Deep Learning and Machine Learning Techniques for Change Detection in Behavior Monitoring

Giovanni Diraco, Alessandro Leone, Andrea Caroppo and Pietro Siciliano

CNR—National Research Council of Italy, IMM—Institute for Microelectronics and Microsystems, Lecce 73010, Italy
{giovanni.diraco, alessandro.leone, andrea.caroppo}@cnr.it,
pietro.siciliano@le.imm.cnr.it

Abstract. Nowadays, smart living environments are equipped with various kinds of sensors which enable enhanced assisted living services. The availability of huge data volumes coming from heterogeneous sources, together with emerging of novel artificial intelligence methods for data processing and analysis, yields a wide range of actionable insights with the aim to help older adults to live independently with minimal supervision and/or support from others. In this scenario, there is a growing demand for technological solutions to monitor human activities and physiological parameters in order to early detect abnormal conditions and unusual behaviors. The aim of this study is to compare state-of-the-art machine learning and deep learning approaches suitable for detecting early changes in human behavior. At this purpose, specific synthetic datasets are generated, which include activities of daily living, home locations and vital signs. The achieved results demonstrate the superiority of deep-learning techniques over traditional supervised/semi-supervised ones in terms of detection accuracy and lead-time of prediction.

Keywords: Change prediction; machine learning; deep learning; ambient assisted living; human behavior.

1 Introduction

Frail subjects, such as elderly or disabled people, may be at risk when their health conditions are amenable to change, as it is quite common in case of chronic conditions. That risk can be reduced by early detecting changes in behavioral and/or physical state, through sensing and assisted living technologies, nowadays available in smart-living environments. Such technologies, indeed, are able to collect huge amounts of data by days, months, and even years, providing important information useful for early detection of changes. Moreover, early change detection makes it possible to alert formal/informal caregivers and health-care personnel in advance when significant changes or anomalies are detected, before critical levels are reached and so preventing chronic diseases. The huge amounts of heterogeneous data collected by different devices require automated analysis; thus there is a growing interest in automatic systems for detecting abnormal activities and behaviors in the context of smart living and elderly monitoring [1]. Health monitoring can benefit from knowledge held in long-term time series of daily activities and behaviors as well as physiological parameters [2]. A lot of research has been done in the general area of human behavior understanding, and more specifically in the area of daily activity/behavior recognition and classification as normal or abnormal [3, 4]. However, very little work is reported in the literature regarding the evaluation of machine learning (ML) techniques suitable for data analytics in the context of long-term elderly monitoring in smart living environments. The purpose of this paper is to investigate the most representative classical machine learning and deep learning (DL) techniques, by comparing them in detecting/predicting changes in human behavior.

The rest of this paper is organized as follows. The following Section 2 contains related works, some background and state-of-the-art in abnormal activity/behavior detection, with special attention paid to elderly monitoring through heterogeneous data collected with multi-sensor systems distributed over indoor environments. Section 3 describes materials and methods used in this study, and provides an overview of the system architecture, the long-term data generation, and the ML/DL techniques. The findings are presented and discussed in Section 4 and Section 5, respectively. Finally, Section 6 draws conclusions and some final remarks.

2 Related Work

Today's available sensing technologies enable long-term continuous monitoring of activities of daily living (ADLs) and physiological parameters (e.g., heart rate, respiration rate, blood pressure, etc.) in the home environment. Normally, both wearable and ambient sensing are used, either alone or combined, as multi-sensor systems. Wearable motion sensors incorporate low-cost accelerometers, gyroscopes and compasses, whereas physiological parameter sensors are based on some kind of skin-contact biosensors (e.g., heart and respiration rates, blood pressure, electrocardiography, etc.) [5]. These sensors need to be attached to a wireless wearable node, carried or worn by the user, needed to process raw data and to transmit/store detected events/signals. Although wearable devices have the advantage of being usable anywhere and their detection performance is generally good (if the signal-to-noise ratio is sufficiently high), nevertheless their usage is extremely limited by battery life time (shortened by the intensive use of wireless communication and on-board processing, both high energy-demanding tasks) [6], by the need to remember to wear a device, and by the discomfort of the device itself.

On the other hand, ambient sensing devices are not intrusive in terms of body obstruction, since they require the installation of sensors around the home environment, such as cameras (monocular/stereo, time-of-flight, Lidar, etc.), microphones, sonars, pyroelectric infrared (PIR) sensors, radar sensors, and pressure/vibration sensors. Such solutions, blending into the home environment, are generally well-accepted by end-users [7].

