**Noname manuscript No.**
(will be inserted by the editor)

# Detection of Abnormal Behaviour for Dementia Sufferers using Convolutional Neural Networks

**Damla Arifoglu, Abdelhamid Bouchachia**

Received: date / Accepted: date

**Abstract** In recent years, there is a rapid increase in the population of elderly people. However, elderly people may suffer from the consequences of cognitive decline, which is a mental health disorder that primarily affects cognitive abilities such as learning, memory, etc. As a result, the elderly people may get dependent on caregivers to complete daily life tasks. Detecting the early indicators of dementia before it gets worsen and warning the caregivers and medical doctors would be helpful for further diagnosis. In this paper, the problem of activity recognition and abnormal behaviour detection is investigated for elderly people with dementia. First of all, the paper presents a methodology for generating synthetic data reflecting on some behavioural difficulties of people with dementia given the difficulty of obtaining real-world data. Secondly, the paper explores Convolutional Neural Networks (CNNs) to model patterns in activity sequences and detect abnormal behaviour related to dementia. Activity recognition is considered as a sequence labelling problem, while abnormal behaviour is flagged based on the deviation from normal patterns. Moreover, the performance of CNNs is compared against the state-of-art methods such as Naïve Bayes (NB), Hidden Markov Models (HMMs), Hidden Semi-Markov Models (HSMM), Conditional Random Fields (CRFs). The results obtained indicate that CNNs are competitive with those state-of-art methods.

**Keywords** Smart Homes · Sensor based Activity Recognition · Convolutional Neural Networks · Long Short Term Memory Recurrent Neural Networks · Dementia · Abnormal Behaviour Detection

Department of Computing and Informatics
Faculty of Science and Technology
Bournemouth University, UK
E-mail: {darifoglu, abouchachia}@bournemouth.ac.uk

## 1 Introduction

Studies indicate that by year 2030, the number of people aged 65 to 74 will be about 3% of the total population [1]. Elderly people may suffer from the consequences of dementia, which is a condition that causes problems with mobility, physical and mental abilities such as memory and thinking [2]. It may also cause decrease in the ability of speaking, writing, distinguishing objects, performing motor activities and performing complex functional tasks (paying bills, preparing a meal, etc.) [3]. An elderly person having such cognitive decline loses independence in daily life and requires care and support from caregivers. On the other hand, the use of assisted living technologies such as smart homes can substantially help a person with dementia to live independently. Unfortunately, currently there are no dementia friendly smart homes addressing elderly people's special needs.

Cognitive diseases, like dementia, need to be detected at an early stage so that early treatment will be possible. However, research shows that 75% of dementia cases go unnoticed [4] and many cases are diagnosed only when the impairment reaches moderate or advanced stage. The best markers of cognitive decline may not necessarily be detected based on a person's performance at any single point in time, but rather by monitoring the trend over time and the variability of change in a duration [5]. Most common types of dementia (Alzheimer, Parkinsons disease) can be identified by behavioural changes like sleep disturbances, difficulty of walking and inability to complete tasks. Thus, such changes can provide key information about memory, mobility and cognition of a person. For instance, an old person suffering from Alzheimer may forget to have his lunch, take multiple lunches instead, wake up in the middle of the night, go to the toilet frequently, or have dehydration problems because of forgetting to drink daily amount of water. In particular, the daily home activity involving basic functions like preparing food, showering, walking, sleeping, etc. can be used to assess the well-being of elderly people.

The development of ambient home assessment environments has begun to provide the opportunity to assess behaviour change unobtrusively in real-time [6,4,5]. Prevention or delay of dementia onset is contingent upon the ability to detect early, meaningful, cognitive change during the life course [6,4]. The identification of early onsets of dementia using non-medical diagnosis methods requires the development of new diagnostic tools. Although a few promising methods have been experimentally validated [6,7,8,9,10], the translation of the current knowledge into smart homes still requires more dedication and work. Current assessment methods mostly rely on queries from questionnaires or in-person examinations, which depend on recall of events or brief snap-shots of function that may poorly represent a person's typical state of function. Also the clinical methods have some limitations such as their episodic nature, and possible biased reporting. The main motivation for our work is that cognitive decline can be observed in daily activities and routines of an elderly person. Real-time monitoring of activities performed by an elderly person in a smart home would be beneficial for early detection of such decline.

In machine learning, a convolutional neural network (CNN) is a class of deep, feed-forward artificial neural networks. Recently, CNNs are popular due to their ability to learn fruitful representations and capture local dependency and spatial information of granular-level patterns. For example, in image recognition, CNNs firstly detect pixels, then edges and shapes, then parts of objects as the layer level increases. Similar to images, there are granular-level patterns in daily life activities. For example, when the activity *preparing coffee* is considered, it is seen that this activity is constructed by many steps such as getting closer to the sink, turning the water on, filling the coffee machine with water and turning the machine on, etc. In [11], granular-level activity patterns, which they call as movement vectors, are extracted by using a decomposition based unsupervised approach. It is shown that the movement vector can distinguish different high-level activities. The occupant tends to have the same routine of performing the same activities, but has different movement patterns in different activities. For example, the occupant may mainly move around the kitchen sink in *Wash Dishes* activity, and stay around the bedroom area during *Sleeping* activity. A combination of some motion sensors are mostly seen in the instances of *relax* activity while usage of some other motion sensors indicates the movement between kitchen range and the sink and this movement pattern is seen in the instances of *wash dishes* activity. CNNs are good at modelling these granular-level patterns and defining their relationship with each other by using spatial information. Thus, in the present study, CNNs are exploited to model sensors and their relationship with each other in daily life activity recognition.

Unfortunately, there exists no publicly available dataset on abnormal behaviour of people with dementia. Producing such a dataset requires time and adequate experimental environment. When there is no real-world dataset available, data simulation can be a solution [12,13,14,15]. Given the scarcity of such data, simulating daily life abnormal behaviours of elderly people suffering from dementia would be helpful for providing automatic assessment methods. Thus, in this paper, a method is proposed to artificially produce abnormal activities reflecting on typical behaviour of elderly people with dementia.

