Abnormal Behaviour Detection for Dementia Sufferers via Transfer Learning and Recursive Auto-Encoders

Damla Arifoglu and Abdelhamid Bouchachia
Bournemouth University
Fern Barrow, Poole BH12 5BB
Dorset, UK
{darifoglu,abouchachia}@bournemouth.ac.uk

Abstract—Cognitive impairment is one of the crucial problems elderly people face. Tracking their daily life activities and detecting early indicators of cognitive decline would be necessary for further diagnosis. Depending on the decline magnitude, monitoring may need to be done over long periods of time to detect abnormal behaviour. In the absence of training data, it would be helpful to learn the normal behaviour and daily life patterns of a (cognitively) healthy person and use them as a basis for tracking other patients. In this paper, we propose to investigate Recursive Auto-Encoders (RAE)-based transfer learning to cope with the problem of scarcity of data in the context of abnormal behaviour detection. We present a method for generating synthetic data to reflect on some behavior of people with dementia. An RAE model is trained on data of a healthy person in a source household. Then, the resulting RAE is used to detect abnormal behavior in a target house. To evaluate the proposed approach, we compare the results with the-state-of-the-art supervised methods. The results indicate that transfer learning is promising when there is lack of training data.

Index Terms—Dementia, Transfer Learning, Recursive Auto-encoders, Abnormal Behaviour Detection

I. 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 cognitive decline, which is a condition that causes problems with physical and mental abilities such as memory and thinking [2]. An elderly person having such cognitive decline requires care and support from caregivers. A continuous monitoring of the daily routine of the elderly can be helpful for clinicians to diagnose the early onset of cognitive decline. 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 trends over time [3]. Most common types of dementia (Alzheimer) can be identified by behavioural changes like sleep disturbances and inability to complete tasks. 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. In particular, the daily home activity involving basic functions like preparing food, showering, sleeping, etc. can be used to assess the well-being of elderly people. Thus, it is beneficial to track elderly people's life over time in a smart home to detect the indicators of dementia at an early stage.

The development of ambient home assessment environments has begun to provide the opportunity to assess behaviour change unobtrusively in real-time. Although a few promising methods have been experimentally validated [3]–[5], 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 that may poorly represent a person's cognitive status. 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 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.

Machine learning based cognitive status assessment studies rely on activity recognition techniques. These methods first learn what is normal from training data and then flag the abnormal activity based on classification confidence values [6]–[10]. They require training data to be manually annotated, which is extremely hard and a time-consuming task. Moreover, these techniques assume that the training data is available prior to the training phase. However, we cannot expect elderly people to annotate the necessary training data. Thus, tackling the activity recognition and abnormal behaviour detection as an unsupervised process would be helpful. Moreover, this monitoring may need to be done over long periods of time. But collecting sequential data of months or years with time dependency is highly time-consuming. Thus, using an existing data from a source household to learn what is normal, and then transferring this knowledge to a target house would be beneficial.

Daily life activities are often composed of several steps [11]. For example; the activity *wash clothes* implies the following actions: *get clothes from basket, fill up washing machine, turn on washing machine*. The anomalies related to dementia may be reflected in the repetition frequency of these steps

and their relation with each other. The elderly people with dementia tend to confuse things and repeat or skip some steps during the completion of a specific activity. Building activities from their granular units hierarchically would be helpful to understand the internal dynamics of the activities. Hence, the problem of activity recognition can be viewed as a hierarchical learning problem which resembles scene parsing or phrase detection [12]. In this paper, we aim to construct activity instances hierarchically from their low-level units to detect abnormal ones reflecting on indicators of dementia. The hierarchical representation of RAE provides an abstraction of activities in a house and then mapping these abstract levels to another house via transfer learning will be useful. Abstraction provides a generalisation of the hierarchical level of information between the houses and it reduces the differences between them, making transfer learning an appealing approach. In this paper, we use unlabelled data collected for normal activities from a source house to train the RAE model. Then, we transfer this model to a target house to detect abnormal behavior related to dementia. In a nutshell, the present paper introduces two contributions. Firstly, a method is proposed to generate synthetic data that simulates the abnormal behavior of people with dementia. Secondly, RAEs are exploited to model activities based on their low-level structures and detect abnormal behavior related to dementia. The rest of the paper is organised as follows. Section II provides an overview of the related literature. Section III presents the details of the proposed methodology together. Section IV describes the experimental set-up and results of the experiments followed by a discussion. Finally, Section V concludes the paper.