The learning setups for detecting/predicting behavioral changes can be categorized into three main categories: supervised, semi-supervised, and unsupervised approaches. In the supervised case, abnormalities (i.e., changes) are detected via binary classification in which both normal and abnormal behaviors (i.e., activity sequences) are labelled and used to learn a model [8, 9, 10]. This model is then applied on real-life data in order to classify unlabeled behaviors as normal or abnormal. The problem with this approach is that abnormal behaviors are extremely rare in practice, and thus the easiest way to collect them is the laboratory simulation or synthetic generation. In the semi-supervised case, only one kind of labels, i.e., normal behaviors, are used to train a one-class classifier [11, 12]. Behaviors that do not comply with the learned model are labeled as outliers during the testing phase. The advantage here is that normal behaviors, i.e., real-life data, observed during the execution of common ADLs, are used to train the semi-supervised model (and not simulated or synthetic data as needed in the supervised case). The last but not least important category includes unsupervised classifiers, whose training phase does not need any labeling information (i.e., neither normal nor abnormal behaviors) and any separation into a training and testing phase [13]. In unsupervised learning, only a small fraction of the observed behaviors are assumed to be outliers which exhibit a rather different nature than normal behaviors. In this case, the PRO is that unsupervised-based detection can easily adapt to various real-life environmental and user’s conditions (where no labeling information is available); but the disadvantage is that the unsupervised-based detection requires a quite large amount of initial observations to be fully operational [14, 15].

3 Materials and Methods

For each category of learning setup, i.e., supervised, semi-supervised, and unsupervised, one ML-based and one DL-based technique are evaluated and compared in terms of detection performance and prediction lead-time at the varying of both normal behaviors (NB) and abnormal behaviors (AB). All investigated ML and DL techniques are summarized in Table 1. For that purpose, synthetic datasets are generated by referring to common ADLs and taking into account how older people perform such activities at their home environment (i.e., instructions and suggestions provided by geriatricians and leading researches were taken in careful consideration). The synthetic dataset includes six basic ADLs, four home locations in which these activities usually take place, and five levels of basic vital signs (i.e., heart and respiratory rates) associated with the execution of each ADL. The six ADLs are activity of eating (AE), housekeeping (AH), physical exercise (AP), resting (AR), sleeping (AS), toileting (AT). The considered home locations (LOCs) are bedroom (BR), kitchen (KI), living room (LR), toilet (TO). Regarding vital signs, for simplicity and without loss of generality, the only hear rate signal has been considered by dividing it into five levels (HRLs): very low (VL) < 50 beats/min, low (LO) $\in [50–80]$ beats/min, medium (ME) $\in [80–95]$ beats/min, high (HI) $\in [95–110]$ beats/min, very high (VH) > 110 beats/min.

The objective of this study is to deeply evaluate the learning techniques reported in Table 1 by considering abnormal datasets, obtained by including the following perturbations:

1. Changing of the starting time of one ore more activities (St). This is a change in the starting time of an activity, e.g., having breakfast at 9 AM instead of 7 AM as usual.
2. Changing of the duration of one or more activities (Du). This change refers to the duration of an activity, e.g., resting for 3 hours in the afternoon, instead of 1 hour as usual.
3. Disappearing of one or more activities (Di). In this case, after the change, one activity is no more performed by the user, e.g., having physical exercises in the afternoon.
4. Swapping of two activities (Sw). After the change, two activities are per-formed in reverse order, e.g., resting and then housekeeping instead of housekeeping and resting.
5. Changing the location of one or more activities (Lo). One activity usually performed in a home location (e.g., having breakfast in the kitchen), after the change is performed in a different location (e.g., having breakfast in bed).
6. Changing in heartrate levels when performing one or more activities (Hr). This is a change in heartrate during an activity, e.g., changing from a low to a high heartrate during the resting activity in the afternoon.

Although, the sporadic presence of above mentioned changes is not enough to determine an abnormal condition, nonetheless a sustained change over days or months in activities, locations, and heartrate levels may be linked to an AB. Hence, the aim of this study is to evaluate, the ability of ML and DL techniques in predicting such sustained changes, with the objective to notify caregivers/doctors who can use historical sensor data to make decisions within the application domain of ambient assisted living.