In a nutshell, the present paper introduces the following contributions.

1. A method is proposed to generate synthetic data that simulates the abnormal behaviours of people with dementia.
2. To the best of our knowledge, our study is the first to apply CNNs, thanks to their ability to model granular-level patterns, for daily life activity recognition and dementia related anomaly detection task.

The rest of the paper is organised as follows. Section 2 provides an overview of literature work. Section 3 presents the details of the proposed methodology together with the datasets and models used. Section 4 describes the experimental set-up and results of the experiments followed by a discussion. Finally, Section 5 concludes the paper.

## 2 Literature Review

In-home automatic assessment of cognitive decline has been the subject of many studies [16,17,6,7,18,19,20]. Many machine learning approaches such as SVMs and Naïve Bayes methods [21,22], Restricted Boltzmann Machines (RBMs) [19] and Markov Logic Networks [7,18,20], Hidden Markov Models (HMMs) [14], Random Forest methods [15], and Hidden Conditional Random Fields [23] have been exploited.

In some studies [16,24], assessment of cognitive status is done by providing patients some instructions during the completion of pre-defined tasks (e.g., sweeping the kitchen). In the end, the patients receive scores which are calculated based on the time spent, the frequency of the sensor triggered, etc. These scores are used to assess the cognitive status of elderly people. In [16], cognitive decline assessment is done by asking elderly people to complete a sequence of scripted instrumental activities of daily living (IADLs). The participants are monitored via a camera while they perform tasks such as cooking oatmeal on the stove and in the end, they receive scores by trained experts. In [24], the authors first extract sensor based features (the duration the activity and the number of sensors triggered) and then use SVMs, NB and neural networks to assess the activity quality and cognitive status of elderly people in smart homes. However, participants are provided with a brief description of each sub-task that they should refer to during the simulated activities such as planning a bus route, finding a recipe in the recipe book. These studies fail to provide an unobtrusive way of assessment since they are not done in the natural flow of daily living and in real life scenarios. Moreover, using rule-based systems, an expert is needed to manually integrate specific rules to the system since every person has own daily life routines. For example, waking up and drinking water in the middle of the night might be normal for a person, while abnormal for some other person. However, our approach does not require any expert knowledge, since it learns what is normal and abnormal from the training data automatically. Specifically, we aim in this study to detect anomalies in the natural flow of daily living without giving any instruction and considering not only some time interval, but everyday living scenario.

Some studies [25,26] focus on anomalies related to the duration and the timing of performed activities and other type of anomalies related to dementia such as repetition of activities are not taken into account. In [25], the authors introduce *activity curves* which models an individual's generalised daily activity routines based on automatically recognized activities. Deviations in behavioural routines are detected by comparing *activity curves* in order to do health assessment. In [26], the authors use a probabilistic model based on the location and outing interference of each activity. Then cross-entropy measure is used to detect anomalies such as staying in bed for a long time or not using the bedroom for sleeping during the night. In [14], the authors exploit HMM and fuzzy rules to detect duration, time and frequency related anomalies.

In the literature, there is some work dedicated to the synthesis of activity related data [12,14,15]. In [14], the authors modified real-world dataset in order to synthesise health related abnormal behaviours for their experiments. 8 daily activities such as sleeping, waking up, walking, eating are chosen and health related abnormal behaviours like frequent toilet visit, no exercise, slept without dinner are synthesised. In [15], more data is synthesised using HMMs based on a small set of real data collected. To increase the realism of data simulation, the sensor events were modelled by a combination of Markov chains and the Poisson distribution. However, in both [14,15], it is not mentioned in detail how the data synthesis was done. In [12], the authors modified a real-life dataset of an older adult converting basically the rooms into activities. The authors focused on walking and eating in conjunction with the sleeping activity and samples of these activities are manually inserted in the XML data set.

In [23], the authors exploit Hidden State Conditional Random Field (HCRF) method to detect abnormal activities that often occur in homes of elderly people by considering sub-activity relations. First, HCRF is used to recognise activities by producing a recognition confidence value for each activity. Then, a threshold based method is used to decide the activities as normal and abnormal. In [7], the authors detect anomalies of mild cognitive impairment by exploiting Markov Logic network. They use a hybrid technique including supervised learning, rule-based reasoning and probabilistic reasoning. However, they construct their model prior by defining each steps of each action. Those rules strongly depend on the specific home environment, on the used sensors, and on the particular habits of the elderly people; hence, their definition is time-expensive, and rules are not portable to different environments. In order to address this issue, the same authors propose a method to automatically learn the rule-based definitions of behavioural anomalies [18]. They exploit formal rule induction methods and a training set of normal and abnormal behaviours. However, the authors claim that their proposed rule learning method infers deterministic rules, which are prone to generate anomaly mispredictions in the presence of noise from the sensor infrastructure. In our study, normal daily life patterns are learnt for each individual from training data automatically and without the integration of any rules. Similar to [23,7], in our proposed work, anomaly is defined not activities alone but defined in the context of sequences, with other activities happened before and after.

Recently, there has been growing interest in CNNs [27,28,29,30,31,32,33,34,35,36], Deep Belief Networks (DBN) [37], Restricked Boltzman Machines (RBMs) [38,19,37,39] and Recurrent Neural Networks (RNNs) [29,30,40,41]. In [38], RBMs are used for feature extraction and selection from sequential data. In [39], results with RBM on CASAS dataset outperformed HMMs and NB in most of the cases. In [29], the authors use a combination of CNNs and Long Short Term (LSTM) RNNs to do multi-modal wearable activity recognition. In [30], the authors explore deep, convolutional and recurrent approaches on movement data captured with wearable sensors. Moreover, they describe how to train recurrent models in this setting and introduce a novel regularisation. In [34], the authors utilised convolutional networks to classify activities using time-series data collected from smart phone sensors. In [35], in a real world setting, an automatic stereotypical motor movement in Autism detection systems is developed exploiting CNNs. The discriminating features from multi-sensor accelerometer signals are learnt via CNNs and this knowledge is transferred to a new dataset. In [36], CNNs are exploited to learn features from raw physiological signals in an unsupervised manner analysis and then using multivariate Gaussian distribution, anomalies are detected to identify latent risks. In our previous work [41], RNNs are exploited to detect anomalies related to dementia in a daily living scenario.