II. RELATED WORK

In-home automatic assessment of cognitive decline has been the subject of many machine learning approaches such as Support Vector Machines (SVMs) and Naïve Bayes (NB) [6], Restricted Boltzmann Machines (RBMs) [7], Hidden Markov Models (HMMs) [8], Random Forests [9] and Recurrent Neural Networks (RNNs) [10]. In [10], the authors exploit RNNs to detect abnormal behavior of dementia sufferers in a daily living scenario. The abnormal behavior are flagged based on their classification confidence values. Unsupervised methods such as auto-encoders are also being used for anomaly detection in time-series literature [13], [14]. Many studies [4], [7] have relied on rule-based systems to assess the cognitive status of elderly people. In [15], the assessment is done by asking elderly people to complete a sequence of scripted actions. The participants are monitored via Web camera while they perform tasks and they receive scores by trained experts. In [4], the authors detect anomalies by exploiting Markov Logic network. They use a hybrid technique including supervised learning, rule-based reasoning and probabilistic reasoning. 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. Specifically, we aim in this study to detect anomalies without giving any instruction and considering not only some time interval, but everyday living scenario. Moreover, in

rule-based systems, an expert is needed to manually integrate specific rules to the system. The proposed approach does not require any expert knowledge.

In transfer learning literature, most of the activity recognition models are supervised models that require labelled data to learn the model parameters [16]–[18]. Good results are obtained using generative models such as HMM [16], [18] and discriminative models such as CRF [17], [18]. In [19], a method is proposed to learn the parameters of a HMM using labelled data from the source domain, and unlabelled data from the target domain. The study ignores the activities' important features such as the activity structure and related temporal features. They also assume that the structure of HMMs is given and pre-defined. Later they extend this work to learn hyperparameter priors for HMM instead of learning the parameters directly [20].

In transfer learning, sensors and activities in different households are needed to be mapped. In [19], a comparison of feature mappings was done. The mapping that combined sensor readings in a single feature based on their function (e.g. sensors used during cooking) gave the best results. In some cases, meta-features are first manually introduced into the feature space and then the feature space is automatically mapped from the source domain to the target domain [20]. In [21], the authors first assign a location label to each sensor indicating in which room or functional area the sensor is located. Then activity templates are constructed from the data for both the source and target data, finally a mapping is learnt between the source and target datasets based upon the similarity of activities and sensors.

III. PROPOSED WORK

The proposed work consists of the following steps: 1) Time-slice chunks are extracted from sequential sensor reading data using a sliding window. 2) *Last-fired* features are extracted from time-slices as in [22]. 3) RAE is trained on a source household dataset to learn the parameters for *normal* behavior. 4) These parameters are then transferred into a target household to detect abnormal behavior.

A. Dataset

The evaluation of the proposed method is done on households A and C of Kasteren datasets [20]. We chose these two households since they span more days (25 days and 18 days respectively). The activities performed in household A and C are used to reflect normal behavior. However, some of the data in household C is modified (Section III-B) to generate samples representing abnormal behaviour of dementia sufferers. Here, household A will be used as the source house while household C will be used as the target house.

B. Synthesis of Abnormal Behaviour

Given the scarcity of data reflecting abnormal behavior of dementia sufferers, we need to synthesise some activities that can be observed in daily-life routines of elderly people

with dementia. We focus on the generation of two kinds of anomalies: 1) Repeating activities and 2) Disruption in sleep

1) Repeating activities: Elderly people with dementia may forget whether they performed a particular activity or not, so they may repeat that activity (having multiple lunches, e.g.) [23]. To reflect on this scenario, we generate synthetic abnormal activities by manually inserting a specific set of actions within 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.

