



TraMiner: Vision-Based Analysis of Locomotion Traces for Cognitive Assessment in Smart-Homes

Samaneh Zolfaghari¹ · Elham Khodabandehloo² · Daniele Riboni¹

Received: 3 September 2020 / Accepted: 29 December 2020
© The Author(s) 2021

Abstract

The rapid increase in the senior population is posing serious challenges to national healthcare systems. Hence, innovative tools are needed to early detect health issues, including cognitive decline. Several clinical studies show that it is possible to identify cognitive impairment based on the locomotion patterns of the elderly. In this work, we investigate the use of sensor data and deep learning to recognize those patterns in instrumented smart-homes. In order to get rid of the noise introduced by indoor constraints and activity execution, we introduce novel visual feature extraction methods for locomotion data. Our solution relies on locomotion trace segmentation, image-based extraction of salient features from locomotion segments, and vision-based deep learning. We carried out extensive experiments with a large dataset acquired in a smart-home test bed from 153 seniors, including people with cognitive diseases. Results show that our system can accurately recognize the cognitive status of the senior, reaching a macro- F_1 score of 0.873 for the three categories that we target: cognitive health, mild cognitive impairment, and dementia. Moreover, an experimental comparison shows that our system outperforms state-of-the-art methods.

Keywords Pervasive healthcare · Artificial intelligence · Cognitive decline · Deep learning

Introduction

Nowadays, different societal factors are determining a significant reduction in fertility and an increase in life expectancy. As a consequence, we are observing a rapid and constant growth of the average population age. This phenomenon determines increasing pressure on healthcare systems. Indeed, more and more seniors affected by noncommunicable diseases (NCDs) need to be hospitalized for long periods, and this trend is projected to continue for the foreseeable future [1]. Unfortunately, the recent COVID-19 pandemic has dramatically exposed the fragility of national healthcare systems when dealing with increasing numbers

of patients [2], and the need for innovative technological solutions to cope with global health crises has emerged. Among the NCDs that affect the elderly population, one of the most prevalent is dementia. For instance, the prevalence of dementia is estimated to 14% in the U.S. population aged 70 or older, with a total cost estimated from \$157 billion to \$215 billion, and this cost is projected to triple by 2050 [3]. Of course, in addition to monetary costs, dementia and other cognitive issues disrupt the quality of life of patients and of those caring for them. Hence, in order to prolong independent and healthy aging, it is urgent to devise novel tools to continuously monitor the mental wellness of seniors and to early detect abnormal cognitive decline in an unobtrusive and privacy-conscious manner.

In this paper, we contribute to this grand challenge by proposing *TraMiner*, a novel system based on Trajectory Mining for continuous cognitive assessment in smart-homes. Our system relies on clinical indicators that characterize cognitive decline in terms of abnormal locomotion patterns of the elderly. Different solutions have been recently proposed to recognize those patterns outdoors, including [4–6]. However, most seniors, and especially those experiencing cognitive issues, normally spend a large part of their time at home. Hence, indoor locomotion traces provide invaluable information for assessing

✉ Daniele Riboni
riboni@unica.it

Samaneh Zolfaghari
samaneh.zolfaghari@unica.it

Elham Khodabandehloo
ekhodabandehloo@mail.kntu.ac.ir

¹ Department of Mathematics and Computer Science,
University of Cagliari, Cagliari, Italy

² Department of Geo-spatial Information Systems, K. N. Toosi
University of Technology, Tehran, Iran

their cognitive health status. Unfortunately, recognizing those patterns at home is challenging, since the inhabitant movements are constrained by the home environment, and they are affected by the execution of daily living activities. The problem of indoor abnormal pattern detection for cognitive assessment has been investigated in a few previous works. Up to date, most of those works rely on manual feature engineering to define statistical measures characterizing locomotion traces [7–9]. However, locomotion traces encode involved spatio-temporal information, which can be only partially captured through a fixed set of numerical features. For this reason, we take a different approach, representing locomotion traces through images and relying on Deep Learning (DL) for the classification of those images according to the cognitive health status of the inhabitant. Our method relies on specific techniques for locomotion data cleaning and segmentation. Each locomotion segment is transformed into two images that capture different aspects related to the clinical indicators of locomotion anomalies. Each couple of images is fed to a DL classifier in charge of predicting the anomaly level of the corresponding locomotion segment. Since it would be unrealistic to provide a hypothesis of diagnosis based on the observation of a single trajectory, our system includes a long-term analysis module. The latter is in charge of processing the history of DL predictions for computing a hypothesis of diagnosis, which is provided to clinicians for supporting the neuropsychological assessment.

We have developed a prototype of TraMiner system, and carried out extensive experiments with a real-world dataset gathered in a smart-home test bed with 153 seniors, including cognitively healthy subjects, people with Mild Cognitive Impairment (MCI) [10], and people with dementia (PwD). Results proved the accuracy of TraMiner in predicting the cognitive health status of the inhabitants. Indeed, based on the observation of the locomotion traces acquired during one day per patient, our system achieved a macro- F_1 score of 0.873. Moreover, an experimental comparison showed the superiority of our approach with respect to state-of-the-art techniques. In order to enable the full reproducibility of our experiments, we have released the code of all our system components, and the used dataset is available on the Web.

A preliminary investigation of this work was presented in [11]. In this paper, we extend our preliminary investigation with: (i) support for recognition of MCI, (ii) a novel locomotion trace segmentation algorithm, (iii) a new visual feature extraction technique, (iv) a novel two-input DL architecture, (v) an additional algorithm for long-term assessment of the cognitive health status, (vi) experiments with a larger dataset, (vii) new experiments to evaluate all the components of our system, (viii) an experimental comparison with state-of-the-art techniques, and (ix) a dashboard accessible on the Web to inspect the visual features and the system predictions.

The rest of the paper is structured as follows. Section 2 introduces preliminary notions and discusses related work. Section 3 illustrates our overall system. Section 4 describes the techniques for trajectory segmentation and visual feature extraction. Section 5 explains our methods for DL-based trajectory classification and long-term analysis. Section 6 reports experimental results. Finally, Section 7 concludes the paper and outlines future research directions.

Preliminaries and Related Work

In this section, at first, we illustrate different clinical indicators to recognize cognitive decline based on locomotion data, which are the theoretical basis of our work. Then, we review existing techniques based on Internet of Things (IoT) technologies to automatically detect abnormal locomotion patterns that may indicate cognitive decline.

Clinical Indicators of Abnormal Locomotion Patterns

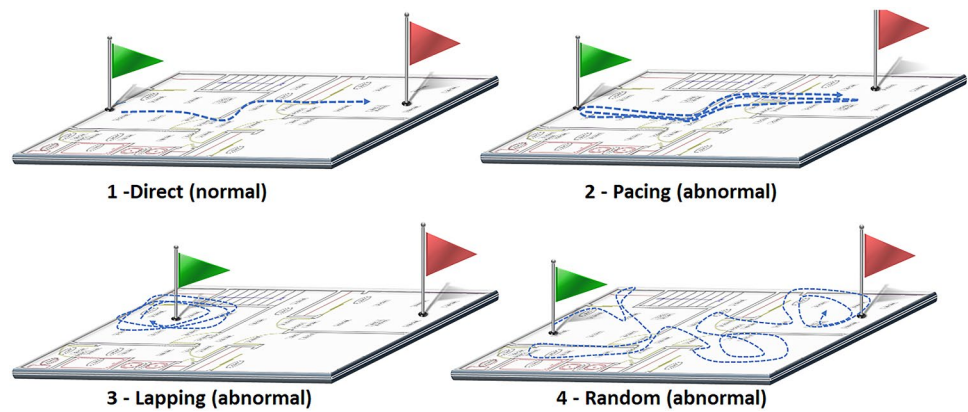
Several studies have attempted to characterize the typical locomotion patterns of people experiencing cognitive decline. Some models relied on the notion of *wandering*, defined by Algase et al. [12] as a “*syndrome of dementia-related locomotion behavior having a frequent, repetitive, temporally disordered, and/or spatially disoriented nature that is manifested in lapping, random, and/or pacing patterns*”. Other researchers characterized cognitive decline based on the observation of subtle gait anomalies. In the following, we present the models adopted in this work.

