plot of the binary data generated by the ambient sensors is shown in Fig. 3. A sample of the interpreted human activities is displayed in Table 1 showing the different attributes of the performed activities such as the start time, end time, location and the activity type.

The datasets are preprocessed by removing instances with missing entries and erroneous readings. The data is filtered and sleeping activity is selected as the activity of interest to be modelled. The extracted features from the sleeping activity listed below are normalised since the features are in different scales.

- *Start time*: represents the starting time of the activity.
- *Duration*: represents the length of the activity in minutes.

- *Day*: represents the day of the performed activity.
- *Day group*: indicate if the activity falls on a weekend or a weekday.
- *Transition*: is the number of transition within the activity (i.e., sleep interruption).
- *Transition length*: indicate the duration of the transitions (i.e. duration of the interruptions)

### 3.2. Experimental scenario

The experimental scenario for the proposed approaches is described in this section. The proposed adaptive system is activity dependent (i.e. for optimal performance, activities of interest should be selected and modelled independently). The sleeping activity is selected and modelled. The system capability of forgetting outdated data while adapting to the newly verified ADL data is evaluated. To achieve this, a cross-dataset evaluation strategy is performed such that the anomaly detection model is trained on one dataset and evaluated for adaptation on the other dataset. For example, the model is trained on the SmartNTU dataset and adapted to the CASAS HH111 dataset and vice versa. This validation strategy is adopted to ensure that the system is environment invariant, i.e. the system can be migrated to a different environment with easy adaptation.

The baseline anomaly detection model in the adaptive system is an ensemble approach for novelty detection models proposed in our work in [23]. The approach termed as "Consensus Novelty Detection Ensemble (CNDE)" identify outliers by estimating the score of an activity (called "Normality Score") with activities whose score exceeds a certain threshold predicted as anomalous. The ensemble approach consists of four models, namely; OC-SVM with Radial Basis Function (RBF) kernel, iForest, Local Outlier Factor (LOF) and Robust Covariance Estimation (RCE). An acceptable result is achieved using 31 days training data with the contamination rate set to 0.1 across all the four models in the ensemble. This baseline ensemble model is not adaptive by default, therefore, novel (new) data entries are always predicted as outliers. A more detailed discussion of the model and the obtained results can be found in [23].

To validate the adaptive approach proposed in this paper, the cross dataset validation technique is used. For example, to evaluate the approach on the SmartNTU dataset, the first 31 days data is used for training while the remaining data (referred to as a segment of the training dataset) is used for validating the error rate of the model. The CASAS HH111 data is then used to validate the model adaptation rate, and the data is split into an adaptation set and a validation set with a duration of 30 and 35 days data respectively. The adaptation set is injected into the system incrementally using different batch sizes (i.e. 5, 10 and 15 days data). After each incremental injection of the adaptation data, the two validation sets (i.e. a segment of the training data for evaluating the model error rate and a segment of the cross-dataset for evaluating the adaptation rate) are used. The system is expected to adapt to the new dataset after a certain number of incremental injection of the adaptation data, therefore, the adaptation rate is expected to increase over time while the error rate is expected to decrease. To reduce the effect of class imbalance, the datasets with insufficient sample entries for the adaptation and validation are oversampled using Synthetic Minority Over-sampling Technique (SMOTE) proposed in [33].