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 [23], [24]. We simulate these anomalies by inserting some synthetic activities in the normal night-time activity sequences.

For example; assume that S is a sequence of activities occurring in a day such as $S = d_1, \ldots, d_x, e_1, \ldots, e_t, d_{x+1}, \ldots, d_n$ where each d_i is a time-slice of some activity and each e_j is a time-slice of *eating* activity. Then, time-slices of e are injected into the sequence S to have the abnormal version. Then S becomes $d_1, \ldots, d_m, e_1, \ldots, e_t, d_{m+1}, \ldots, d_a, e_1, \ldots, e_k, d_{a+1}, \ldots, d_n$. Many instances of *getting drink*, *taking shower*, *use toilet* activities are injected. In result, 162 abnormal time-slices are synthesised in dataset C.

C. Feature Extraction and Sensor Mapping

After the synthesis, the datasets are processed in the following way. Firstly, 1 minute slices are extracted from datasets using a sliding window [22]. Then time-slices are mapped into *last-fired* feature representation [20]. *Last-fired* feature [20] 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. The last representation gives an indication of the location of an inhabitant. As people start moving, the sensors are triggered based on the location of the movement, which provides an update of their current location [22]. The updates, in the form of a time-series data, provide fine-grained patterns about the activity performed. Such patterns are hierarchical and they follow grouping rules at multiple levels of abstraction. Findings in [25] support our approach. The authors extract location-based patterns in daily-life routines. Hence, we employ RAE as a hierarchical model to organise the steps in an activity and record their relative ordering. We exploit *last-fired* feature to model location based granular level information, since such feature allows for capturing execution details of the activities. Next, RAEs will be exploited to model all this low-level information in a hierarchical representation.

There are 14 sensors in dataset A and 21 sensors in dataset C. We map these sensors to each other by using meta features as described in [20]. We use the mapping that combined sensor readings in a single feature based on their function (e.g. sensors used during cooking). In [19], different mapping strategies, such as union, intersection and duplicate are investigated. We use *union* mapping since it gave the best results. Using *union* mapping for each function group, the union of all the sensors in the group is taken, resulting in one sensor output per group per house. For example, the front and back door in the target house are combined into a single sensor and matched with the front door sensor in the source house. This results in 7 sensor groups, which will be treated as features. Moreover, the activities in two datasets are mapped and 9 similar activities are used [20].

D. Recursive Auto-Encoders

Auto-encoders are unsupervised artificial neural networks that compress the input into a representation and then reconstructs the output from it. They are self-supervised because they use the input instances as labels and use training data to learn the parameters for the model. An auto-encoder has 3 parts: an encoding function, a decoding function, and a loss function. The encoder compresses the input and produces a representation, the decoder then reconstructs the input from this representation. Loss function calculates the error between the actual input and the reconstructed input.
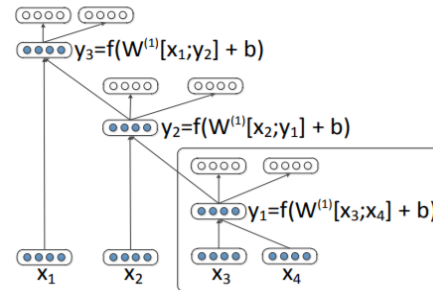

Fig. 1. A recursive auto-encoder

As depicted in Figure 1 (retrieved from [26]), RAEs merge each instance with its next neighbour to construct the parent node [27]. For example, the children x_3 and x_4 are merged and the parent y_1 is constructed by an encoding function f. Then the parent y_1 is merged with another child x_2 and this goes on in the upper layers in a bottom-up fashion. The process yields a hierarchical organisational structure for children. The final code representing the entire tree is decoded to recover the children and the entire hierarchy by the following inverse process. The first parent vector y_1 is computed from the children x_3, x_4, so that $y_1 = f(W^{(1)}[x_3; x_4] + b^{(1)})$ where a matrix of parameters weights W is multiplied with the children vectors. After adding a bias term, an element wise activation function such as $tanh$ is applied to the resulting vector. To see how well this function is doing, the model reconstructs the children in a reconstruction layer: $[x_3; x_4] = g(W^{(2)}y_1 + b^{(2)})$. At each level of the tree, the same encoding and decoding function is used recursively. During training, the goal is to minimise the reconstruction errors of the input pairs. For each pair, the Euclidean distance between the original input and its