In this study, both normal and abnormal long-term (1-year) datasets are realistically generated by using a probabilistic model based on Hidden Markov Model (HMM) and Gaussian process (GP). The evaluation metrics adopted in this study are sensitivity (SEN) and specificity (SPE), defined as follows:

$$SEN = \frac{TP}{TP+FN}, \; SPE = \frac{TN}{TN+FP}, \tag{1}$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives.

Table 1. Machine learning (ML) and deep learning (DL) techniques compared in this study.

Category	Type	Technique
Supervised	Machine learning	Support vector machine (SVM)
Supervised	Deep learning	Convolutional neural network (CNN)
Semi-supervised	Machine learning	One-class support vector machine (OCSVM)
Semi-supervised	Deep learning	Stacked auto-encoders (SAE)
Unsupervised	Machine learning	K-means clustering (KM)
Unsupervised	Deep learning	Convolutional auto-encoder (CAE)

The lead-time of prediction (LTP) is defined as follows: *maximum number of days, before the day at which the change becomes stable, within which the future change can be predicted with the highest performance*, *i.e.*, *maximizing TP and TN, and minimizing FP and FN*. Thus, the higher the lead-time (in number of days), the better is the overall prediction performance.

3.1 Data Generation

In this study, since HMM is used for data generation (in contrast to other studies where HMM is used for detection purposes [16, 17]), the probabilistic model should be able to take into account the influence of circadian rhythms on motivated behaviors (e.g., sleep, hunger, exercise, etc.) [18]. Although suitability of HMM to model ADLs is encouraged by previous authors' findings [19], nonetheless further research is needed to confirm this choice. The user's physical state bearing diverse ADLs during the daily circadian cycle is modelled by using three hidden states, i.e., Tired (T), Hungry (H), and Energized (E), as depicted in Figure 1. Furthermore, each state can lead to different activities depending on the time of the day (e.g., the state Tired may lead to Sleeping activity in the night and to Resting activity in the afternoon). Each arrow of the graph reported in Figure 1 is associated with a probability parameter, which determines the probability that one state π_i follows another state π_{i-1}, i.e., the transition probability:

$$a_{qr} = P(\pi_i = q \,|\, \pi_{i-1} = r), \tag{2}$$

where $q, r \in \{T, H, E\}$. The HMM output is a sequence of triples $(a, b, c) \in ADL \times LOC \times HRL$, with ADL={AE,AH,AP,AR,AS,AT}, LOC={BR,KI,LR,TO}, and HRL={VL,LO,ME,HI,VI} representing, respectively, all possible ADLs, home locations, and HR levels as previously discussed.

The temporal dependency of activities generated from hidden states is handled by subdividing a day into four time intervals and modeling the activities in each time interval with a dedicated HMM sub-model. For each sub-model M_i, thus, the first state being activated starts at a time T_i modeled as a GP, while the other states within the same sub-model M_i start in consecutive time slots whose durations are also modeled as GPs.

The ADL, LOC and HRL signals are, normally, sampled at different rates according to the specific variability during day time of each signal. For example, the minimum duration of ADLs is of about 10 min, so it is not useful to sample the ADL signal at 1 min interval. Nonetheless, a unique sampling rate can be adopted for all measurements. In this study, the sampling rate of 0.2 sample/min (i.e., 5 min interval between two samples) is selected for all signals. Each dataset is represented in matrix form with rows and columns equal to the total amount of observed days (i.e., 365 days) and to the total amount of samples per day (i.e., 288 samples), respectively. For each dataset, the matrix cells can take 120 different values obtained by combining 6 ADLs, 4 locations, and 5 HR levels. For example, the cell value AE-KI-ME indicates that the subject is eating her meal in the kitchen and her HR level is medium (i.e., between 80 and 95 beats/min). Finally, each 1-year dataset is represented by an image of 365 × 288 pixels with 120 levels of which an example is reported in Figure 2. Additionally, for the sake of understanding, each dataset can be represented by splitting it into three different images, referring to ADLs (6 levels), locations (4 levels), and HR (5 levels), as shown in Figure 3.

Furthermore, to assess the ability of ML and DL techniques (reported in Table 1) to detect behavioral abnormalities and changes, the model parameters (i.e., transition probabilities, emission probabilities, starting times, and durations) were randomly perturbed in order to generate various kind of abnormal datasets. Without loss of generality, each abnormal dataset included only one of the abovementioned changes (i.e., St, Du, Di, Sw, Lo, Hr) at a time or pairs of them, taken without repetitions (i.e., StDu, StDi, StSw, StLo, StHr, etc.).