CNNs have been exploited for activity recognition using movement datasets that are generated by wearable sensors [38,27,28,29,31,34,37]. Except the work by Fang et al. [39,41], none of these studies focus on daily activity datasets collected by sensors placed at home. Previous work on activity recognition based on wearable sensor datasets shows that CNNs and RNNs are useful to recognise activities, but leaves a lot of room for improvement. In this work, CNNs and their combination with LSTMs are investigated on daily activities datasets, namely Aruba [42] and WSU testbeds of CASAS smart home datasets [22] since the activities in these datasets are good examples to reflect daily life patterns of elderly person and to synthesise anomalies related to dementia.

## 3 Methodology

To assess CNNs in daily life activity recognition and abnormal behaviour detection tasks, the following steps are proposed: Firstly, a real-world dataset is modified in order to simulate abnormal behaviours related to dementia. Secondly, this dataset is segmented into time-slices by using a sliding window approach as described in [43]. Thirdly, sensor-based raw data is mapped into *last-fired* representations as described in [43]. Fourthly, CNNs are trained to recognise daily activities and encode daily-life behaviour routines. Lastly, the trained model is used to detect anomalies deviating from the normal daily-life sequences. In the following, the datasets are described as well as the methodology used to generate artificial dataset that reflects on the typical behaviour of a person with dementia.

### 3.1 Dataset

In this study, two datasets are used to evaluate activity recognition and abnormal behaviour detection. These datasets are namely Aruba [42] and WSU testbeds of CASAS smart home project [22].

In Aruba testbed, motion, door and temperature sensors are used. However, temperature sensors are excluded in this study and other 34 sensors (3 door and 31 motion sensors) are used. The data is provided as a list of (sensor, time-stamp) sensor measurements. In this dataset, there are 11 daily activities performed by a single user and it spans 224 days. These activities are *Meal*

*Preparation* (1606 instances), *Relax* (2910 instances), *Eating* (257 instances), *Work* (171 instances), *Sleeping* (401 instances), *Wash dishes* (65 instances), *Bed to toilet* (157 instances), *Enter home* (431 instances), *Leave home* (431 instances), *Housekeeping* (33 instances) and *Respirate* (6 instances). The activities performed in this dataset are totally normal and some of these normal activities will be modified for anomaly detection.

In WSU testbed [22], there are 5 activities, which are *Make a phone call*, *Hand washing*, *Meal Preparation*, *Eating*, *Cleaning*. There are 20 instances of each activity performed by 20 students in both *adlerror* and *adlnormal* versions. The *adlnormal* version consists of totally normal behaviours while in the *adlerror* version, there are specific errors in the task completion of these activities. Errors were selected to reflect common difficulties that can compromise everyday functional independence. The participants are told to include these errors during their performance. These errors can be seen in daily life activities and activity patterns of elderly people who are suffering from the consequences of cognitive decline.

### 3.2 Synthesis of Abnormal Activities Related to Dementia

This study aims to detect the following 3 different kinds of anomalies that can be seen in daily-life routines of elderly people with dementia: 1) Repeating activities, 2) Disruption in sleep, and 3) Confusion (getting confused during the activities).

**1) Repeating activities:** Elderly people suffering from dementia may forget whether they performed a particular daily activity or not, so they may repeat that activity. Frequency sensitive activities such as having a snack or drink, brushing teeth, taking medicine multiple times, etc. are the ones only the number of occurrences matters in terms of medical assessment. For instance, an elderly person suffering from Alzheimer may forget to have lunch, take multiple lunches instead [44], may forget to have dinner and start to prepare it in the middle of the night.

To reflect on this cognitive problem, we generate this kind of abnormal activities by manually inserting a specific set of actions within a random area of the normal activity sequence. This will result in multiple occurrences of that activity, which will occur in some inadequate time of the day such as having dinner in the middle of the night. We inject the instances of the following activities: *brushing teeth, preparing dinner, eating, getting snack* into the normal activity sequences to generate abnormal activities related to the frequency. For example; let's assume that $S$ is a sequence of activities occurring in a day such as $S = d_1, d_2, d_3, ..., d_x, b_1, b_2, ..., b_t, d_{x+1}, ..., d_n$ where each $d_i$ is a time-slice of some activities and each $b_j$ is a time-slice of *brushing teeth* activity. Here, there are $t$ time-slices of *brushing teeth* activity which consecutively results in only one instance of *brushing teeth* activity in the whole day sequence. Then, time-slice instances of *brushing teeth* activity are injected into the sequence $S$ to have the abnormal version. Then modified $S$ becomes

$S = d_1, d_2, d_3, ..., d_m, b_1, b_2, ..., b_t, d_{m+1}, ..., d_a, b_1, b_2, b_3, ..., b_k, d_{a+1}, ..., d_n$. As a result, we have two occurrences of *brushing teeth* activity in the sequence.

**2) Disruption in sleep:** Degeneration of the sleep-waking cycle and night time wandering are among the most severe behavioural symptoms of dementia. For example, elderly people may wake up many times in the night to use the toilet and go back to sleep or may forget to take daily amount of water [44, 45].

We simulate these anomalies by inserting some specific synthetic activities in the normal night-time activity sequences of a person. More specifically, we inject *Eating, Bed to Toilet* into a random area of *sleeping* activity in the normal daily activity sequence. This will emulate the activities of getting drink and going to the toilet frequently in the middle of the night. For example; given a sequence of *sleeping* activity such as $S = s_1, s_2, s_3, ..., s_n$ where each $s_i$ is a time-slice; time-slice instances of *getting drink* are injected into a random area of $S$. Then modified $S$ becomes $S = s_1, s_2, s_3, ...., s_m, d_1, d_2, d_3...d_k, s_{m+1}, ..., s_n$ where each time-slice $d_j$ is from *getting drink* activity. As a result, we simulate disruption in sleep anomaly where the person wakes up in the middle of the night and gets drink.