reconstruction is calculated: $E = P_{rec}([x_3; x_4])$. The process repeats until the full tree is constructed and a reconstruction error is obtained at each non-terminal node. The encoding and decoding weight parameters are learnt by using the training set and applying back-propagation through structure [28] to update the network weights.

E. Abnormal Behaviour Detection

First, house A is used to learn the parameters $(W^{(1)}, b^{(1)})$ for encoding function and $(W^{(2)}, b^{(2)})$ for decoding function of RAE. These parameters are then used to construct RAE trees to test instances of house C. In each level of the tree, two children are merged to form a feature vector as their parent, which encodes the information coming from the children. Thus, the feature vector at the root node summarises all the information coming from the children in the tree and their hierarchical orderings are learnt by RAE. The feature vectors at parent nodes can be decomposed into their granular level, hierarchical pieces by using the decoding weights. Anomalies are defined as samples that are deviations from the expected behaviour. Any deviations from this normal can be identified by measuring the reconstruction error E described above. When a new activity is introduced as a test instance, if it is a normal activity, the reconstruction error will be smaller since anomalies that represent any deviations will be poorly reconstructed. If it is an abnormal behaviour, which is not seen in the source household, RAE will reconstruct that instance with a higher error. We exploit these errors to decide if that activity is normal or abnormal based on a threshold.

The proposed method is compared with the following state-of-the-art supervised methods; Long Short Term (LSTM) RNNs, NB, HMM and CRF. First, using the training instances and their corresponding labels in dataset A, these models are trained. Then the instances in dataset C are given to the trained classifiers. The models assign a class label to each instance with a confidence value. When a new test instance in house C is introduced, if the model assigns it to a class with a confidence value which is bigger than a threshold, that instance is considered as a normal activity, otherwise it is flagged as an abnormal activity.

IV. EXPERIMENTS

The experiments with NB, HMM and CRF are performed based on the implementation [19], while LSTM and RAE experiments are performed on Python and Keras. In RAE trees, each child represents 1 minute time-slice of a feature vector (size of 1×7). RAE trees are constructed with a time-step of 5 (chosen experimentally), where 5 time-slices are merged in a RAE tree. To run experiments on LSTM, we used drop-out with a value of 0.5. We also set the batch size to 10 and the epoch to 500 iterations. The internal architecture of LSTM (2 hidden layers consisting of 30 and 50 nodes respectively) and time-step of the sequences (25 activity slices) were empirically set. In supervised models, the activities in house C are evaluated based on the model learnt from the house A.

To assess the abnormal behaviour detection success, True Positive Rate (TPR) and False Positive Rate (FPR) are used. These values for different thresholds are showed on a Receiver Operating Characteristic (ROC) curve. Moreover, Area Under Curve (AUC) is calculated for each model to interpret the results in a better way. TPR refers to the method's ability to correctly detect instances which are abnormal. FPR gives the percentage of mislabelled normal instances, thus reflects the method's ability to differentiate between normal and abnormal.

The performance of the supervised methods to classify the activity instances is measured by precision, recall, F-measure and accuracy. Final precision and recall values are calculated by taking average over classes. Precision and recall give better idea about the performance on imbalanced datasets like the ones 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.