In order to evaluate the detection performance of ML and DL techniques in Table 1, the HMM parameters (e.g., transition and emission probabilities, starting time and duration of activities, etc.) are gradually perturbed between the 90th and 180th day, by randomly interpolating the parameters of the normal and abnormal models. The resulting perturbed dataset consists of three parts: the first one, ranging from day 1 to day 90, is referred to normal behavior; the second one, from day 90 to 180, is characterized by gradual changes, becoming progressively more accentuated; the third one, starting from day 180, is very different from the initial normal period, the change rate is low or absent, and the subject's behavior moves into another stability period. An abnormal dataset, referred to the St change type, is reported in Figure 4.

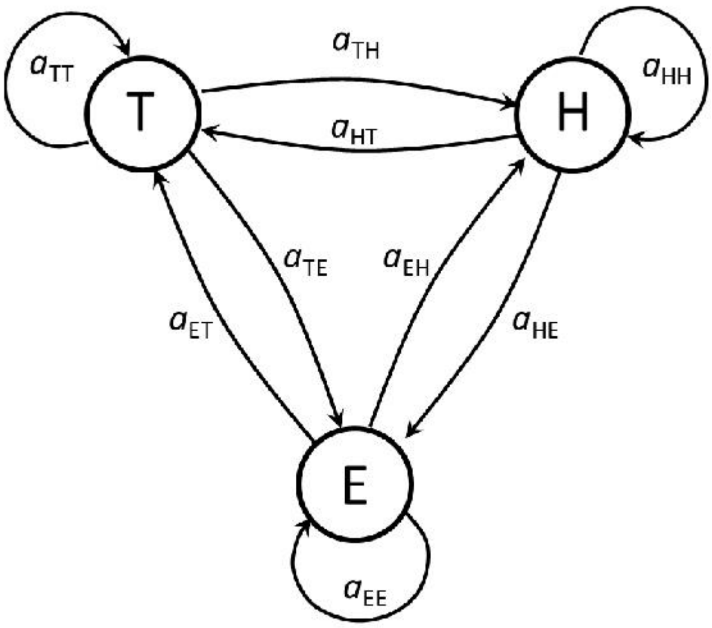

Fig. 1. State diagram of the Hidden Markov Model (HMM) used to generate long-term activity data.

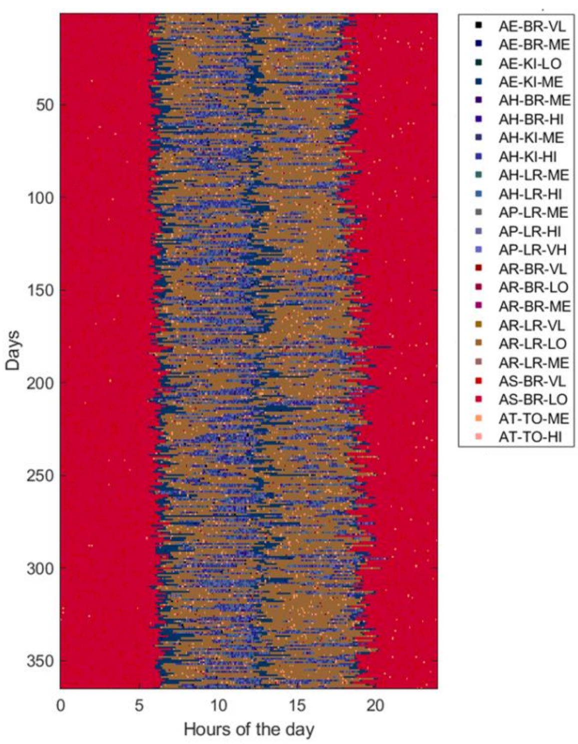

Fig. 2. Example of a normal dataset represented as an image of 365 × 288 pixels and 120 levels (only used levels are reported in the legend).