The performance of the proposed approaches with forgetting factor are computed and compared with other anomaly detection models namely, OC-SVM, iForest, and the aforementioned CNDE ensemble approach (using the same model parameters). A Deep Q-Network (DQN) is also implemented with two actions (normal and outlier) with the rewards of +1 and −1. The datasets are sufficiently over-sampled prior to training and the weights are
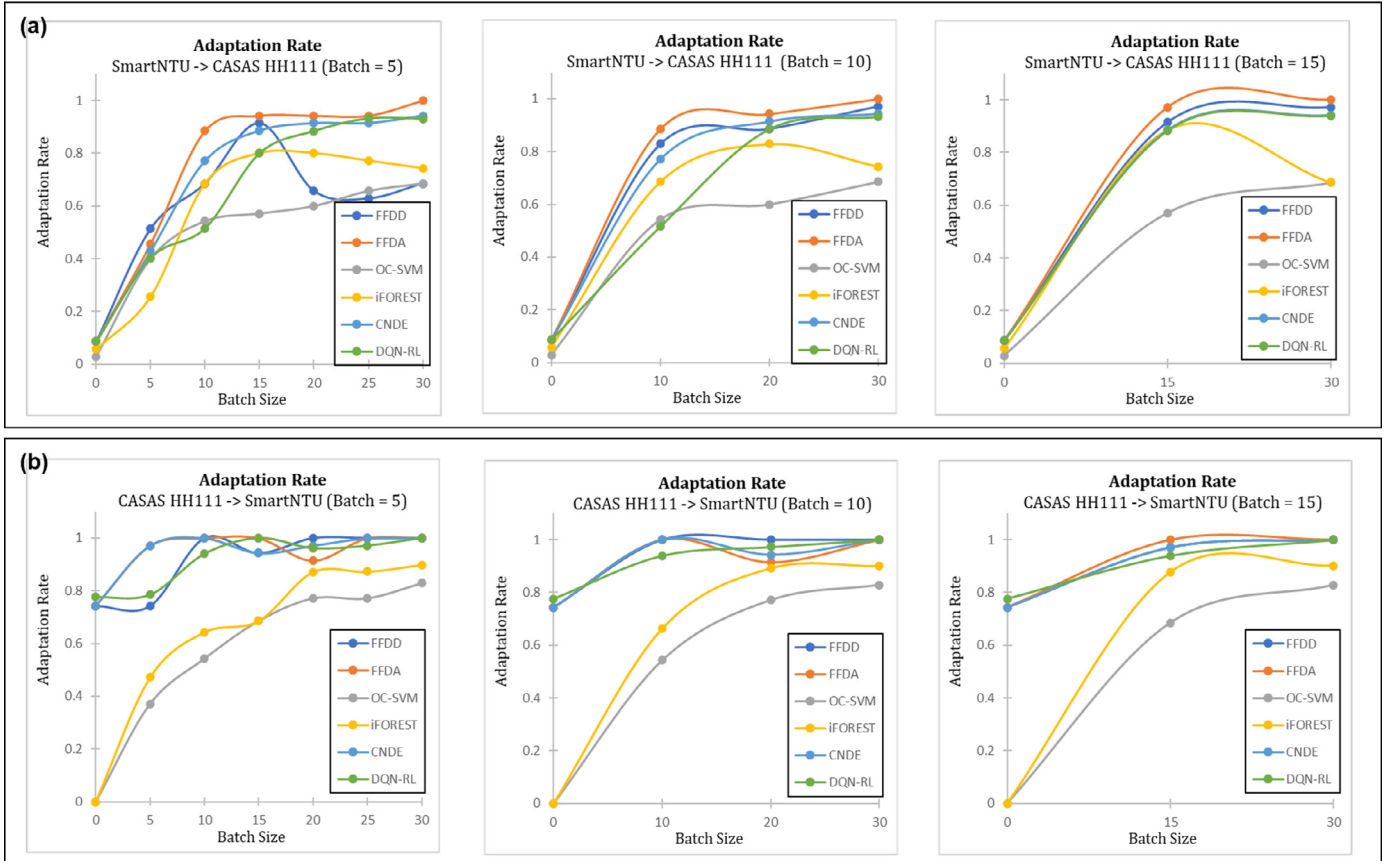
**Fig. 4.** Adaptation rate plots; a) SmartNTU to CASAS HH111, b) CASAS HH111 to SmartNTU.

saved when the training is complete. During the adaptation phase, the adaptation data batch is used to re-train the saved model weights instead of training the model from scratch. For the baseline anomaly detection models, since the models are not adaptive by default, a brute-force injection of the adaptation data is performed, i.e. the adaptation data are appended to the original training data incrementally.