A. Results

The first experiment is conducted to evaluate the classification performance of the supervised methods when using transfer learning. These methods are trained on house A and then tested on house C. Activity recognition accuracy rates are depicted in Table I. These results are very close to activity recognition rates with leave-one-out cross validation presented on the same datasets in [22], where one day of the dataset C is used as testing set, while the remaining days are used training set. However, our results are obtained via transfer learning, where the training is done on dataset A and the testing is done dataset C. In [22], the leave-one-out classification accuracy with NB, HMM and CRF are given as 87.0%, 83.9% and 89.7% as respectively. In our case, the classification accuracy rates are 87.47%, 84.88%, 84.55% and 87.02% with NB, HMM, CRF and LSTM respectively. The similar results show that applying transfer learning is successful to recognise activities. Unfortunately, we cannot test the classification accuracy of RAE model since it is unsupervised.

Moreover, the results in Table I show that the highest accuracy is achieved by NB since NB favours the most frequent class. Analysing precision, recall and F-measures (36.71%, 33.37% and 34.96% respectively), we see that class-based success is not high. However, for the methods which take temporal information into consideration, such as LSTM, HMM and CRF, these metrics are better. The highest precision is achieved by LSTM (48.58%) while the highest recall is achieved by HMM (44.18%). The highest F-measure is achieved by LSTM (42.95%).

The ability of RAE to reconstruct the features is evaluated by k-means clustering. After clustering the reconstructed features into 9 clusters, the dimensions of the features are reduced to 2D by Principal Component Analysis (PCA). As depicted in Figure 2, RAE is successful to reconstruct the features from different classes.

TABLE I
ACTIVITY RECOGNITION RESULTS

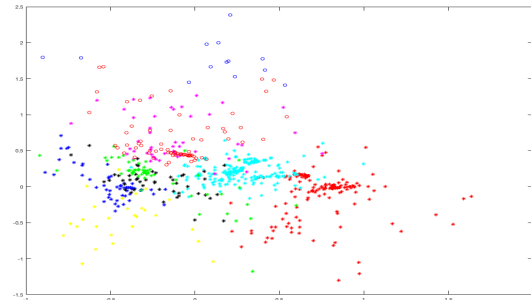| Model	Precision	Recall	F-Measure	Accuracy
NB	36.71%	33.37%	34.96%	87.47%
HMM	37.32%	44.18%	40.46%	84.88%
CRF	42.80%	37.81%	40.15%	84.55%
LSTM	48.58%	38.49%	42.95%	87.02%

Fig. 2. Clustering of RAE re-constructed features. Each colour indicates a different class, while X and Y coordinates are 2D features (2 principal components).
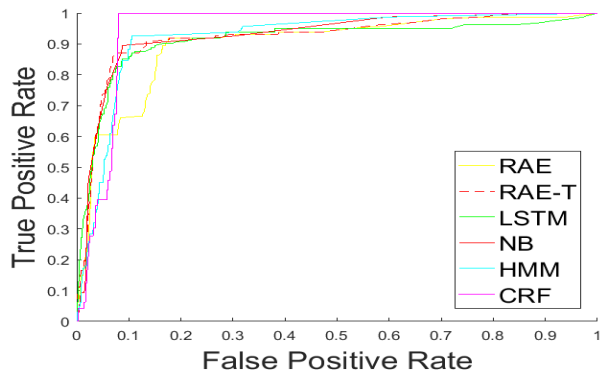

Fig. 3. ROC curves for abnormal behaviour detection.

In the second experiment, the methods are compared in terms of abnormal behaviour detection. The results are depicted as ROC in Figure 3. AUC values for methods CRF, HMM, LSTM, NB, RAE and RAE-T are 94.72%, 92.25%, 91.21%, 92.70%, 89.74%, 92.50% respectively. The results show that RAE based abnormal behaviour detection is competitive with the supervised methods. All methods are good at detecting the abnormal behaviour instances and pruning the false alarms. The proposed RAE based method produces slightly worse TPR and FPR. However its superiority comes from the fact that it doesn't use any labels during the parameter learning process.