**3) Confusion:** Older adults suffering from cognitive decline tend to confuse things and perform some steps of activities more than once during the completion of activities. For example, they may fail to remember how to turn a CD player on, or may forget to turn off the television, air conditioning or house utilities such as kettle, oven, or they may leave the refrigerator door, the main door open. In order to test our methods on these kind of anomalies, the *adlerror* set of WSU dataset is used since confusion and forgetting anomalies are reflected in this dataset. For example, leaving the water running after washing hands, leaving the burner on after cooking the oatmeal, forgetting to take medication with the meal, wiping off the dishes without using running water to clean them are some examples to these kind of anomalies in the WSU *adlerror* set.

The first two types of anomalies are simulated by modifying Aruba testbed. Here, there is only one subject in the dataset. The lifestyle in the training data is taken as a norm and then we synthesise the abnormalities deviating from this norm and introduce these abnormalities in the test data. These activities are totally normal on their own but they become abnormal when they occur at a wrong time of the day and after or before a specific activity. Hence, capturing these abnormalities within the context is important. In all, 150 abnormal activity slices are generated manually. The third anomaly type is already reflected in WSU dataset; thus it is used directly without modifying any sensor reading.

### 3.3 Sensor Reading Representation

Firstly, time-slice chunks are extracted from raw sensor readings via a sliding window approach [43]. Data is discretised using the time-slice length of 60 seconds. A time-series chunk is a matrix of $t \times f$ size, where $t$ is the length of time-slices and $f$ is the number of sensor features. Then raw sensor readings are mapped into *last-fired* representation. Last-fired representation indicates which sensor is fired last. The sensor that changed state last continues to give 1 and changes to 0 when another sensor changes state. Previous work [41] shows that this representation gives better activity recognition accuracy rates than other representations as proposed in [43].

In the following, a description of CNNs used in this work is given.

## 3.4 Convolutional Neural Networks (CNNs)

CNN takes inputs of dimensions $h \times w \times d$, where $h$ is the height of the input matrix, $w$ is the width of the input matrix and $d$ is the number of different channels of the input matrix. In our study, $d$ is 1 since time-slice input matrices has only one channel.

A local filter (kernel) with a size of $n \times m \times q$ is used to extract fruitful feature patterns and capture local dependencies on the given input. Here, $n$ is the number height of the filter, $m$ is the width of the filter, while $q$ is the number of filters used. These values are given as a parameter during the network construction process. The weight of these filters are initialised randomly in the beginning and then CNN learns these weights on its own during the training process by optimising the values. In this study, random uniform initialisation is used to initialise the filters and Stochastic gradient descent is used to optimise the values during the training. An additional operation called activation function has been used after every convolution operation. In this study, Rectified Linear Unit (ReLU) is used as the activation function. Then a max-pooling layer, which is followed by a fully connected layer is added to the network. The fully connected layer used in our network is a traditional Multi Layer Perceptron that uses a softmax activation function in the output layer. The purpose of this layer is to use these features for classifying the input image into various classes based on the training dataset.

CNNs can contain one or more pairs of convolutional and max-pooling layers, where higher layers use broader filters to process more complex parts of the input. The top layers in CNNs are stacked by one or more fully connected normal neural networks. These fully connected neural network are expected to combine different local structures in lower layers for final classification purpose. In the training stage, CNN parameters are estimated by standard forward and backward propagation algorithms to minimise objective function.

### *3.4.1 Activity Recognition and Abnormal Behaviour Detection*

In order to recognise daily activities, training instances of the datasets and their corresponding labels are fed into CNNs to be trained. The models assign a class label to each instance with a confidence value. Firstly, the mean of confidence values of training instances for each class is calculated as follows.

$$m_j = 1/N \sum_{t=1}^{N} p_t \tag{1}$$

where $m_j$ is the mean confidence value of class $j$ and $p_t$ is the confidence value for training instance $t$ of that class and $N$ is the total number of instances in that class.

Then when a new test instance is introduced, if the model assigns it to a class with a confidence value which is bigger than the mean of that class ($m_j$), that instance is considered as a normal activity, otherwise it is flagged as an abnormal activity.

In order to test the affect of convolutions on different dimensions and different architectures, the following networks are tested on Aruba dataset (see Figure 1). Here, the input matrix is $N \times M$, where the rows are sensor readings for each time-slice and columns are the values of each sensor as time passes.

**1D Convolution:** In this model, convolution is done on temporal dimension. As depicted in Figure 1a, in the convolutional layer, 100 filters with a length of 10 is used. 1D convolution is followed by a max-pooling layer, which has a stride of 2. Then another convolutional layer (with 50 filters and a length of 5) and a max-pooling layer are added. After the extracted features are flattened, these features are fed into dense layers (3 hidden layers having 512, 128 and 50 units respectively) and then the final decision is given by a softmax layer producing the confidence values of assigned class labels.

**2D Convolution:** In this model, convolution is done on both of the dimensions, specifically on feature and temporal dimension. 100 filters with a size of $10 \times 34$ are used in the first convolutional layer which is followed by a $2 \times 2$ max-pooling. Then another 2D convolution operator is added this time with 20 filters with the size of $5 \times 34$. The flattened features are fed into the same dense layer and the softmax layer described above.

**CNN and LSTM (2D CNN + LSTM):** CNNs can learn spatial relationships on a given $N \times M$ input but they cannot relate a current input to the next one in the occurrence order of the input sequence. To overcome such limitation, LSTMs are used at the end of the CNN network. In this combination, firstly, the 2-layer 2D-CNN described above is used to learn the fruitful feature representation. And then the extracted feature maps are fed into LSTM layers which will be taking further temporal information of the slices into account. LSTM has hidden layers of size $30 \times 50$ respectively. LSTM layer is followed by a dense layer with 128 hidden units and then another dense layer with 50 units. Eventually, softmax layer classifies the input into one of the activity classes with a probability value.