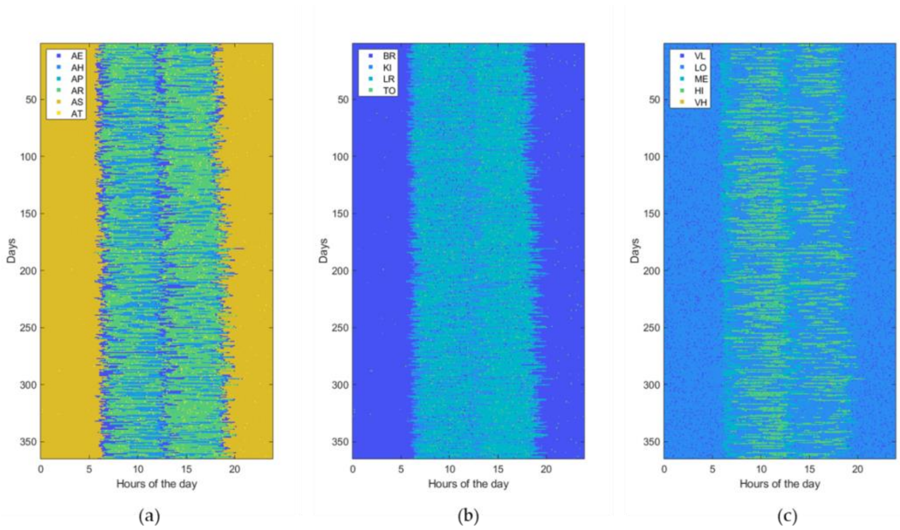

Fig. 3. The same normal dataset shown in Figure 4 but represented with different images for (a) activity of daily living (ADL), (b) home locations (LOCs), and (c) heartrate levels (HRLs).

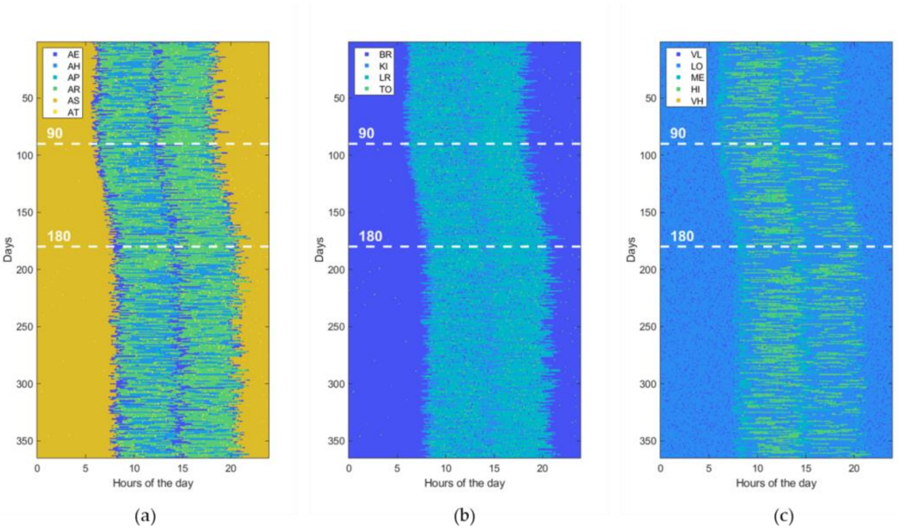

Fig. 4. Example of an abnormal data set, due to a change in the "Starting time of activity" (St). The change gradually takes place from the 90th day on. (a) ADL, (b) LOCs, and (c) HRLs.

3.2 Abnormal behavior detection

As already anticipated, the learning problem can be addressed by means of different learning setups, depending on the label availability. Correspondingly, there are three main detection approaches, i.e., supervised, semi-supervised and unsupervised, and which are taken into account in this study as discussed below.

Supervised detection

Supervised detection is based on learning techniques (i.e., classifiers) requiring fully labelled data for training. This means that both positive samples (i.e., ABs) and negative samples (i.e., normal behaviors) must be observed and labelled during the training phase. However, the two label classes are typically strongly unbalanced, since abnormal events are extremely rare in contrast to normal patterns that instead are abundant. As a consequence, not all classification techniques are equally effective for this situation. In practice, some algorithms are not able to deal with unbalanced data, whereas others are more suitable thanks to their high generalization capability, such as support vector machine (SVM) [20] and artificial neural networks especially those with many layers like convolutional neural networks (CNNs) which have reached impressive performances in detection of AB from videos [21].

Semi-supervised detection.

In real-world applications, the supervised detection workflow described above is not applicable due to the assumption of fully labelled data, on the basis of which abnormalities are known in advance and correctly labeled. However, dealing with elderly monitoring, abnormalities are not known in advance and obviously cannot be purposely performed to train the detection algorithms. Semi-supervised detection is usually achieved by introducing the concept of one-class classification, whose state-of-the-art implementations—as experimented in this study—are one-class SVM (OC-SVM) [22] and auto-encoders (AEs) [23] within ML and DL fields, respectively. DL techniques learn features in a hierarchical way: high-level features are derived from low-level ones by using layer-wise pre-training, in such a way structures of every higher level are represented in higher layers of the network. After pre-training, a semi-supervised training provides a fine-tuning adjustment of the network via gradient descent optimization. Thanks to that greedy layer-wise pre-training followed by semi-supervised fine-tuning [24], features can be automatically learned from large datasets containing only one-class label, associated with normal behavior patterns.