Martino-Saltzman Model

One of the most well-known and widely accepted models for categorizing the patterns of wandering behaviors in dementia was proposed by Martino-Saltzman [13]. This model categorizes the trajectories into one of four distinctive patterns of movement: direct, random, lapping, or pacing, which are shown in Fig. 1:

- **Direct:** a simple or uncomplicated trajectory from one location to another one.
- **Pacing:** at least three consecutive back-and-forth movements between two locations along very similar paths.
- **Lapping:** at least two circular movements between at least three distinct locations.
- **Random:** a continuous and aimless movement across numerous locations with multiple directional changes, that generally passes through more than four locations.

Fig. 1 Travel patterns of people according to the Martino-Saltzman model of wandering behavior [13]



Based on the Martino-Saltzman model, *random*, *pacing*, and *lapping* patterns are indicators of cognitive issues. As explained in Introduction, several random, pacing, and lapping patterns can be observed in the home, which are actually due to the normal execution of daily living activities by cognitively healthy inhabitants. Hence, in this work, we do not aim at explicitly recognizing those patterns, but we rely on a supervised learning approach for recognizing abnormal locomotion patterns.

Low-Level Motion Indicators

In addition to clinical indicators based on wandering behavior, different low-level indexes were introduced in other works as indicative of cognitive decline. Those indicators have been used to automatically assess the cognitive status in IoT systems [14]:

- **Jerk** [15] is the rate at which a person's acceleration changes with respect to time.
- **Sharp angles** [6] are defined as vector angles in a trajectory being equal to or more than 90 degrees.
- **Straightness** [16] is defined as the ratio of the distance between two consecutive trajectory segments and the distance between the start and end point of these segments.
- **Turning angle** [9] is defined as the sum of the absolute angles between any two subsequent lines in a trajectory.
- **Path efficiency** [9] is defined as the ratio between the distance from the start to the end of a trajectory and the trajectory length.

IoT Techniques for Detecting Locomotion Anomalies

In the last few years, different methods have been proposed to recognize cognitive issues based on artificial intelligence (AI) [17], and several AI-based solutions have been proposed to support active and healthy aging [18–20].

Riboni et al. relied on activity recognition algorithms to detect overt errors in the execution of activities of

daily living, which indicate cognitive decline [21]. Other works, including those of Seelye et al. [22] and Dawadi et al. [23], rely on sensor activations and activity patterns to distinguish PwD subjects from cognitively healthy ones. However, activity recognition systems determine privacy issue, especially when deployed in homes. Moreover, those systems require a dense sensing infrastructure, while in this work we rely solely on an indoor positioning system.

Indeed, a promising direction consists in tracking the movements of the elderly through unobtrusive positioning technologies, and analyzing location traces according to clinical indicators as the ones mentioned in Section 2.1. Methods based on low-level motion indicators usually rely on body-worn sensors. Mc Ardle et al. used accelerometers worn by the senior to detect symptoms of cognitive impairment [24]. In that work, the authors extracted spatiotemporal features including gait variability, rhythm, asymmetry, and postural aspects of gait. An accelerometer-based method to detect challenging behaviors typical of PwD was proposed by Goerss et al. [25]. That method relies on the extraction of actimetric movement features from raw sensor data. Experiments carried out with 17 seniors in a retirement home showed that the long-term observation of those features provided relevant information to identify the challenging patterns of PwD. The correlation between low-level motion behaviors and dementia was assessed in different studies, including the one carried out by Kirste et al. relying on ankle-mounted three-axis accelerometers [26]. However, body-worn systems determine usability issues, especially for the senior population [27]. Hence, in our work we disregard body-worn sensors, and we rely on fixed positioning infrastructures to detect indoor location traces.

Most existing works aimed at recognizing cognitive impairment based on locomotion patterns rely on outdoor movements. Lin et al. [6] analyzed GPS trajectories outdoors for detecting wandering according to the Martino-Saltzman model. A smart GPS tracker was devised by Ng and Kong for supporting secure and independent outdoor walking of the elderly [5]. That system is based on wandering

detection and activity recognition and can adapt to personal locomotion wandering patterns. Schaaf et al. investigated the use of GPS and accelerometer data for real time detection of disorientation in PwD [4].

Since many elderly people, and especially those who are experiencing cognitive decline, spend large part of their time at home, indoor locomotion data can be a rich source of information to assess their cognitive status. Unfortunately, indoor wandering recognition is particularly challenging, since movements are constrained by the ambient shape, and are impacted by the execution of activities of daily living. A few previous works have addressed this issue. Vuong et al. applied supervised machine learning methods for automatically classifying indoor trajectories according to the Martino–Saltzman model [28]. Lin et al. proposed a method to identify repetitive locomotion episodes in the home according to well-known models of wandering [29]. Khodabandehloo and Riboni proposed a collaborative mining approach to assess the cognitive status of smart-home inhabitants based on statistical features extracted from their trajectories [7]. Kearns et al. proposed the use of accurate localization technologies deployed in a common indoor living space of a retirement home, and measured the tortuosity of trajectories to recognize wandering episodes [8]. In a subsequent study on the same setup, the authors found out that other features, including speed, path efficiency, and angle-turn, were predictive of dementia [9]. Other studies showed that the relation between in-home walking velocity and activity patterns was significantly correlated to the cognitive status of the inhabitant [30].

However, feature engineering from trajectory data is not a trivial task. Indeed, the abundance of spatiotemporal information encoded by movement traces, and the presence of noise introduced by positioning technologies, make it challenging to summarize trajectory data through numerical features. In order to overcome the feature engineering challenge, a recent research direction consists in representing complex information through images, and using Deep Neural Networks (DNN) for images classification to solve the AI task. This approach is used in different domains, including financial forecasting [31], activity recognition [32], and predictive maintenance [33], just to name a few. In this paper, we pursue this approach. The rationale of our choice is that trajectory images may effectively capture discriminative features without the need of sophisticated feature engineering efforts. This approach was applied to trajectory data in few previous studies. In a recent work, Wang et al. [34] proposed an action recognition method based on the representation of 3D skeleton sequences as 2D images. Sequences of movements are represented using different color distributions. They used DL, in particular, convolutional neural network (CNN), to detect human actions. Kieu et al. used trajectories in the form of 3D images

containing geographical features, together with driving behavior features, to predict the identity of drivers [35]. They applied unsupervised auto encoder neural networks to avoid over-fitting, and supervised neural networks to identify drivers. Endo et al. used DNN as an automatic feature extraction method, along with handcrafted features, for estimating users' transportation modes from images depicting their movement trajectories [36]. Transportation modes were then recognized using supervised learning algorithms.

In this paper, we investigate the application of image-based trajectory classification to the domain of cognitive assessment. To the best of our knowledge, the only previous work that adopted a similar approach to the same domain was proposed by Gochoo et al. in [37]. In that work, the authors represent trajectories in a two-dimensional grid, in which one dimension represents time, while the other dimension represents mono-dimensional space. They map each three-dimensional spatiotemporal trajectory point into the two-dimensional grid. The corresponding binary image is then classified using deep convolutional neural networks to recognize different patterns of wandering behavior. However, as a consequence of the transformation from two-dimensional to one-dimensional space, the spatial information is partially disrupted, since metric operations and topological relationships are not preserved. Indeed, points that are close in the two-dimensional geographic space are not necessarily close in the one-dimensional space of the grid.

In our work, we pursue a different direction, retaining the spatial information in trajectory images, and enriching them with additional visual features encoding low-level motion indicators of cognitive decline. To the best of our knowledge, apart from the preliminary investigation that we carried out in [11], this is the first work that investigates this method for cognitive assessment.

TraMiner System Overview

In this paper, we assume a smart-home infrastructure capable of continuously monitoring the inhabitant's position at a fine-grained level. For the sake of this work, we consider the case of a person living alone in the home. This is a common situation for elderly people. Moreover, seniors living alone may have particular benefits from remote monitoring and assessment of cognitive functions. In order to support seniors living with other people or pets, our system could be easily extended by adopting an identity-aware positioning system, or by applying a *data association* algorithm in charge of associating each location reading to the individual that triggered the corresponding sensor [38]. Our system

may extract position information not only from localization infrastructures, but also based on the inhabitant's interaction with sensorized objects and appliances with a fixed position in the home.