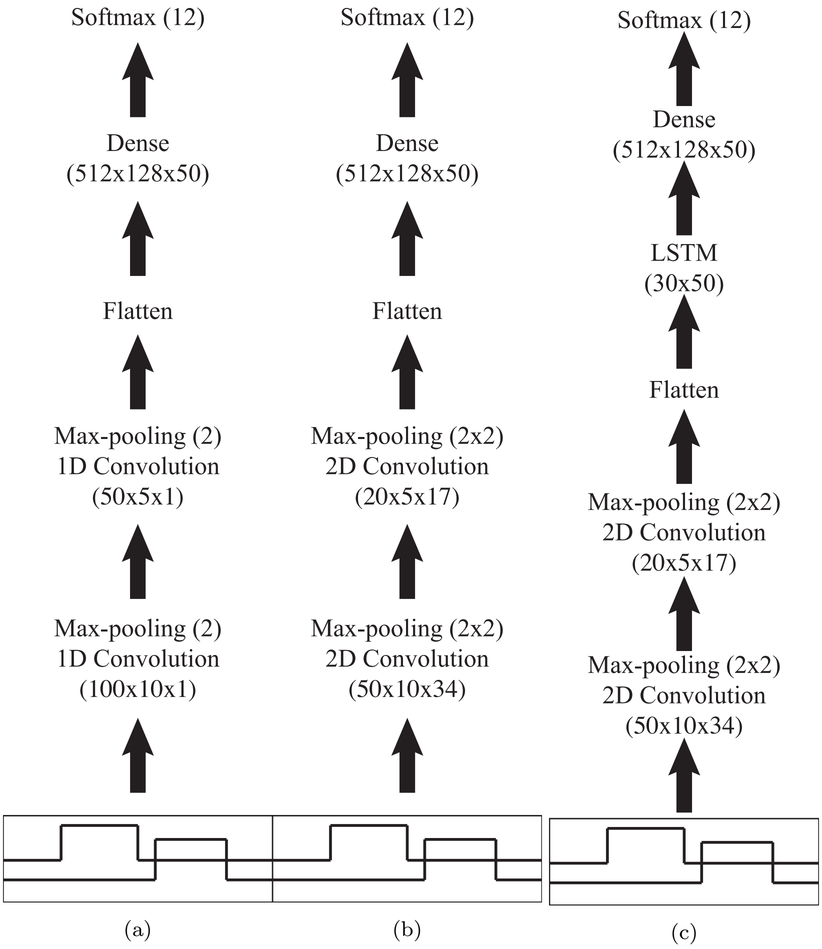



Fig. 1: Convolutional neural network architectures used. (a) 1D convolutional along temporal dimension (b) 2D convolutional both along temporal dimension and feature dimension (c) 2D convolution followed by an LSTM layer.

## 4 Experiments

In order to evaluate our methods, first the datasets are splitted (see Sec. 3.1) into train and test sets. However, the split is not done with a traditional split method since dividing daily activity datasets based on a fixed time period such as day is more meaningful [17]. Aruba testbed was collected in 224 days, thus 70 days are used as test, 15 days for validation and the remaining days are used for training. The first WSU set *adlnormal*, representing normal behaviours, are used to train the classifiers, while the second set, *adlerror*, containing the abnormal activity, is used for test set.

Keras Deep Learning library's [46] and Theano's [47] implementations of the CNNs and LSTM are used in this study. Moreover for the sake of comparison, results with NB, HMM, HSMM and CRF are provided which are based on the implementation provided in [43]. In the CNN experiments below, Adam optimiser [48] is used and the instances are fed into the system with a batch size of 20. In the following, the evaluation metrics are explained further.

## 4.1 Evaluation Metrics

In order to assess the activity recognition success, the following metrics are used: Precision, Recall, F-measure and Accuracy. As seen in Formula 2 and 3, final precision and recall values are calculated by taking average over classes. Precision and recall measures are used in order to show how well the models perform on imbalanced datasets like the one in this study. On the other hand, the accuracy represents the percentage of correctly classified time slices, therefore more frequently occurring classes have a larger weight in this measure. Here, $TP$ is true positive, $TT$ is total number of instances, $TP$ is total true labels, $TI$ is total of inferred labels, $N$ is the number of classes in a specific class of the dataset and $Total$ is the total number of instances of all classes in the dataset

$$\text{Precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TI_i} \tag{2}$$

$$\text{Recall} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TT_i} \tag{3}$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} TP_i}{\text{Total}} \tag{5}$$

Abnormal behaviour detection success rate is evaluated by sensitivity and specificity metrics. Sensitivity or True Positive Rate (TPR) refers to the method's ability to correctly detect instances which are abnormal. Specificity or True Negative Rate (TNR) gives the percentage of correctly recognized normal instances, thus reflects the method's ability to differentiate between normal and abnormal.