Unsupervised Detection.

The most flexible workflow is that of unsupervised detection. It does not require that abnormalities are known in advance but, conversely, they can occur during the testing phase and are modelled as novelties with respect to normal (usual) observations. The main idea is that extracted features are scored solely on the basis of their intrinsic properties. In order to decide what is normal and not, unsupervised detection is based on appropriate metrics of either distance or density. The distance used in this study is defined as follows:

$$D(\bar{s}) = \sum_{i=1}^{N} |s_i| \ln(i), \tag{3}$$

where $N = 288$ and $\bar{s} = (s_1, s_2, \dots, s_N) \in \mathbb{R}^N$ is a day of observation, i.e., a row of the matrix dataset.

Clustering techniques can be applied in unsupervised detection. In particular, K-means is one of the simpler unsupervised algorithms that address the clustering problem by grouping data based on their similar features into K disjoint clusters. However, K-means is affected by some shortcomings: (1) sensitivity to noise and outliers; (2) initial cluster centroids (seeds) are unknown (randomly selected); and (3) there is no criterion for determining the number of clusters. The Weighted K-Means [25], also adopted in this study, provides a viable way to approach clustering of noisy data. The last two problems are addressed by implementing the intelligent K-means suggested by [26], in which the K-means algorithm is initialized by using the so-called anomalous clusters, extracted before running the K-means itself.

3.3 Experimental setting

For the experimental purpose, 31500 datasets were generated, i.e., 1500 random instances for each of the 21 abnormalities, obtained by considering the abnormalities (St, Du, Di, Sw, Lo, Hr), together with all pairs of these abnormalities taken without repetitions. Each dataset represented a 1-year data collection, as a matrix (image) of 365 rows (days) and 288 columns (samples lasting 5 min each), for a total amount of 105120 values (pixels) through 120 levels. The feature extraction process was carried out by considering a 50%-overlapping sliding window lasting 25 days, then leading to a feature space of dimension D = 7200.

Each dataset was divided into three parts: Upper (1st–90th days), middle (90th–180th days), and lower (180th–365th days) regions. The feature vectors (i.e., ACT-LOC-HRL sequences) belonging to the upper regions were negative samples (i.e., normal behavior), whereas those belonging to the lower regions were positive ones (i.e., AB). The middle regions were, instead, considered as prediction regions, characterized by gradual changes becoming progressively more accentuated. The aim is to classify the feature vectors belonging to the middle regions in order to predict the likelihood of a future change which will become increasingly relevant and stable from the 180th day onwards (lower region).

In both supervised and semi-supervised settings, regarding the SVM classifier, a radial basis function (RBF) kernel was used. The kernel scale was automatically selected using a grid search combined with cross-validation on randomly subsampled training data. Regarding the CNN-based supervised detection, the network structure included eight layers: four convolutional layers with a kernel size of 3 × 3, two subsampling layers, and two fully connected layers. Finally, the two output units represented, via binary logical regression, the probability of normal and abnormal pattern behaviors.

The stacked auto-encoder (SAE) network was structured in four hidden layers, and the sliding-window feature vectors were given as input to the first layer, which thus included 7200 units. The second hidden layer was of 900 units, corresponding to a compression factor of 8 times. The following two hidden layers were of 180 and 60 units, respectively, with compression factors of 5 and 3 times. In supervised detection settings, the six abnormal datasets were joined in order to perform a 6-fold cross-validation scheme. In semi-supervised detection settings, instead, only normal data from the same dataset were used for training, while testing was carried out using data from day 90 onwards.

Regarding the convolutional auto-encoder (CAE) structure, the encoder included three convolutional layers with a kernel size of five, five, and three, respectively, followed by a fully connected layer. The decoder structure was a mirror of the encoder one. All experiments were performed on an Intel i7 3.5 GHz workstation with 16GB DDR3 and equipped with GPU NVidia Titan X using Keras [27] with Theano [28] toolkit for DL approaches, and Matlab [29] for ML approaches.