Figure 2 illustrates the TraMiner architecture. Since the focus of this paper is on processing locomotion data, the core methods of our contribution are largely independent from the available sensor infrastructure. We assume that a SMART-HOME SENSOR INFRASTRUCTURE is in charge of continuously gathering data about the inhabitant's position, e.g., through passive infrared (PIR) sensors. The smart-home system communicates raw sensor data to a STREAM PROCESSING SOFTWARE PLATFORM (e.g., Apache Kafka) for integration and temporal synchronization.

Each time a sensor fires, the platform sends a **raw sensor event** $rse = \langle t, s_id, v \rangle$ to the TraMiner system, where t is the timestamp of firing, s_id is the sensor's unique identifier, and v is the generated value. For instance, the raw sensor event:

$$rse = \langle sens5371, 2020-08-30\ 16:46:18.323, open \rangle,$$

states that the sensor identified as 'sens5371' (for instance, a sensor attached to the fridge door) fired a value 'open' at timestamp '2020-08-30 16:46:18.323'. Since we assume that the home is inhabited by a single individual, we are not interested in associating the sensor record to the person that triggered it.

The INTEGRATED POSITIONING SYSTEM is in charge of deriving spatiotemporal information from raw sensor events. To this aim, it relies on a **sensor position table** storing the relative position of each sensor in the home. A record of sensor position in that table is a triple: $\langle s_id, x, y \rangle$, where (x, y) are the relative coordinates of the sensor identified by 's_id' in the home. For the sake of simplicity, we consider only static sensors in the home; hence, we assume that the position of sensors does not change with time. Each time the integrated positioning system receives a raw sensor event, it joins the corresponding record with the sensor position table to obtain the (x, y) coordinates, producing a **user's position record** $r = \langle p, t \rangle$, where $p = (x, y)$ are the relative coordinates of the sensor that fired at time t .

The TRAJECTORY SEGMENTATION module is in charge of reducing noise in the data and partitioning the temporal stream of position records into **trajectories**. A trajectory is a temporally contiguous sequence of positions which corresponds to a locomotion episode. Each trajectory is passed to the modules FOR TRAJ and SPEED FEATURE EXTRACTION. Those modules represent the trajectory as an image each. The two images represent the walked trajectory in a two-dimensional space and highlight different visual features, such as speed and intersection points.

The TRAJECTORY IMAGES CLASSIFICATION module is in charge of classifying each trajectory as either walked by

a cognitively healthy person, MCI individual, or PwD. To this aim, it uses a DL classifier processing the two images corresponding to each trajectory. The classifier is trained on the cloud using a set of anonymous trajectories labeled according to the cognitive health status of the individual. According to the history of predictions, the LONG-TERM TRAJECTORY ANALYSIS module computes a hypothesis of diagnosis for the individual. The hypothesis may be *cognitively healthy*, *MCI*, or *PwD*. Finally, the hypothesis of diagnosis is communicated to the remote healthcare center, to support the clinicians in the evaluation of the patient.

Trajectory Segmentation and Visual Feature Extraction

This section reports the details of modules in charge of trajectory segmentation and visual feature extraction.

Position Data Cleaning and Trajectory Segmentation

As explained before, the integrated positioning system continuously provides TraMiner with user's position records. These records instantiate the **position history** $H = \langle r_1, r_2, \dots, r_n \rangle$, which is the temporal sequence of user's positions records. Inevitably, the position history contains inaccuracies because, in real-world conditions, sensor data are affected by a relevant level of noise. In this regard, noise reduction is an inevitable step in preparing the data. For addressing this problem, we manually analysed the spatiotemporal information of the dataset used in our experiments by plotting movement traces over the home layout. We observed several unfeasible deviations from the expected trajectories, which were due to wrong position detection by the positioning system. Hence, the TRAJECTORY SEGMENTATION module performs the following preliminary steps for noise reduction:

- We set a threshold T_v for the maximum possible velocity v of a person moving in the home. If the speed between any two consecutive position records $\langle r_i, r_{i+1} \rangle \in H$ is higher than T_v , the record r_{i+1} is considered a noisy reading; hence, it is deleted from H . For the sake of this work, we set T_v to 15 m/s.
- By considering the arrangement of sensors in the test-bed of our experiments, the maximum distance between any two adjacent sensors is below 3 m. By carefully analyzing the persons trajectories, we observed that some paths spatially deviated from the expected trajectory due to abrupt movements between non-contiguous sensors. In this regard, if the distance between the positions of two

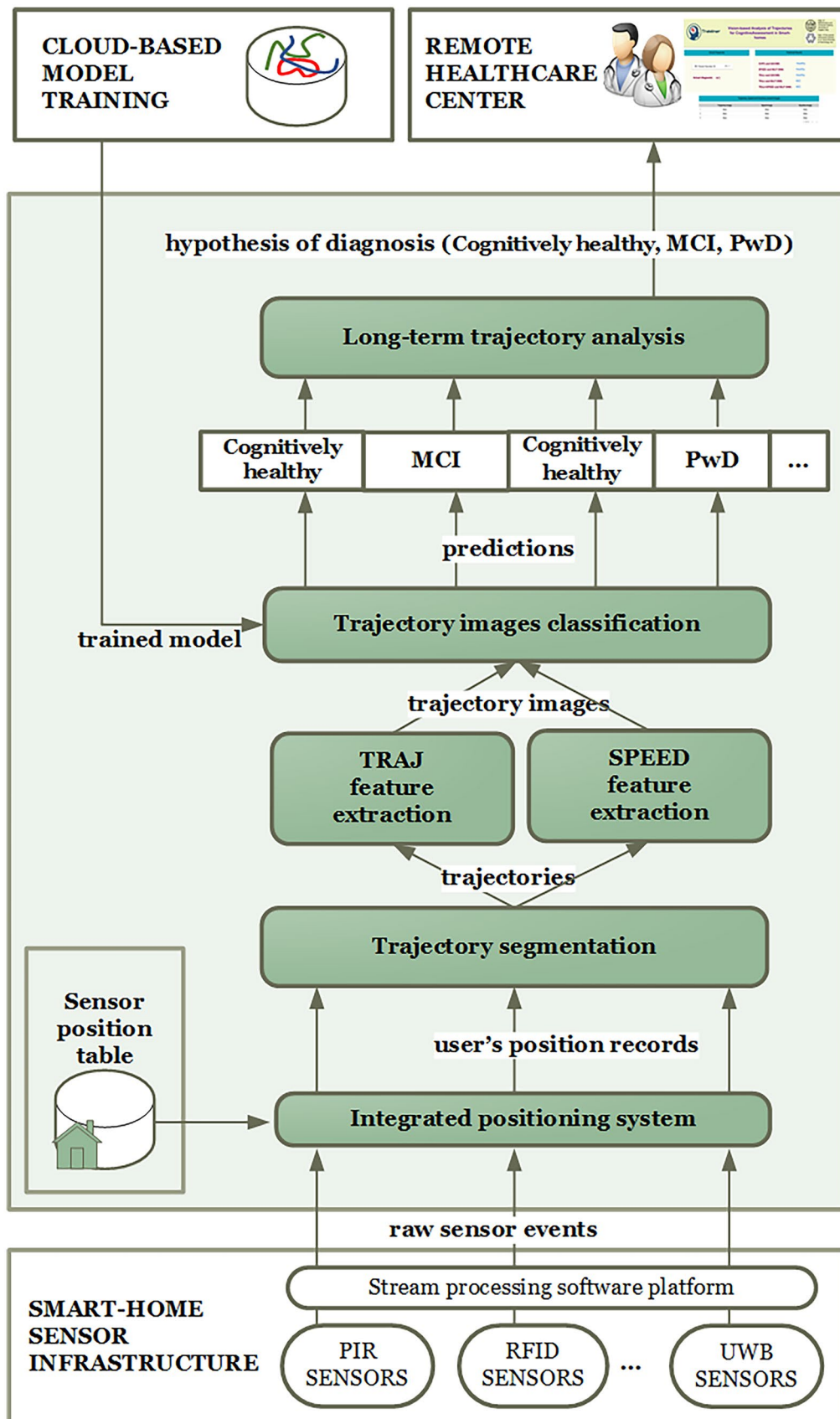


Fig. 2 Overview of the TraMiner system

consecutive records $\langle r_i, r_{i+1} \rangle \in H$ exceeds a threshold T_d (which is set to $5m$ in this work), we remove r_{i+1} from H .