$$\text{Sensitivity (TPR)} = TP/(TP + FN) \tag{6}$$

$$\text{Specificity (TNR)} = TN/(TN + FP) \tag{7}$$

### 4.2 Results

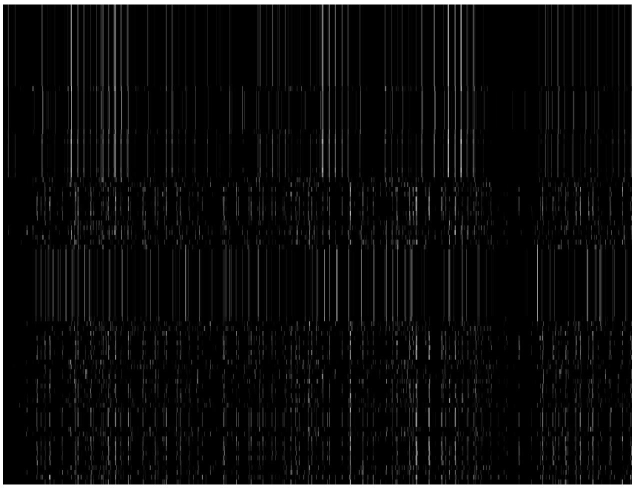
Two types of experiments are performed: 1) Activity recognition, and 2) Abnormal activity detection. Activity recognition success rates by both generative and discriminative methods on Aruba set are depicted in Table 1. The results indicate that CNNs with 2D convolution (accuracy of 89.67%) and also CNN-2D followed by an LSTM classifier (accuracy of 89.72%) outperforms CRF (accuracy of 88.58%). The reason is CNNs extract their own fruitful features while CRF only relies on the given input. HMM and HSMM give the worst accuracy results (77.90% and 77.98% respectively). NB gives better accuracy result (84.37%) than HMM and HSMM but it results in lower precision (42.87%) and recall (61.04%) rates. Although HMM and HSMM give the best recall rates (72.03% and 71.56% ), they fail in giving good precision rates (43.66% and 43.97% respectively). It is seen that CNN-1D network has an accuracy of 87.50% while it fails in high precision (31.42%) and recall (36.78%) values. CNN-1D extracts features on temporal dimension, so it takes temporal information within a time-slice chunk into account but on the other hand, it ignores the relationship between sensors since it doesn't do convolution on the feature dimension. Thus, it doesn't learn class specific feature maps to differentiate between different classes resulting in low precision and recall. When 2D convolution is used, both temporal and spatial information are taken into account and the networks learn more informative features. Thus, it gains the ability to learn class specific features, which results in higher precision and recall values (46.84% and 41.68%) and high accuracy results (89.67%). CNNs cannot remember the previous and the next inputs, but feeding the feature maps into an LSTM layer helps us process the temporal dimension further. In result, CNN with 2D convolution followed by LSTM (CNN-2D + LSTM) achieves a precision rate of 51.20% and a recall rate of 50.55% and an accuracy of 89.72%.

Table 1: Activity recognition results with *last-fired* representation on Aruba dataset

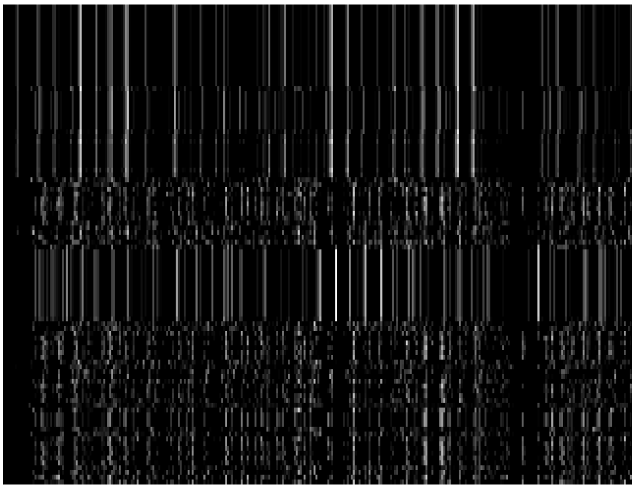
| **Model** | **Precision** | **Recall** | **F-measure** | **Accuracy** |
|---|---|---|---|---|
| NB | 42.87% | 61.04% | 50.36% | 84.37% |
| HMM | 43.66% | 72.03% | 54.36% | 77.90% |
| HSMM | 43.97% | 71.56% | 54.47% | 77.98% |
| CRF | 50.24% | 52.83% | 51.50% | 88.58% |
| LSTM | 38.65% | 41.29% | 39.92% | 89.00% |
| CNN-1D | 31.42% | 36.78% | 33.89% | 87.50% |
| CNN-2D | 46.84% | 41.68% | 44.11% | 89.67% |
| CNN-2D + LSTM | 51.20% | 50.55% | 50.87% | 89.72% |

In Figure 2, extracted feature maps from first and second layer and the flatten layer are visualised for CNN-2D network as described in Figure 1. It is seen that noise is reduced and more informative features are learnt as the layer level increases. The x-axis represents features while y-axis is time axis

(a)



(b)



(c)

Fig. 2: a) Extracted features from the first layer. b) Extracted features from the second layer. c) Extracted features from the flatten layer. The x-axis shows time while y-axis represents features. Successive model layers learn deeper intermediate representations. The features get more discriminative and visible in the last layer.

and the white pixels are activations of neurons. It is seen that as the times passes, different activities give different features.

Moreover, we calculate Cohen's Kappa statistics in order to show the robustness of the proposed method, CNN-2D classifier. Kappa statistics is measure that can handle well on both multi-class and imbalanced class problems. It tells how much better the classifier is performing over the performance of a classifier that simply guesses at random according to the frequency of each class. It is thought to be a more robust measure than simple percent agreement calculation, since Kappa takes into account the possibility of the agreement occurring by chance [49]. The calculated Kappa statistics for CNN-2D classifier is 0.64431, which is a substantial agreement according to [49].

Results on Aruba dataset show that classifiers are mostly successful in detecting the instances of *leave home* and *enter home* activities since they are the only activities involving door sensors, thus they are not confused with any other activities. Moreover, *meal preparation* activity is confused with *wash dishes* activity most of the time since they involve same kind of sensors and they both take place in the kitchen. Also *house keeping* activity is generally confused with *work* activity since they may take place in the same room and may involve same sensors.

Table 2: Abnormal behaviour detection results on Aruba Modified dataset

| | **Aruba Modified** | | **WSU** | |
|---|---|---|---|---|
| **Model** | **Sensitivity** | **Specificity** | **Sensitivity** | **Specificity** |
| NB | 99.33% | 33.89% | 46.17% | 98.42% |
| HMM | 45.54% | 27.71% | 100% | 50.55% |
| HSMM | 100% | 35.61% | 100% | 42.89% |
| CRF | 100% | 66.03% | 47.87% | 72.17% |
| LSTM | 98.67% | 75.48% | 86.50% | 77.89% |
| CNN -2D | 85.33% | 33.89% | 88.70% | 67.46% |

The second experiment, abnormal activity detection is performed firstly on modified Aruba set. As a representative of CNN networks, the results are presented with the CNN-2D network. After training the models with normal behaviours, test set which includes the abnormal behaviours is introduced to the classifier and activity instances which are assigned a label with low confidence values are flagged as abnormal. In Table 2, it is seen that the highest specificity is achieved by LSTM networks giving an accuracy of 75.48% (and sensitivity rate of 98.67%). Although NB, HMM and HSMM models gives higher sensitivity rates (99.33%, 100%, 100% respectively), the specificity rates are smaller (33.89%, 27.71% and 35.61% respectively). HMM gives the worst results (a sensitivity rate of 45.54% and a specificity rate of 27.71%). CNN-2D gives a sensitivity rate of 85.33% and a specificity of 33.89%. This shows that LSTMs are more suitable to detect repetition and order related abnormal activities since it can relate current input with the upcoming ones what CNN cannot do.