4 Results and discussion

The achieved results are reported in Table 2 in terms of specificity, sensitivity and LTP for all learning techniques evaluated in the present study. Note that the latter is related to SEN and SPE since LTP refers to the average number of days, before the 180th day (from which the changed behavior becomes stable), at which the change can be detected with such sensitivity and specificity. The longer the LTP the earlier the change can be predicted.

Generally, detecting ABs by using supervised methods presents the shortcoming of requiring both positive samples (i.e., changing activity sequences) and negative samples (i.e., habitual activity sequence) for model training. The SVM-based and CNN-based detection have been evaluated by training models with positive and negative samples taken from different datasets, in order to reproduce more accurately the real-life conditions mentioned above. Due to the lack of real data for training discussed above, the supervised approaches achieved the lowest detection performances.

The problem of training data availability is mitigated with semi-supervised techniques, since only negative samples (i.e., normal behaviors) are required, which are quite abundant in everyday activities. However, the main difficulty is to select training samples that are most representative of normal behaviors. The semi-supervised approaches evaluated in this study, i.e., OC-SVM and SAE, achieved intermediary detection performances, although with lower prediction performance (LTP) due to the difficulty selection of suitable (negative) samples for training.

The most promising results were obtained with the unsupervised learning methods, i.e., K-means (KM) and DC, in which no labeled data were necessary, allowing the easy adaptability to different environmental conditions as well as to users' physical characteristic and habits [13]. The KM detection, however, required an initial observation period during which the system was unable to detect changes from usual activity, negatively affecting the resulting prediction performance.

Classical ML methods, such as SVM and OC-SVM, have to deal with the problem of learning a probability distribution from a set of samples, which generally means to learn a probability density that maximizes the likelihood on given data. Conversely, such density does not always exist, as what happens when data lie on low-dimensional manifolds, which is the case of change types involving a narrow range of values (e.g., if the change regards only few heartrate levels), or when training and testing data come from a different probability distribution (e.g., as in the case of supervised learning). Under such a point of view, conversely, DL methods are more effective because they follow an alternative approach. Instead of attempting to estimate a density, which may not exist, they define a parametric function (deep network) able to generate samples closer to data samples taken from the original data distribution (by hyper-parameter tuning).

Table 2. Detection performance and lead-time of prediction for each technique.

	SVM	**CNN**	**OCSVM**	**SAE**	**KM**	**CAE**
SEN (%)	87	90	89	93	96	98
SPE (%)	91	92	90	92	97	99
LTP (days)	10	11	6	9	14	22

5 Conclusions

The contribution of this study is twofold. First, a common data model able to represent and process simultaneously ADLs, home locations and vital signs as image matrices is presented. Second, the performance of state-of-the-art ML-based and DL-based detection techniques have been evaluated by considering large datasets, synthetically generated, including both normal and abnormal behaviors. The achieved results are promising and show the superiority of DL-based techniques in dealing with huge datasets characterized by different kinds of data distribution. Future and ongoing activities are focused on the evaluation of ML/DL learning techniques in different domains, such as clinical decision support system and predictive maintenance.