After position data cleaning, the TRAJECTORY SEGMENTATION module is in charge of partitioning the denoised H into **trajectories**. In this regard, the position history H is partitioned into a set T of non-overlapping trajectories:

$$T = \{t_1, t_2, \dots, t_m\},$$

where each trajectory $t \in T$ is a temporal sequence of consecutive position records:

$$t = \langle p_j, p_{j+1}, \dots, p_k \rangle \in H.$$

A trajectory consists of locomotion and non-locomotion phases. The non-locomotion phases happen during a time interval between any two consecutive sensor activations that do not exceed a given threshold T_s . In our experiments, we evaluate different values of T_s ranging from 30s to 180s. The rationale of our segmentation algorithm is that a phase of non-locomotion may happen in a trajectory if no sensor is triggered for less than T_s . Otherwise, if no user movement is detected for more than T_s , the previous trajectory is completed, and a new one is initialized.

Our segmentation algorithm initializes the first trajectory with the first entry of H , and each trajectory t_i is a temporal sequence of consecutive position records from H such that the time interval between any two consecutive position records does not exceed T_s . The next trajectory is initialized with the last entry of the previous trajectory, and continues until the threshold condition is met. The algorithm continues until the end of H .

Visual Feature Extraction

In order to prepare input data for the DNN, the modules for TRAJ and SPEED FEATURE EXTRACTION are in charge of visually representing the salient features of each trajectory through images. To prepare the images, we use the RGB color model. Hence, each image has three colors: red, green, and blue. For every pixel, the values related to these colors are represented by a byte each. Hence, each pixel is characterized by three integer values ranging from 0 to 255. In our work, the maximum spatial extent of all the trajectories is the same, and corresponds to the home layout.

Choosing the image size is an important aspect for feature extraction. By increasing the resolution of the image, the sparsity of non-zero pixels in the image increases, and may decrease the learning capability of the DNN [36]. Low-resolution images can alleviate this issue; however, important information in the image could be lost, since the values of multiple contiguous pixels could be mixed. After considering different resolutions for the images,

for our experimental setup, we chose image length and height of 100 by 130 pixels. By considering the size of our smart-home test bed, each pixel in the image correspond to approximately $0.1m^2$.

TRAJ Feature Extraction

Figure 3 shows an example of image obtained using the TRAJ feature extraction method from a person's trajectory in our smart-home test bed.

The trajectory path is shown through a monochrome line string. The weight of each line is set according to the number of times that the corresponding path has been walked. The line weight is 1 if the path has been walked only once; the weight is 2 if it has been walked twice, and so on. Consequently, lines corresponding to paths walked several times are bolder than those of paths walked sporadically. This feature extraction method allows us to visually capture locomotion anomalies corresponding to the pacing and lapping patterns defined by the Martino–Saltzman model.

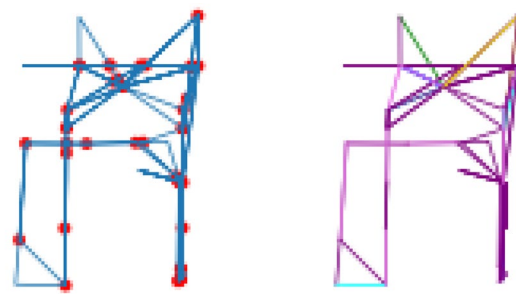
For emphasizing the intricacy of the trajectory, which may indicate random walk according to the Martino–Saltzman model, the intersection points between the trajectory lines are shown in the image as a red circle. By considering the image size, the radius of that circle corresponds to $0.25m$.

Moreover, the trajectory line string naturally encodes low-level motion indicators such as sharp angles, straightness, turning angle, and path efficiency.

SPEED Feature Extraction

As explained in Section 2.1.2, patterns of locomotion speed are important indicators of cognitive decline. However, speed information is not captured by the TRAJ feature extraction method.

Hence, TraMiner produces a second image of the trajectory, using the SPEED feature extraction method. Figure 3 shows an image obtained using the SPEED method from a person's trajectory in our experimental



(a) TRAJ feature extraction (b) SPEED feature extraction

Fig. 3 Example of visual feature extraction from a trajectory

test bed. With this method, the trajectory is partitioned in sections, such as the speed of the inhabitant is steady in each section. The color of each section depends on its speed range. When a section is walked multiple times, the most recent speed of that section is considered for drawing the image. For the sake of this work, we consider seven different speed ranges, with 7 associated colors. In particular, the speed ranges are 0-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-14, and more than 15 m/s, and the corresponding colors are purple, violet, blue, cyan, green, yellow, orange, and red, respectively. Of course, some of those speed ranges are exaggerated when referred to people moving inside an apartment. The explanation is that the positioning infrastructure used in our experimental test bed has approximately 1 meter resolution, and (being based on passive infrared technology) the detection range of sensors partially overlaps. For this reason, the computed speed is inevitably approximated and subject to errors. However, as confirmed by our experimental results, those approximated speed values are useful to improve the recognition performance of our system. The SPEED feature extraction method allows to capture the jerk low-level motion indicator.

DNN Trajectory Classification and Long-Term Analysis

In the following of this section, we present the DNN trajectory image classification method and the long-term analysis module.

Cloud-Based Model Training

The goal of the CLOUD-BASED MODEL TRAINING module is to train a DNN model to classify every trajectory as either walked by a cognitively healthy subject, by a person with MCI, or by PwD. To this aim, we take a collaborative approach. That module periodically receives a training set of trajectory images from the local instances of the TraMiner system, running in the individual homes. Each image is labeled with the cognitive status of the anonymous inhabitant (i.e., either 'cognitively healthy', 'MCI', or 'PwD'). Trajectory images are locally computed on the edge by the different instances of TraMiner, as explained in Section 4. Based on that training set, the trusted cloud-based module is in charge of training a DNN model for trajectory classification. The TraMiner local instances receive the model from the cloud, and use it for classifying the trajectories of the inhabitant. Note that the TraMiner instances do not receive any trajectory image. Indeed, for the sake

of privacy, trajectory images are processed only by the trusted cloud module.