The second part of anomaly detection experiments are performed on WSU testbed. 30 second time-slice chunks are extracted from sensor readings from WSU. This dataset is not collected in a daily life scenario, thus sensor readings are not in a sequential order. Thus the sensor readings are available only for activities labelled in the dataset. The *adlnormal* set is used as training set and the *adlerror* set is used as test dataset. The aim here is to see how successful the classifiers are to detect the anomalies, given normal behaviours. The results in Table 2 indicate that the highest sensitivity rate is given by HMM and HSMM (both 100%), while HMM gives a specificity rate of 50.55% and HSMM achieves specificity of 42.89%. The highest sensitivity rate is achieved by CNN-2D classifier (86.70%), but LSTM gives a very close sensitivity rate (86.50%) and a higher specificity rate (77.89%) where CNN-2D achieves a specificity rate of 67.47% only.

As a comparison, in [22], the authors present their results as follows. The number of correctly detected activities are 95 for *adlnormal* and 76 for *adlerror*, both out of 100. Experiments in our study are performed on activity slices, on the other hand. in [22] they take whole activity and extract features from that activity and then try to decide if it is normal or abnormal. The problem here that is in real-life scenario, it cannot be known, where an activity starts and ends. Thus using slice-based detection is more meaningful.

LSTM is better to capture repetition related activities, while CNN is better to detect “confusion related activities”. CNN can detect changes in feature patterns. Even though it is not explicitly defined in daily activity datasets each activity is formed by steps. The steps in this dataset are based on the motion sensors triggered. For example, when the *sleeping* activity is considered, it is seen that the resident first goes out of bed, then goes to the middle of the room and then goes to the bathroom in the *bed to toilet activity*. CNN doesn’t need to extract them, but it exploits them hierarchically in each layer. In the end, model cannot identify the steps involved, but it detects the anomaly in the higher level. Thus, whenever the orders of sensors or these steps change, input matrix changes which leads different feature maps extracted by CNNs.

Our method cannot detect anomalies such as incorrectly measuring the oatmeal, not using soap when cleaning, washing hands multiple times, confusing the location of items, and using too much soap or leaving kitchen utilities on. Because there were no specialised sensors for the items involved in these steps, the algorithm could not detect these errors and future research will be needed to deal with these errors. Moreover, our current approach may fail to detect abnormalities, when there is gradual deterioration regarding the health of an elderly person. This issue will be taken into consideration in future while collecting real-world data in which gradual deterioration can be observed. Moreover, it is planned to extract sub-activities involved in daily life activities and model their relations hierarchically. Then, this information can be used in order to provide more robust and accurate cognitive assessment tools.

## 5 Conclusion

This paper introduces a method of recognising sensor based activities and detecting anomalies related to dementia in smart homes. CNNs are exploited as well as their combination with LSTM in order to achieve these tasks. Our results on activity recognition shows that these methods are better than their competitors such as NB, HMMs, HSMMs and CRFs. Moreover, results on anomaly detection gives promising results to detect most of the abnormal behaviours simulating the daily life behaviour of elderly people suffering from dementia.

*Acknowledgments*

We thank to Tim Van Kasteren for providing the code for NB, HMM, HSMM and CRF. We also thank to Diana Cook and her team for making Aruba and WSU datasets publicly available.

## References

1. S. Wild, G. Roglic, A. Green, R. Sicree, and H. King. Global prevalence of diabetes estimates for the year 2000 and projections for 2030. pages 1047–1053, 2004.
2. M. S. Albert and et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging. *Alzheimer's & Dementia*, Volume 7:270–279, 2011.
3. W. Thies and L. Bleiler. 2013 Alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association vol. 9, no. 2*, pages 208–245, 2013.
4. M. R. Hodges, K. Kirsch, M. W. Newman, and M. E. Pollack. Automatic assessment of cognitive impairment through electronic observation of object usage. In *Pervasive*, pages 192–209, 2010.
5. K. Wild. Aging changes. In *Geraotechnology, Vol. 9 No 2*, pages 121–125, 2010.
6. P. Dawadi, D. Cook, and M. Schmitter-Edgecombe. Smart home-based longitudinal functional assessment. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1217–1224, 2014.
7. D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui. Fine-grained recognition of abnormal behaviours for early detection of mild cognitive impairment. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 149–154, 2015.
8. D. Lara and M. A. Labrador. A mobile platform for real-time human activity recognition. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, pages 667–671, 2012.
9. T. Kirste, A. Hoffmeyer, P. Koldrack, A. Bauer, S. Schubert, S. Schrder, and S. Teipel. Detecting the effect of Alzheimer's disease on everyday motion behaviour. *Journal of Alzheimer's Disease*, pages 121–132, 2014.
10. M. Ermes, J. Parkka, and L. Cluitmans. Advancing from offline to online activity recognition with wearable sensors. *$30^{th}$Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4451–4454, 2008.
11. T. Zhang, W. Fu, J. Ye, and M. Fischer. Learning movement patterns of the occupant in smart home environments: an unsupervised learning approach. *Journal of Ambient Intelligence and Humanized Computing*, 8(1):133–146, 2017.
12. Gilles Virone. Assessing everyday life behavioural rhythms for the older generation. *Pervasive Mob. Comput.*, 5(5):606–622, 2009.