### 3.3. Evaluation and discussion

The results obtained for the evaluation of the proposed adaptive approaches are shown in Figs. 4 and 5. The performance metrics measured are the adaptation rate and error rate. The adaptation rate is the rate at which the system adapts to the newly verified data (adaptation). This is achieved by calculating the prediction accuracy of the validation set. The adaptation rate is expected to increase over time as more data representing the new behavioural routine are injected into the model. The error rate (forgetting inverse) is the rate at which the system forgets the initial training data (old data). This is obtained by calculating the error rate of the model prediction on the segment of old training data. As highlighted in the previous section, the initial training data is split into a training set and a validation set prior to the model training. This remaining segment (validation set from the initial training data), is used to compute the error rate. It should be noted that this data segment is not used for model training. The model is expected to predict this data segment as outliers, and an inlier prediction is considered an error. The error rate is expected to decline over time since the newly injected data (adaptation data) is from a different dataset and differ from the training data segment causing the model to gradually forget the characteristic features of the ini-

tial training data. Obtaining a lower error rate is the final goal as this indicate that the model predicts the segment of the training as outliers after adaptation.

It can be seen from Fig. 4 that proposed approaches (FFDA and FFDD) outperformed the Reinforcement Learning (RL) model as well as the classical novelty detection models. The adaptation rate of the CNDE and DQN-RL is comparably close to that of the proposed approaches. The FFDD reaches the peak of its performance faster than FFDA. This is because the FFDD remove dissimilar data at each iteration while FFDA only discards data based on the data age. The error rate plots in Fig. 5 shows that only the proposed approaches in this paper can forget (discard) the outdated data while the other models (i.e. CNDE, OC-SVM and iForest) performed poorly. This is expected since the models are not adaptive and the new adaptation data are appended using a brute-force approach. The RL model adapts to the adaptation data without retraining on the entire dataset by using the pre-trained weights, but still retains the properties of the old data. Although the error rate can be improved by fine-tuning the model parameters, the properties of the old data may still be retained (this is against our goal of forgetting the data in its entirety). The FFDD approach converges faster than the FFDA since dissimilar entries are removed on each iteration. The FFDA will eventually converge when the old data entries are discarded based on their age with longer convergence time for large datasets. The best results obtained for the adaptation and error rates are summarised in Tables 2 and 3 respectively.