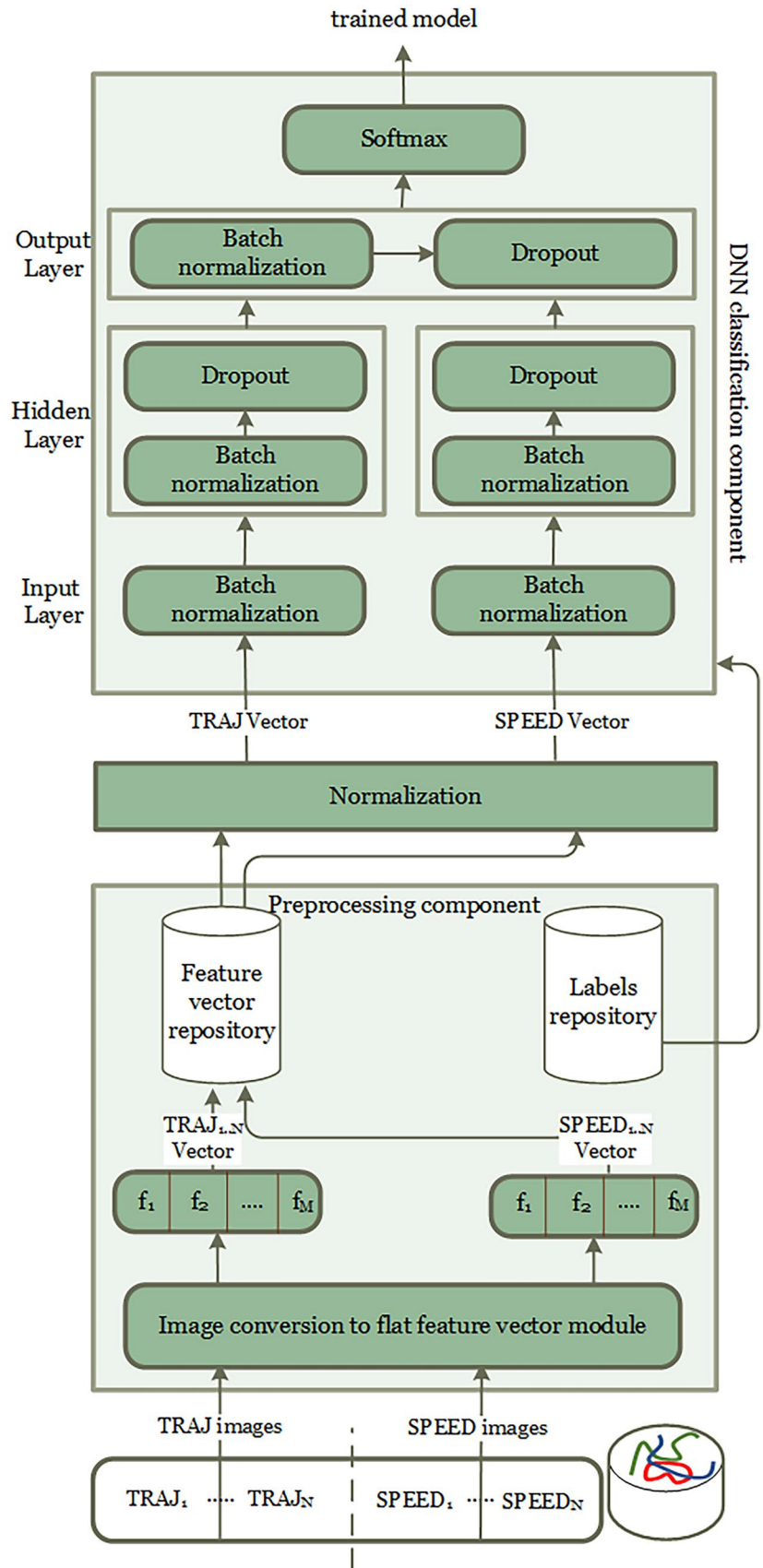
Figure 4 shows the multilayer perceptron (MLP) DNN architecture used by the cloud-based infrastructure. For each trajectory, the MLP DNN takes as input two images. The first one is obtained through the TRAJ feature extraction method as shown in Fig. 3, while the second one is obtained through SPEED feature extraction as shown in Fig. 3. Since we use two images with different features for each trajectory, our model relies on 'mixed data'. Developing systems capable of handling mixed data is still an open area of research, and can be challenging since each input with different feature representation may require separate preprocessing steps. In order to prepare the trajectories images for training, we convert them to flat feature vectors based on their width, height, and color values of each pixel. Hence, since images have 100 by 130 pixels, and each pixel is represented by three color values, each feature vector has size 39,000. So, each feature corresponds to the color value of a single pixel. Then, we apply binarization to the color values of the image to reduce the computational cost of training. Finally, we add the labels to the feature vectors before feeding them to the DNN for training.

The used MLP DNN for each input is composed of two fully connected (dense) layers: one input layer and one hidden layer, with 32 neurons for both. The layers used the Rectified Linear Units function (ReLU) as the activation function. Furthermore, for both inputs each layer is followed by batch normalization to speed up learning and increase the stability of the neural network. In the last layer for each of them, there is also a dropout layer in order to significantly reduce over-fitting. The fraction of drop units has been set to 0.5. Then, we combine the output of both inputs (TRAJ and SPEED) and apply one more fully connected layer with three neurons followed by batch normalization and dropout with same drop units. Since we are facing a three-class classification problem, we have chosen the softmax function as activation function on the final output layer. This function is used to find a probability distribution of the mentioned categories as follows:

$$p(k) = \frac{g_s}{\sum_{j=1}^{N_c} g_j}, g_i = \max(0, \sum_i f_i \cdot w_{ij} + h_j)$$

where $p(k)$ is the probability of a trajectory to belong to the k -th class, g_s is a standard exponential function applied to each element of the input vector. This function gives a positive value above zero, which will be very large if the input was large and very small if the input was negative. Furthermore, N_c denotes the number of classes, f_i is a value of i th neuron in the last fully connected layer, w_{ij} and h_j are coefficients of the softmax function [37].

Fig. 4 Cloud-based model training MLP DNN architecture



Furthermore, we have used a low learning rate, set to 0.00001, with adam optimizer. In this field, this solution achieved major improvements over other regularization methods [39, 40]. As a loss function, we have used category cross-entropy, which is an effective loss function for classification problems which have softmax activation function in the output layer. This loss function is based on the maximum likelihood estimate approach and it is used for single label categorization [41]. In our case, this ensures that a trajectory belongs to exactly one class.

We have used a batch size approach to calculate the model error and to update the model coefficients. We divided the training dataset into small batches of 64 samples which are utilized in one epoch. In order to fine-tune the number of epochs, we have used the *early stopping* procedure. We tested different numbers of epochs, ranging from 20 to 100 epochs. According to this procedure, the training is stopped when the generalization error increases. Therefore, we evaluated the model during training on a holdout validation set after each epoch. When the performance of the model starts to degrade (i.e., the loss begins to increase), the training process is stopped. In the presented configuration, the number of epochs was set to 27. With this approach, we improved the computational efficiency of the learning process, without the need of keeping all training data in the main memory. The approach also provides robust convergence, avoiding local minima [42].

Long-Term Trajectory Analysis

As shown in Fig. 2, the trained model is communicated to the TRAJECTORY IMAGES CLASSIFICATION module of the individual smart-home systems, which uses it for trajectory image classification as soon as new trajectories are observed in the home. Finally, the classification predictions are processed by the LONG-TERM TRAJECTORY ANALYSIS module, which is in charge of computing a hypothesis of diagnosis regarding the inhabitant's cognitive health (i.e., cognitively healthy, MCI, or PwD).

The aim of that module is to generate a hypothesis of diagnosis based on the long-term predictions of the vision-based trajectory classifier. Of course, in order to produce a reliable hypothesis, the long-term analysis module needs a sufficient quantity of trajectory classifications. Therefore, it is assumed that the module considers the whole data acquired in a given time period; e.g., all the predictions about the trajectories observed in the previous 30 days. The module analyses the whole history *VTC* of vision-based trajectory classifications produced in that period:

$$VTC = \{class_1, class_2, \dots, class_m\},$$

where the value of $class_i$ can be either '*cognitively healthy*', '*MCI*', or '*PwD*'. Also, there is a need to compute the

number of predictions of each class in *VTC*. Then, the module outputs the hypothesis corresponding to the most frequent class fc as follows:

$$fc = \arg \max_{class \in D} |\{cl \in VTC : cl = class\}|,$$

where $D = \{\text{'cognitively healthy'}, \text{'MCI'}, \text{'PwD'}\}$.

Experimental Evaluation

In this section, we report our experimental evaluation which was carried out with real-world locomotion data acquired in an instrumented smart-home from a large set of seniors, including cognitively healthy seniors, PwD, and persons with MCI. Therefore, in the following subsections, we explain the used dataset, the experimental setup, and the achieved results. We also experimentally compare our technique with an existing state-of-the-art method.

Dataset

The experiments were carried out considering real-world trajectories acquired from 153 individuals in a smart-home of the CASAS test-bed [44] and annotated by the researchers of the Center for Advanced Studies in Adaptive Systems (CASAS) at Washington State University (WSU) [22]. The smart-home layout is represented in Fig. 5. The smart-home is a two-story apartment equipped with different kinds of sensors, and its floor plan includes a living/dining room, three bedrooms, a kitchen, and a bathroom. For our experiments, we relied on passive infrared (PIR) motion sensors and door sensors to track the movements of the individuals in the home. PIR sensors are mounted on the ceiling and their accuracy is about one meter. In total, the apartment included 51 motion sensors and 16 door sensors.

Participants were recruited by advertisement and physician referrals. After obtaining informed consent, participants underwent multidimensional clinical assessment by neuropsychologists, in order to assess the cognitive health status. For the sake of anonymity, explicit identifiers of the individuals involved in the study were removed, and quasi-identifier personal data, such as age, were generalized to age ranges. The protocol of recruiting and data collection was approved by the Institutional Review Board of WSU [22]. As a consequence of clinical examination, each participant was classified as either PwD, person with MCI, or cognitively healthy person.