Moreover, some of the data of target household is used to re-train the learnt RAE model. In this way, we can tailor the RAE model for the resident by re-tuning the parameters of previously trained RAE using user-specific training examples. This strategy is similar to inductive transfer learning or self-taught learning [29] when none or few data labels are available in the target domain. However, in our case, which is unsuper-

vised, we use only some of the unlabelled data coming from the target dataset. Although we need some data from the target house, this still allows us to reduce several weeks or months of data collection and annotation in the target space to only a few days. For this purpose, RAE learnt on instances of source household dataset is re-trained over 10-days data from house C. The results are shown in the ROC curve (Figure 3 with the abbreviation RAE-T). These results indicate that re-tuning the parameters and considering the house specific behaviour improve the results.

Moreover, we calculate Cohen's Kappa statistics to show the robustness of RAE to detect abnormal behavior. Kappa statistics is a measure that handles both multi-class and imbalanced class problems. It tells how good the classifier is performing over the performance of a classifier that simply guesses at random according to the frequency of each class [30]. However, we use weighted Kappa statistics, since detecting abnormal behaviour is more important than pruning normal ones in our case. In health-care problems, missing a true positive may cause more serious problems than retrieving a high number of false positives. Thus, we assign a higher weight to true positive than false positive in the weight matrix of Kappa. The calculated Kappa for RAE is 0.53, which is a moderate agreement according to [30].

Supervised models, especially deep learning methods such as LSTMs, require too much training data. Moreover, providing labelled data just once would not be enough since observation of dementia sufferers is a task which can be up to years. Also, these models need activity classes to be fixed.However, in a time lapse of years, users may change their behavioural patterns and they may introduce new activity labels. On the other hand, when transfer learning is used, just changing the mapping of sensors and activities would be enough to adapt the model to the new data. Moreover, supervised methods such as NB, HMM, CRF, LSTM doesn't encode time-slices in a hierarchical representation. RAE encodes hierarchy via merging in a bottom-up tree structure. The use of hierarchical models might be a better fit for transfer learning because the different levels of the hierarchy allow a better abstraction between houses.

Although we re-train the learnt RAE model on partial data steming from the source house, when there is no source data available prior, domain adaptation would not be possible. Then, there will be a problem to detect resident specific abnormal behavior. Transfer learning generalises the behavior of inhabitants between different houses and does not take resident specific behavior into account. For example, going to toilet during sleep might be normal for a person, while it is abnormal for another person. A prior distribution can be learnt from the source house and used to provide a sensible initial value for the model parameters of the target house. The behaviour across different houses are transferable under the condition that the resident profiles such as age, gender and lifestyle are similar in both households. Also the different sensors and activities in these houses should be mapped into each other.

The proposed system would improve life experience of dementia sufferers in the following way. The system detects possible candidates for abnormal behaviour to inform the caregiver or the medical doctor. The decision maker will analyse the detected abnormal behavior by considering the person's personal life style. Thus, the proposed method can be used as a decision supporting system rather than a decision making system. Detecting high amount of false positives will not introduce any risk related to the health of the person. Detecting true positives in an early stage would trigger further analysis and would be helpful for an early treatment. The important advantage of the proposed system would be to provide a cognitive status assessment in the natural flow of daily living without annoying elderly people.

V. CONCLUSION

In this paper, we proposed an RAE-based method to detect abnormal behavior of elderly people with dementia. Transfer learning can be an interesting option to cope with scarcity of data. The empirical results showed that the proposed method is promising when supervised methods cannot be exploited because of the lack of (labelled) training data. However, the proposed method failed to detect the person's specific abnormal behaviour. In future, we will consider personal habits by learning a prior distribution from the source house to adapt the model to the target house.

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," *Diabetes Care, vol. 27, no. 5*, pp. 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*, vol. Volume 7, pp. 270–279, 2011.

[3] K. Wild, "Aging changes," in *Geraotechnology, Vol. 9 No 2*, 2010, pp. 121–125.

[4] 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)*, 2015, pp. 149–154.

[5] 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*, pp. 121–132, 2014.

[6] F. J. Ordóñez, P. de Toledo, and A. Sanchis, "Sensor-based bayesian detection of anomalous living patterns in a home setting," *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 259–270, 2015.