The batch size for the incremental adaptation has a significant effect on the convergence rate as it can be seen in Figs. 4 and 5. A small batch (e.g. 5 days) will result in regular system update with a slow convergence rate (more prediction errors) and vice versa. From the obtained results, it can be seen that the performance de-
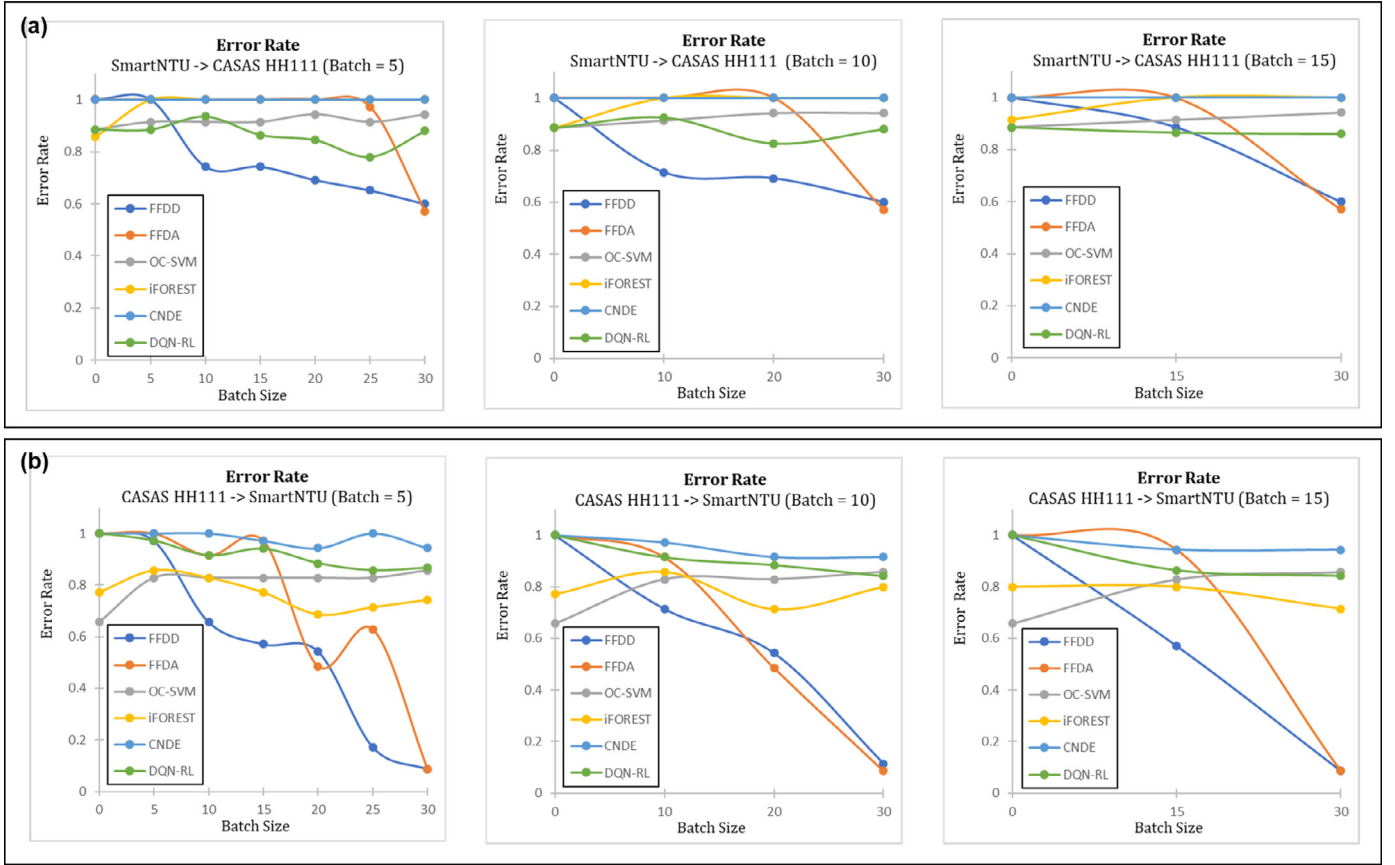
**Fig. 5.** Error Rate (Forgetting Inverse) plots; a) SmartNTU to CASAS HH111, b) CASAS HH111 to SmartNTU.

**Table 2**
A summary of the highest achieved adaptation rates.

| Approach | Training: SmartNTU | Training: CASAS HH111 |
|---|---|---|
| iForest | 0.886 | 0.890 |
| OC-SVM | 0.686 | 0.829 |
| CNDE | 0.943 | 0.982 |
| DQN-RL | 0.940 | 0.982 |
| FFDA | **0.992** | **0.982** |
| FFDD | **0.971** | **0.983** |

**Table 3**
A summary of the lowest achieved error rates.

| Approach | Training: SmartNTU | Training: CASAS HH111 |
|---|---|---|
| iForest | 0.857 | 0.686 |
| OC-SVM | 0.886 | 0.657 |
| CNDE | 0.989 | 0.914 |
| DQN-RL | 0.780 | 0.841 |
| FFDA | **0.571** | **0.086** |
| FFDD | **0.600** | **0.086** |

teriorates after a peak is reached. This is due to overfitting as a result of the continuous model update on a similar data entries. A model update can be stopped when the characteristic features of the adaptation data are similar to the existing ones. While the model update is performed continuously in this paper, stopping the update can be realised in a real-life scenario through the human confirmation mechanism.