Participants were asked to individually execute scripted Day-Out Tasks (DOTs) in the smart-home test-bed. DOTs are naturalistic tasks involving the execution of interleaved activities for reaching a certain goal [45]. Each participant executed the DOTs in a single day for an average of three hours. Data collection occurred in the morning or in the

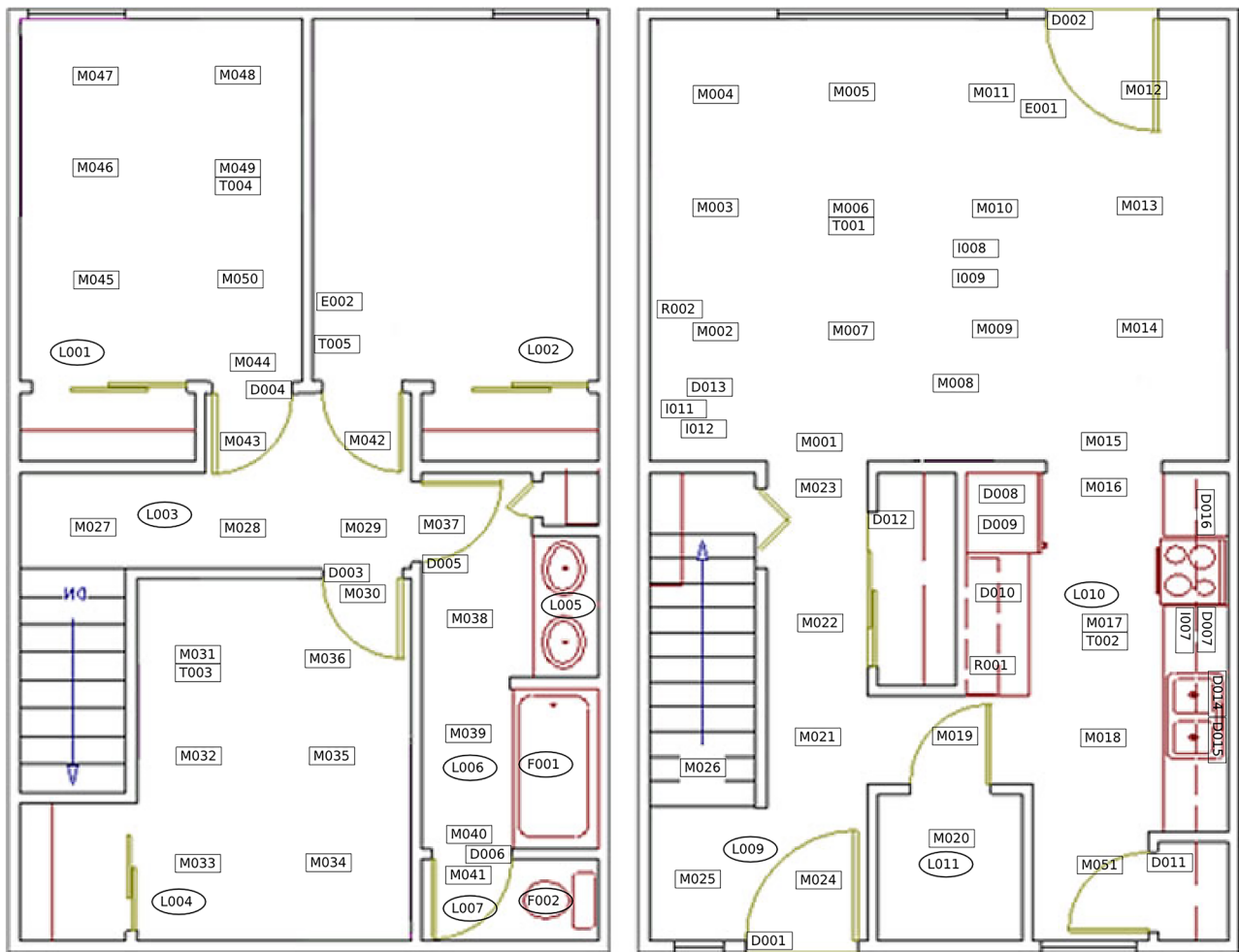


Fig. 5 The smart-home layout used in our experiments [43]

early afternoon. During the execution of DOTs, the sensor infrastructure acquired the sensor data triggered by their actions and movements. The detailed descriptions of DOTs, together with the collected dataset, are available on the Web¹. The smart-home setup is described in detail in [46].

While 40 PwD were recruited, only part of them were able to participate to the data collection in the smart-home. Hence, in our experiments we considered only PwD who were able to carry out activities in the home (19 individuals). In our experiments, we considered also the data acquired from the 80 seniors aged 60 to 74 years old, and from the 54 persons with MCI.

Comparison with State-of-the-Art Methods

In order to experimentally compare our method with the state of the art, we implemented both a baseline numeric feature extraction method, and the image-based method proposed by Gochoo et al. in [37].

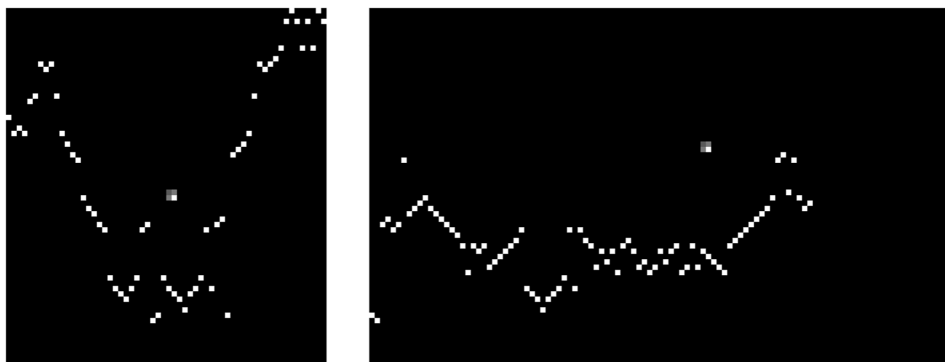
¹ <http://casas.wsu.edu/datasets/assessmentdata.zip>

State-of-the-Art Numeric Feature Extraction (NFE)

As a comparison, we implemented a baseline feature extraction method, in which each feature corresponds to a locomotion-based clinical indicator of cognitive decline proposed in the literature. We call this method *numeric feature extraction* (NFE). We consider the following indicators:

- *Pacing* [13] travel pattern, defined in the Martino-Saltzman model: this feature counts the number of observations of this pattern in the last day.
- *Lapping* [13] travel pattern, defined in the Martino-Saltzman model: This feature counts the number of observations of this pattern in the last day.
- *Random* [13] travel pattern, defined in the Martino-Saltzman model: This feature counts the number of observations of this pattern in the last day.
- *Jerk* [15] is computed as the first time derivative of acceleration. This features represents the average jerk observed in the individual's trajectories in the last day.

Fig. 6 Example of images generated through the GVFE feature extraction method. The left-hand side image is extracted from a trajectory using $T_s = 90\text{ s}$. The right-hand side image is extracted from a trajectory using $T_s = 150\text{ s}$



- *Straightness* [16] represents the average straightness computed on the individual's trajectories of the last day.
- *Sharp angles* [6] feature counts the number of sharp angles observed in the individual's trajectories during the last day.

In order to compute the above-mentioned features, we adopt the algorithms recently presented in [47].

State-of-the-Art Visual Feature Extraction (GVFE)

In the original paper where the method proposed by Gochoo et al. was presented [37], the GVFE technique was used to recognize those abnormal locomotion patterns that are strong indicators of neurocognitive diseases according to the Martino–Saltzman model; i.e., pacing, lapping, and random patterns. The authors experimented their GVFE technique in the same smart-home environment used in our experiments, with data acquired from a cognitively healthy senior during 21 months. They obtained very good results, achieving accuracy above 97%. Those results indicate that the GVFE technique is effective in recognizing abnormal travel patterns that indicate cognitive impairment. For this reason, we experimentally compare the GVFE technique with our visual feature extraction method.

In the GVFE technique, images in each trajectory are prepared based on the sequence of activation of the position sensors. Each trajectory is converted to a binary image, where the x axis represents the temporal order of the sensor activation, and the y axis represents the numeric identifier of the fired sensors. For example, suppose that the temporal sequence of activated motion sensors in a trajectory is: M005, M003, M005, M010, M011, M001. Then, the only non-zero pixels in the image correspond to the following coordinates: (1,5), (2,3), (3,5), (4,10), (5,11), (6,1).

Also for this feature extraction method, we apply trajectory segmentation using different values of the threshold T_s . The width of the corresponding images is chosen based on the maximum length of trajectories. For example, in our dataset, using $T_s = 60\text{ s}$, the maximum

length of trajectory for all considered individuals is 32. Hence, images computed using that threshold value have 32 pixels width.