13. K. S. Park J. H. Shin, B. Lee. Detection of abnormal living patterns for elderly living alone using support vector data description. *IEEE Transactions on Information Technology in Biomedicine 15(3):*, page 438448, 2011.
14. Abdur Rahim Mohammad Forkan, Ibrahim Khalil, Zahir Tari, Sebti Foufou, and Abdelaziz Bouras. A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. *Pattern Recognition*, 48(3):628 – 641, 2015.
15. Jens Lundstrm, Eric Jrpe, and Antanas Verikas. Detecting and exploring deviating behaviour of smart home residents. *Expert Systems with Applications*, 55:429–440, 2016.
16. D. J. Cook A. Crandall Adriana M. Seelye, M. Schmitter-Edgecombe. Naturalistic assessment of everyday activities and prompting technologies in mild cognitive impairment. *Journal International Neuropsychologly Soc. 2013 (4):*, pages 442–52, 2013.
17. N.K. Suryadevara, S.C. Mukhopadhyay, R. Wang, and R.K. Rayudu. Forecasting the behavior of an elderly using wireless sensors data in a smart home. *Engineering Applications of Artificial Intelligence*, 26(10):2641–2652, 2013.
18. Zaffar Haider Janjua, Daniele Riboni, and Claudio Bettini. Towards automatic induction of abnormal behavioral patterns for recognizing mild cognitive impairment. In *Proceedings of the* $31^{st}$ *Annual ACM Symposium on Applied Computing*, pages 143–148, 2016.
19. N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plotz. PD disease state assessment in naturalistic environments using deep learning. pages 1742–1748, 2015.
20. K. S. Gayathri, Susan Elias, and Balaraman Ravindran. Hierarchical activity recognition for dementia care using markov logic network. *Personal and Ubiquitous Computing*, 19(2):271–285, 2015.
21. Fco Javier Ordóñez, Paula de Toledo, and Araceli Sanchis. Sensor-based bayesian detection of anomalous living patterns in a home setting. *Personal and Ubiquitous Computing*, 19(2):259–270, 2015.
22. Diana Cook and M. Schmitter-Edgecombe. Assessing the quality of activities in a smart environment. *Methods Inf. Med.*, 48:480–485, 2009.
23. Y. Tong, R. Chen, and J. Gao. Hidden state conditional random field for abnormal activity recognition in smart homes. *Entropy*, 17(3):1358, 2015.
24. P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe. Automated cognitive health assessment using smart home monitoring of complex tasks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(6):1302–1313, 2013.
25. Prafulla N. Dawadi, Diane J. Cook, and Maureen Schmitter-Edgecombe. Modeling patterns of activities using activity curves. *Pervasive Mob. Comput.*, 28(C):51–68, June 2016.
26. Oya Aran, Dairazalia Sanchez-Cortes, Minh-Tri Do, and Daniel Gatica-Perez. Anomaly detection in elderly daily behaviour in ambient sensing environments. *Human Behaviour Understanding:* $7^{th}$ *International Workshop*, pages 51–67, 2016.
27. J. Yang, M. Nguyen, P. San, X. Li Li, and S. Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. pages 3995–4001, 2015.
28. M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In $6^{th}$ *International Conference on Mobile Computing, Applications and Services*, pages 197–205, 2014.
29. F. J. Ordonez and D. Roggen. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.
30. Nils Y. Hammerla, Shane Halloran, and Thomas Plötz. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 1533–1540, 2016.
31. S. Ha and S. Choi. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 381–388, 2016.

32. Charissa Ann Ronao and Sung-Bae Cho. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, 59:235 – 244, 2016.
33. Rui Yao, Guosheng Lin, Qinfeng Shi, and Damith Chinthana Ranasinghe. Efficient dense labeling of human activity sequences from wearables using fully convolutional networks. *CoRR*, abs/1702.06212, 2017.
34. A. Charissa and C. Sung-Bae. Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors. *In Proceedings of the KIISE Korea Computer Congress*, pages 858–860, 2015.
35. Nastaran Mohammadian Rad, Andrea Bizzego, Seyed Mostafa Kia, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello. Convolutional neural network for stereotypical motor movement detection in autism. *CoRR*, abs/1511.01865, 2015.
36. Kai Wang, Youjin Zhao, Qingyu Xiong, and et al. Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals. *Scientific Programming*, 2016.
37. S. Choi, E. Kim, and S. Oh. Human behavior prediction for smart homes using deep learning. In *2013 IEEE RO-MAN*, pages 173–179, 2013.
38. T. Plötz, N. Hammerla, and P. Olivier. Feature learning for activity recognition in ubiquitous computing. volume 2, pages 1729–1734, 2011.
39. H. Fang and C. Hu. Recognizing human activity in smart home using deep learning algorithm. In *$33^{rd}$ Chinese Control Conference*, pages 4716–4720, 2014.
40. H. Fang, H. Si, and L. Chen. Recurrent neural network for human activity recognition in smart home. *In Proceedings of 2013 Chinese Intelligent Automation Conference*, pages 341–348, 2013.
41. Damla Arifoglu and Abdelhamid Bouchachia. Activity recognition and abnormal behaviour detection with recurrent neural networks. In $14^{th}$ *International Conference on Mobile Systems and Pervasive Computing*, pages 86–93, 2017.
42. D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan. CASAS: A smart home in a box. *Computer*, 46(7):62–69, 2013.
43. T. Van Kasteren, G. Englebienne, and B. J. A. Kröse. Human activity recognition from wireless sensor network data: Benchmark and software. *Activity Recognition in Pervasive Intelligent Environments*, pages 165–186, 2011.
44. J. Saives, C. Pianon, and G. Faraut. Activity discovery and detection of behavioral deviations of an inhabitant from binary sensors. *IEEE Transactions on Automation Science and Engineering*, 12(4):1211–1224, 2015.
45. M. Amiribesheli and A. Bouchachia. Smart homes design for people with dementia. In *2015 International Conference on Intelligent Environments*, pages 156–159, 2015.
46. C. François. Keras. https://github.com/fchollet/keras, 2015.
47. Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, may 2016.
48. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
49. G.G. Landis, J.R.; Koch. The measurement of observer agreement for categorical data. *Biometrics 33 (1)*, pages 159–174, 1977.