## 4. Conclusion

This paper presents a data-driven adaptive system pipeline for detecting abnormalities in human activities. The adaptive system is based on data ageing and data dissimilarity forgetting factor approaches referred to as FFDA and FFDD. These approaches enable the anomaly detection model in the system to adapt to changes in human behavioural routine while discarding the characteristic features of old routines. Experimental evaluation of the proposed approaches achieved an overall average adaptation accuracy of 98% on both validation datasets, thereby outperforming the conventional anomaly detection models. The FFDD discard dissimilar data by measuring the similarity of the existing data with new activity data while the FFDA discard data based on its age. This results in faster convergence time for the FFDD since dissimilar data are discarded at each iteration. The batch size of the adaptation data also have an effect on the forgetting factor approaches as small batch size results in a regular model update but leads to a high rate of false prediction while a large batch size results in less error rate with slower model update. Future work plans include the deployment and validation of the system in a real home environment, as well as performing trend analysis on a longitudinal ADL data.

## Declaration of Competing Interest

I, Mufti Mahmud, the corresponding author of this presented work declare, on behalf of the authors, that the authors have nothing that can construct to a possible conflict of interest.

## Acknowledgements

# References

[1] Y. Li, Q. Pu, S. Li, H. Zhang, X. Wang, H. Yao, L. Zhao, Machine learning methods for research highlight prediction in biomedical effects of nanomaterial application, Pattern Recognit. Lett. 117. (2019) 111–118.

[2] M.R.M. Talabis, R. McPherson, I. Miyamoto, J.L. Martin, D. Kaye, Analytics defined, in: Information security analytics, Syngress (2015) 1–12.

[3] M. Mahmud, M.S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, IEEE Trans. Neural Netw. Learn. Syst. 29. (6) (2018) 2063–2079.

[4] M. Mahmud, M.S. Kaiser, T.M. McGinnity, A. Hussain, Deep learning in mining biological data, Cogn. Comput. 13. (1) (2021) 1–33.

[5] J.R. Pasillas-Díaz, S. Ratté, An unsupervised approach for combining scores of outlier detection techniques, based on similarity measures, Electron. Notes Theor. Comput. Sci. 329. (2016) 61–77.

[6] M.A. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, Signal Process. 99. (2014) 215–249.

[7] Z. Wang, C. Zhao, S. Qiu, A system of human vital signs monitoring and activity recognition based on body sensor network, Sens. Rev. 34. (1) (2014) 42–50.

[8] P.N. Dawadi, D.J. Cook, M. Schmitter-Edgecombe, Automated cognitive health assessment from smart home-based behavior data, IEEE J. Biomed. Health Inform. 20. (4) (2016) 1188–1194.

[9] A. Alberdi, A. Weakley, M. Schmitter-Edgecombe, D.J. Cook, A. Aztiria, A. Basarab, M. Barrenechea, Smart home-based prediction of multidomain symptoms related to Alzheimer's disease, IEEE J. Biomed. Health Inform. 22. (6) (2018) 1720–1731.

[10] A.A. Aramendi, A. Weakley, A.A. Goenaga, M. Schmitter-Edgecombe, D.J. Cook, Automatic assessment of functional health decline in older adults based on smart home data, J. Biomed. Inform. 81. (2018) 119–130.

[11] S.W. Yahaya, A. Lotfi, M. Mahmud, Detecting anomaly and its sources in activities of daily living, SN Comput. Sci. 2. (1) (2021) 1–18.

[12] M.J.A. Nahian, et al., Towards artificial intelligence driven emotion aware fall monitoring framework suitable for elderly people with neurological disorder, in: Proc. Int. Conf. Brain Inform., 2020, pp. 275–286.

[13] M. Nahiduzzaman, et al., Machine learning based early fall detection for elderly people with neurological disorder using multimodal data fusion, in: Proc. Int. Conf. Brain Inform., 2020, pp. 204–214.