As mentioned before, the y axis represents the numerical identifier of the fired sensor. Since in our dataset we consider 67 sensors (i.e., 51 PIR sensors and 16 door sensors), the y axis includes 67 different values. Consequently, the images have 67 pixels height, irrespectively from the value of the threshold T_s . For thresholds T_s set to 30 s, 60 s, 90 s, 120 s, and 150 s, the image size is (14 × 67), (32 × 67), (60 × 67), (110 × 67) and (169 × 67), respectively. Figure 6 shows two samples of images obtained using different threshold values.

State-of-the-Art DNN (DCNN)

In order to compare our DNN architecture with a state-of-the-art one, we consider the Deep Convolutional NN (DCNN) used by Gochoo et al. in [37]. As illustrated in Fig. 7, that network has three zero padding convolution layers and three fully connected layers which are followed by max-pooling layers and feature filters size of 5 × 5. The pooling window size is set to 22 and, since max-pooling creates smaller version of input maps, the output images become two times smaller than the input. The first convolution layer has 32 kernels, the second one, which receives the output of the first max-pooling layer as inputs, has 128 feature filters, and the last convolutional layer receives the second max pooling layer output and convolutes them with 256 feature filters.

Finally, in the fully connected part, the layers are flattened, and the output of the third max-pooling layer is converted into a feature vector. In this part, the first, second, and third fully connected layers have 512, 128 and 64 neurons, respectively, and neurons of the last fully connected layer are connected to all three outputs; i.e., cognitively healthy, MCI, PwD. In order to find probability the distribution of the classes, the softmax function is applied. Since the reference paper [37] does not mention

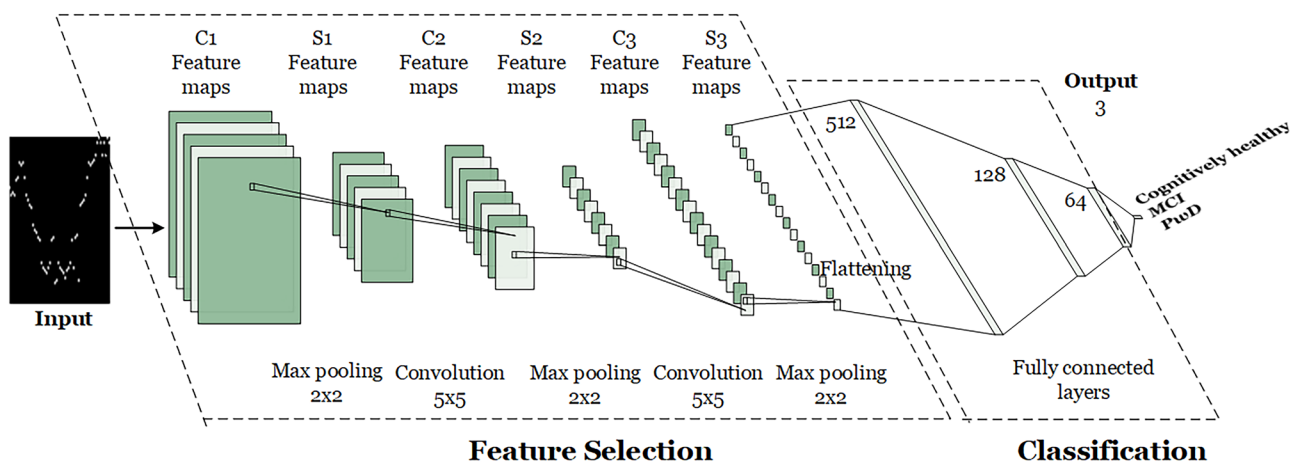


Fig. 7 The state-of-the-art deep convolutional neural network (DCNN) used by Gochoo et al. in [37]

internal details such as the used optimizer and its rate, or the chosen loss function, we used the same parameters chosen for our proposed DNN configuration described in Section 5.

Experimental Setup

We developed all the algorithms in Python. The code is available on the Web². For experimenting the NFE technique, we used the machine learning algorithms implemented by the Weka toolkit [48]. The code for extracting the NFE features is available online³. We have used the Python Keras neural network library⁴ to develop the proposed DNN

classification systems. In order to support scalability, in the general architecture of our system we envision the use of a cloud-based system for training the DNN model. However, given the relatively small size of the training set used in our experiments, we trained the DNN on a departmental server. We have run experiments on a Linux server with four NVIDIA Tesla p6 graphic boards, a single NVIDIA Pascal GP104 graphics processing unit (GPU), and 16 GB GDDR5 memory. To evaluate the effectiveness of our TRAJ and SPEED visual feature extraction techniques, we have experimented with different values of the T_s threshold for trajectory segmentation, ranging from 30 seconds up to 180

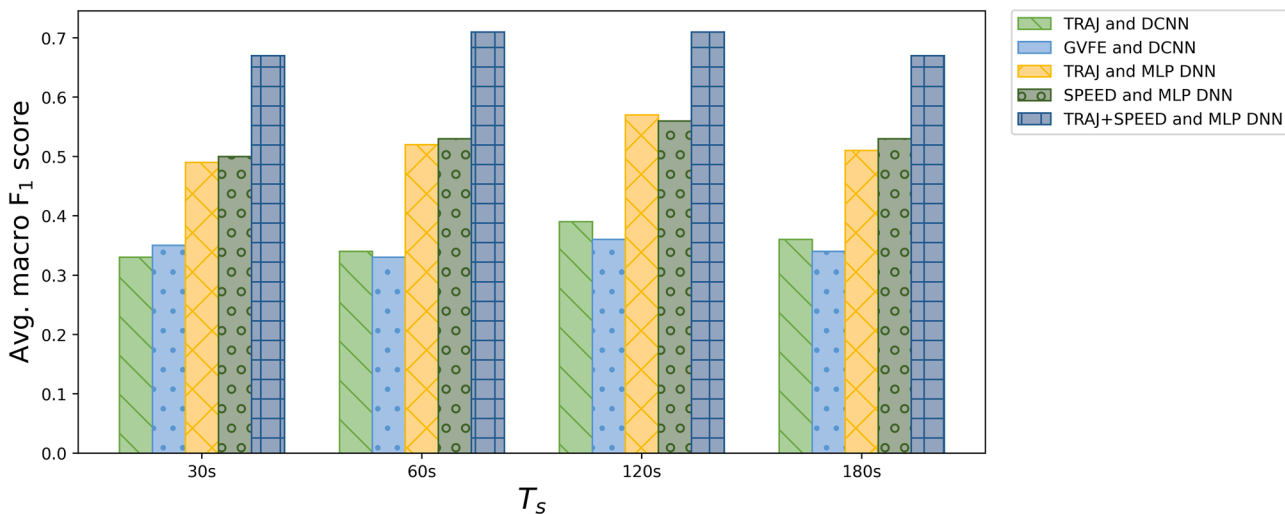


Fig. 8 Trajectory images classification: macro- F_1 score for the different techniques

² <https://sites.unica.it/domusafe/traminer/>

³ <https://sites.unica.it/domusafe/healthxai/>

⁴ <https://keras.io/>

Table 1 Results of numeric feature extraction (NFE) method

Class	Measure	Naive Bayes	Logistic regr.	MLP	SVM		
Cognitively healthy	F ₁ score	0.50	0.69	0.65	0.69		
	Precision	0.56	0.55	0.53	0.53		
	Recall	0.45	0.91	0.84	1.00		
MCI	F ₁ score	0.46	0.32	0.25	0.04		
	Precision	0.38	0.57	0.39	1.00		
	Recall	0.59	0.22	0.19	0.02		
PwD	F ₁ score	0.00	n/a	n/a	n/a		
	Precision	0.00	n/a	n/a	n/a		
	Recall	0.00	0.00	0.00	0.00		
Avg.	F ₁ score	0.32	n/a	n/a	n/a		
	Precision	0.31	n/a	n/a	n/a		
	Recall	0.35	0.38	0.34	0.34		
Class	Measure	kNN	Ripper	C4.5	Rand. tree	RF	
Cognitively healthy	F ₁ score	0.61	0.68	0.59	0.54	0.64	
	Precision	0.55	0.53	0.47	0.55	0.56	
	Recall	0.68	0.93	0.79	0.54	0.74	
MCI	F ₁ score	0.43	0.21	0.00	0.34	0.39	
	Precision	0.46	0.50	0.00	0.35	0.44	
	Recall	0.41	0.13	0.00	0.33	0.35	
PwD	F ₁ score	0.00	n/a	0.00	0.00	0.08	
	Precision	0.00	n/a	0.00	0.00	0.20	
	Recall	0.00	0.00	0.00	0.00	0.05	
Avg.	F ₁ score	0.35	n/a	0.20	0.29	0.37	
	Precision	0.34	n/a	0.16	0.30	0.40	
	Recall	0.36	0.35	0.26	0.29	0.38	

seconds. We also developed the feature extraction method and the DNN used by Gochoo et al. [37] to experimentally compare our methods with the state of the art.