References

1. Sharma, R.; Nah, F.; Sharma, K.; Katta, T.; Pang, N.; Yong, A. Smart living for elderly: De-sign and human-computer interaction considerations. Lect. Notes Comput. Sci. 2016, 9755, 112–122.
2. Parisa, R.; Mihailidis, A. A survey on ambient-assisted living tools for older adults. IEEE J. Biomed. Health Inform. 2013, 17, 579–590.
3. Mabrouk, A.B.; Zagrouba, E. Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Syst. Appl. 2018, 91, 480–491.
4. Bakar, U.; Ghayvat, H.; Hasanm, S.F.; Mukhopadhyay, S.C. Activity and anomaly detection in smart home: A survey. Next Gener. Sens. Syst. 2015, 16, 191–220.
5. Taraldsen, K.; Chastin, S.F.M.; Riphagen, I.I.; Vereijken, B.; Helbostad, J.L. Physical activity monitoring by use of accelerometer-based body-worn sensors in older adults: A systematic literature review of current knowledge and applications. Maturitas 2012, 71, 13–19.
6. Min, C.; Kang, S.; Yoo, C.; Cha, J.; Choi, S.; Oh, Y.; Song, J. Exploring current practices for battery use and management of smartwatches. In Proceedings of the ISWC '15 Proceedings of the 2015 ACM International Symposium on Wearable Computers, Osaka, Japan, 07–11 September 2015.
7. Stara, V.; Zancanaro, M.; Di Rosa, M.; Rossi, L.; Pinnelli, S. Understanding the Interest Toward Smart Home Technology: The Role of Utilitaristic Perspective. In ForItAAL; Springer: Berlin, Germany, 2018.
8. Miao, Y.; Song, J. Abnormal Event Detection Based on SVM in Video Surveillance. In Proceedings of the IEEE Workshop on Advance Research and Technology in Industry Applications, Ottawa, ON, Canada, 29–30 September 2014.
9. Forkan, A.R.M.; Khalil, I.; Tari, Z.; Foufou, S.; Bouras, A. A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. Pattern Recognit. 2015, 48, 628–641.
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in neural information processing systems, Lake Tahoe, Nevada, 3–6 December 2012.
11. Hejazi, M.; Singh, Y.P. One-class support vector machines approach to anomaly detection. Appl. Artif. Intell. 2013, 27, 351–366.
12. Krizhevsky, A.; Hinton, G.E. Using very deep autoencoders for content-based image retrieval. In Proceedings of the 19th European Symposium on Artificial Neural Networks, Bruges, Belgium, 27–29 April 2011.
13. Otte, F.J.P.; Saurer, R.B.; Stork, W. Unsupervised Learning in Ambient Assisted Living for Pattern and Anomaly Detection: A Survey. CCIS 2013, 413, 44–53.
14. De Amorim, R.C.; Mirkin, B. Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering. Pattern Recognit. 2012, 45, 1061–1075.

15. Chiang, M.M.T.; Mirkin, B. Intelligent choice of the number of clusters in k-means clustering: An experimental study with different cluster spreads. J. Classif. 2010, 27, 3–40.
16. Forkan, A.R.M.; Khalil, I.; Tari, Z.; Foufou, S.; Bouras, A. A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. Pattern Recognit. 2015, 48, 628–641.
17. Kim, E.; Helal, S.; Cook, D. Human activity recognition and pattern discovery. IEEE Pervasive Comput. 2010, 9.1, 48.
18. Antle, M.C.; Silver, R. Circadian insights into motivated behavior. In Behavioral Neuroscience of Motivation; Springer: Berlin, Germany, 2015; pp. 137–169.
19. Caroppo, A.; Diraco, G.; Rescio, G.; Leone, A.; Siciliano, P. Heterogeneous sensor platform for circadian rhythm analysis. In Proceedings of the 6th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), Gallipoli, Italy, 10 August 2015, pp. 187-192.
20. Hu, W.; Liao, Y.; Vemuri, V.R. Robust anomaly detection using support vector machines. In Proceedings of the international conference on machine learning, Los Angeles, California (USA), 23-24 June 2003; pp. 282–289.
21. Sabokrou, M.; Fayyaz, M.; Fathy, M.; Moayed, Z.; Klette, R. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. Comput. Vis. Image Underst. 2018, 172, 88–97.
22. Hejazi, M.; Singh, Y.P. One-class support vector machines approach to anomaly detection. Appl. Artif. Intell. 2013, 27, 351–366.
23. Krizhevsky, A.; Hinton, G.E. Using very deep autoencoders for content-based image retrieval. In Proceedings of the 19th European Symposium on Artificial Neural Networks, Bruges, Belgium, 27–29 April 2011.
24. Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.A.; Vincent, P.; Bengio, S. Why does unsupervised pre-training help deep learning? J. Mach. Learn. Res. 2010, 11, 625–660.
25. De Amorim, R.C.; Mirkin, B. Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering. Pattern Recognit. 2012, 45, 1061–1075.
26. Chiang, M.M.T.; Mirkin, B. Intelligent choice of the number of clusters in k-means clustering: An experimental study with different cluster spreads. J. Classif. 2010, 27, 3–40.
27. Chollet, F. Keras. GitHub repository. Available online: https://github.com/fchollet/keras (accessed on February 12, 2019).
28. Bastien, F.; Lamblin, P.; Pascanu, R.; Bergstra, J.; Goodfellow, I.J.; Bergeron, A.; Bouchard, N.; Bengio, Y. Theano: New Features and Speed Improvements. Deep Learning and Unsupervised Feature Learning NIPS Workshop: Lake Tahoe, California/Nevada (USA) 2012.
29. Matlab R2014; The MathWorks, Inc.: Natick, MA, USA. Available online: https://it.mathworks.com (accessed on March 21, 2014).