[14] M.J.A. Nahian, et al., Towards an accelerometer-based elderly fall detection system using cross-disciplinary time series features, IEEE Access 9 (2021) 1–19 In press, doi:10.1109/ACCESS.2021.3056441.

[15] S. Chernbumroong, S. Cang, A. Atkins, H. Yu, Elderly activities recognition and classification for applications in assisted living, Expert Syst. Appl. 40 (5) (2013) 1662–1674.

[16] A. Lotfi, C. Langensiepen, S.M. Mahmoud, M.J. Akhlaghinia, Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour, J. Ambient Intell. Humaniz. Comput. 3. (3) (2012) 205–218.

[17] Q. Ni, A.B.G. Hernando, I.P. De la Cruz, The Elderlys independent living in smart homes: a characterization of activities and sensing infrastructure survey to facilitate services development, Sensors 15 (5) (2015) 11312–11362.

[18] P. Rashidi, A. Mihailidis, A survey on ambient-assisted living tools for older adults, IEEE J. Biomed. Health Inform. 17. (3) (2013) 579–590.

[19] E. Hoque, R.F. Dickerson, S.M. Preum, M. Hanson, A. Barth, J.A. Stankovic, Holmes: a comprehensive anomaly detection system for daily in-home activities, in: Proc. DCOSS, 2015, pp. 40–51.

[20] K.Z. Haigh, L.M. Kiff, G. Ho, The independent lifestyle assistant: lessons learned, Assistive Technol. 18. (1) (2006) 87–106.

[21] D. Arifoglu, A. Bouchachia, Activity recognition and abnormal behaviour detection with recurrent neural networks, Procedia Comput. Sci. 110 (2017) 86–93.

[22] V. Jakkula, D.J. Cook, Detecting anomalous sensor events in smart home data for enhancing the living experience, in: Proc. 7th AAAI Conf. Artif. Intell. Smarter Living: Conqu. Complex., 2011, pp. 33–37.

[23] S.W. Yahaya, A. Lotfi, M. Mahmud, A consensus novelty detection ensemble approach for anomaly detection in activities of daily living, Appl. Soft Comput. 83 (2019) 105613.

[24] T. Wang, Z. Miao, Y. Chen, Y. Zhou, G. Shan, H. Snoussi, AED-Net: an abnormal event detection network, Engineering 5. (5) (2019) 930–939.

[25] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly detection and localization in crowded scenes, IEEE Trans. Pattern Anal. Mach. Intell. 36. (1) (2013) 18–32.

[26] D.J. Weller-Fahy, B.J. Borghetti, A.A. Sodemann, A survey of distance and similarity measures used within network intrusion anomaly detection, IEEE Commun. Surv. Tutor. 17. (1) (2015) 70–91.

[27] L.-Y. Hu, M.-W. Huang, S.-W. Ke, C.F. Tsai, The distance function effect on k-nearest neighbor classification for medical datasets, SpringerPlus 5 (2016) 1304.

[28] S.W. Yahaya, A. Lotfi, M. Mahmud, A similarity measure approach for identifying causes of anomaly in activities of daily living, in: Proc. PETRA, 2019, pp. 575–579.

[29] S.W. Yahaya, A. Lotfi, M. Mahmud, A framework for anomaly detection in activities of daily living using an assistive robot, in: Proc. UKRAS2019, 2019, pp. 131–134.

[30] S.W. Yahaya, A. Lotfi, M. Mahmud, P. Machado, N. Kubota, Gesture recognition intermediary robot for abnormality detection in human activities, in: Proc. SSCI, 2019, pp. 1415–1421.

[31] D.O. Anderez, L.P. Dos Santos, A. Lotfi, S.W. Yahaya, Accelerometer-based hand gesture recognition for human-robot interaction, in: Proc. SSCI, 2019, pp. 1402–1406.

[32] D.J. Cook, A.S. Crandall, B.L. Thomas, N.C. Krishnan, CASAS: a smart home in a box, Computer 46. (7) (2013) 62–69.

[33] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16. (2002) 321–357.