[7] N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plotz, "PD disease state assessment in naturalistic environments using deep learning," 2015, pp. 1742–1748.

[8] A. R. M. Forkan, I. Khalil, Z. Tari, S. Foufou, and A. Bouras, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living," *Pattern Recognition*, vol. 48, no. 3, pp. 628 – 641, 2015.

[9] J. Lundstrm, E. Jrpe, and A. Verikas, "Detecting and exploring deviating behaviour of smart home residents," *Expert Systems with Applications*, vol. 55, pp. 429–440, 2016.

[11] U. Naeem, R. Bashroush, R. Anthony, M. Azam, A. Tawil, S. Lee, and M. Wong, "Activities of daily life recognition using process representation modelling to support intention analysis," *International Journal of Pervasive Computing and Communications*, vol. 11, no. 3, pp. 347–371, 2015.

[10] D. Arifoglu and A. Bouchachia, "Activity recognition and abnormal behaviour detection with recurrent neural networks," in *14th International Conference on Mobile Systems and Pervasive Computing*, 2017, pp. 86–93.

[12] R. Socher, C. D. Manning, and A. Y. Ng, "Learning continuous phrase representations and syntactic parsing with recursive neural networks," in *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*, 2010, pp. 1–9.

[13] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '17, 2017, pp. 665–674.

[14] B. Kiran, D. Mathew Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," vol. 4, 01 2018.

[15] C. D. J. C. A. Adriana M. Seelye, Schmitter-Edgecombe M., "Naturalistic assessment of everyday activities and prompting technologies in mild cognitive impairment," *Journal International Neuropsychologly Soc. 2013 (4):*, pp. 442–52, 2013.

[16] D. J. Patterson, D. Fox, H. Kautz, and M. Philipose, "Fine-grained activity recognition by aggregating abstract object usage," in *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, 2005, pp. 44–51.

[17] W. L. H.L Chieu and L. Kaelbling, "Activity recognition from physiological data using conditional random fields." In SMA Symposium. Singapore-MIT Alliance, 2006.

[18] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate activity recognition in a home setting," in *Proceedings of the 10th International Conference on Ubiquitous Computing*, ser. UbiComp '08, 2008, pp. 1–9.

[19] T. Van Kasteren, G. Englebienne, and B. Krose, "Recognizing activities in multiple contexts using transfer learning," *Proceedings of the AAAI Fall Symposium on AI in Eldercare: New Solutions to Old Problems.*, 2008.

[20] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "Transferring knowledge of activity recognition across sensor networks," in *Proceedings of the 8th International Conference on Pervasive Computing*, ser. Pervasive'10, 2010, pp. 283–300.

[21] P. Rashidi and D. J. Cook, "Activity knowledge transfer in smart environments," *Pervasive Mob. Comput.*, vol. 7, no. 3, pp. 331–343, 2011.

[22] 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*, pp. 165–186, 2011.

[23] 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*, vol. 12, no. 4, pp. 1211–1224, 2015.

[24] M. Amiribesheli and A. Bouchachia, "Smart homes design for people with dementia," in *2015 International Conference on Intelligent Environments*, 2015, pp. 156–159.

[25] 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*, vol. 8, no. 1, pp. 133–146, 2017.

[26] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP '11, 2011, pp. 151–161.

[27] J. B. Pollack, "Recursive distributed representations," *Artif. Intell.*, vol. 46, no. 1-2, pp. 77–105, Nov. 1990.

[28] C. Goller and A. Kuchler, "Learning task-dependent distributed representations by backpropagation through structure," in *Proceedings of International Conference on Neural Networks (ICNN'96)*, vol. 1, 1996, pp. 347–352 vol.1.

[29] S. J. Pan and Q. Yang, "A survey on transfer learning, technical report HKUST-CS08-08, department of computer science and engineering, Hong kong university of science and technology, China."

[30] G. Landis, J.R.; Koch, "The measurement of observer agreement for categorical data," *Biometrics 33 (1)*, pp. 159–174, 1977.