In all the experiments, we applied a *leave one person out* cross-validation approach: we used the data of one individual for the test set, and the data of the other persons for training and validation, iterating on each

person to execute the tests on the whole considered participants. With this approach, the data of the same person is never used both for training/validation and test at the same time. For tuning the hyper-parameters of the DNN, training trajectories are split by a fraction of 10% of each category for validation and 90% of them for training.

Table 2 Trajectory images classification: results obtained using our TRAJ feature extraction method and DCNN vs GVFE and DCNN

Class	Measure	TRAJ and DCNN				GVFE and DCNN			
		30s	60s	120s	180	30s	60s	120s	180s
Cognitively healthy	F ₁ score	0.54	0.52	0.57	0.57	0.57	0.57	0.56	0.55
	Precision	0.57	0.53	0.62	0.6	0.62	0.62	0.6	0.57
	Recall	0.51	0.5	0.52	0.53	0.53	0.52	0.53	0.54
MCI	F ₁ score	0.34	0.36	0.5	0.38	0.36	0.36	0.4	0.35
	Precision	0.33	0.36	0.4	0.37	0.34	0.34	0.39	0.35
	Recall	0.35	0.37	0.41	0.39	0.38	0.38	0.41	0.36
PwD	F ₁ score	0.12	0.13	0.1	0.12	0.11	0.056	0.13	0.11
	Precision	0.1	0.11	0.08	0.08	0.09	0.04	0.1	0.09
	Recall	0.14	0.15	0.13	0.18	0.15	0.081	0.17	0.13
Avg.	F ₁ score	0.33	0.34	0.39	0.36	0.35	0.33	0.36	0.34
	Precision	0.33	0.33	0.37	0.35	0.35	0.33	0.36	0.34
	Recall	0.33	0.34	0.35	0.37	0.35	0.33	0.37	0.34

Table 3 Trajectory images classification: results of our TRAJ vs SPEED feature extraction methods, using our MLP DNN

Class	Measure	TRAJ and MLP DNN				SPEED and MLP DNN			
		30s	60s	120s	180	30s	60s	120s	180s
Cognitively healthy	F ₁ score	0.62	0.67	0.67	0.64	0.62	0.66	0.72	0.67
	Precision	0.56	0.65	0.62	0.59	0.56	0.66	0.7	0.64
	Recall	0.68	0.69	0.72	0.69	0.68	0.65	0.74	0.7
MCI	F ₁ score	0.5	0.55	0.58	0.56	0.53	0.52	0.59	0.57
	Precision	0.46	0.5	0.52	0.52	0.51	0.46	0.55	0.56
	Recall	0.56	0.62	0.65	0.6	0.56	0.61	0.64	0.59
PwD	F ₁ score	0.35	0.34	0.47	0.33	0.37	0.42	0.39	0.35
	Precision	0.58	0.49	0.75	0.54	0.56	0.56	0.51	0.45
	Recall	0.25	0.27	0.34	0.24	0.28	0.34	0.31	0.28
Avg.	F ₁ score	0.49	0.52	0.57	0.51	0.50	0.53	0.56	0.53
	Precision	0.53	0.54	0.63	0.55	0.54	0.56	0.58	0.55
	Recall	0.49	0.46	0.57	0.51	0.51	0.53	0.56	0.52

For the overall performance evaluation, we have used the metrics of macro-precision, macro-recall, and macro- F_1 score. These metrics are standard ones for imbalanced problems. In particular, macro- F_1 score is a reliable metric in imbalanced cases, since it gives equal weights to the different classes, despite their size. Since the 'accuracy' performance measure is inappropriate for imbalanced classification problems like the one we are tackling, we do not consider that measure in our evaluation. Indeed, depending on the degree of imbalance, the majority class accuracy value would overcome the accuracy value of the minority classes.

Results of NFE Technique

At first, we evaluated the numeric feature extraction technique explained in Section 6.2.1. For each participant, we built the feature vector using all the trajectory data collected during the day. Each feature vector was labeled with the cognitive status of the individual; i.e., cognitively healthy, MCI, or PwD. We experimented several classifiers:

- The well-known *Naive Bayes* [49] classifier;
- *Logistic regression* classifier [50], relying on a multinomial logistic regression model with a ridge estimator;
- *Multilayer perceptron (MLP)* feed-forward artificial neural network algorithm;
- *Support Vector Machines* [51]
- k Nearest Neighbours (kNN) [52] lazy classifier, with $k = 5$;
- *Ripper* [53] propositional rule learner;
- *C4.5* [54] decision tree;
- *Random tree* [55] classifier.
- *Random forest (RF)* [56] classifier.

Results are reported in Table 1. Overall, the results achieved by the NFE technique are poor. Indeed, all classifiers achieved a macro-average F_1 score close to the one of a random classifier. The classifier achieving the best performance in this pool of experiments is the Random forest algorithm, with F_1 score of 0.37. These results seem to indicate that, in a home context, numerical statistics about locomotion-based indicators of cognitive decline are ineffective for automatic cognitive assessment. This fact is probably due to the high level of noise introduced by two factors that influence the movement patterns; i.e., obstacles in the home, and execution of daily living activities.

Table 4 Trajectory images classification: results of our TRAJ+SPEED model and MLP DNN

Class	Measure	TRAJ+SPEED and MLP DNN			
		30s	60s	120	180s
Cognitively healthy	F ₁ score	0.79	0.82	0.8	0.79
	Precision	0.73	0.81	0.76	0.74
	Recall	0.86	0.84	0.85	0.84
MCI	F ₁ score	0.74	0.74	0.76	0.72
	Precision	0.67	0.64	0.68	0.69
	Recall	0.84	0.86	0.86	0.78
PwD	F ₁ score	0.49	0.56	0.58	0.52
	Precision	0.81	0.82	0.89	0.39
	Recall	0.35	0.43	0.43	0.77
Avg.	F ₁ score	0.67	0.71	0.71	0.67
	Precision	0.73	0.75	0.77	0.61
	Recall	0.68	0.71	0.71	0.79

Results of Single Trajectory Image Classification

At first, we experimentally compare the effectiveness of our TRAJ feature extraction method with the one achieved using the Gochoo's et al. visual feature extraction method (named GVFE). For this experiment, we used the DCNN used by Gochoo's et al., named DCNN. In this regard, we want to assess the ability of correctly recognizing the cognitive status of a person by the observation of a single trajectory walked by him/her.

According to the results shown in Table 2, for both techniques, the best results are achieved with the trajectories obtained setting $T_s = 120s$. As it can be observed in Fig. 8, the technique using TRAJ and DCNN slightly outperforms the one relying on GVFE and DCNN in terms of macro- F_1 score.

In a second set of experiments, we compare the performance of the TRAJ vs SPEED feature extraction methods, using our MLP DNN. Table 3 shows the achieved results. The two feature extraction methods achieve

comparable results. As it can be observed in Fig. 8, for both methods, the best results in terms of average macro- F_1 score are obtained using $T_s = 120s$. It is evident that the results obtained using the TRAJ feature extraction method with our MLP DNN strongly improved with respect to using the DCNN. Indeed, using the MLP DNN we achieve a macro- F_1 score larger than 0.57 with less computation time, while the best macro- F_1 score obtained using the more complex DCNN was close to 0.36. We believe that this result may depend on the relatively limited size of the training set. Indeed, using a larger training set, the more complex DCNN could possibly outperform the MLP DNN, at the cost of additional time and resource consumption. Due to these results, in the rest of the experiments we use our MLP DNN for performing the classification tasks.

Results of Two Input Trajectory Images Classification

Since the results obtained using separately the TRAJ and SPEED techniques are encouraging, we perform additional

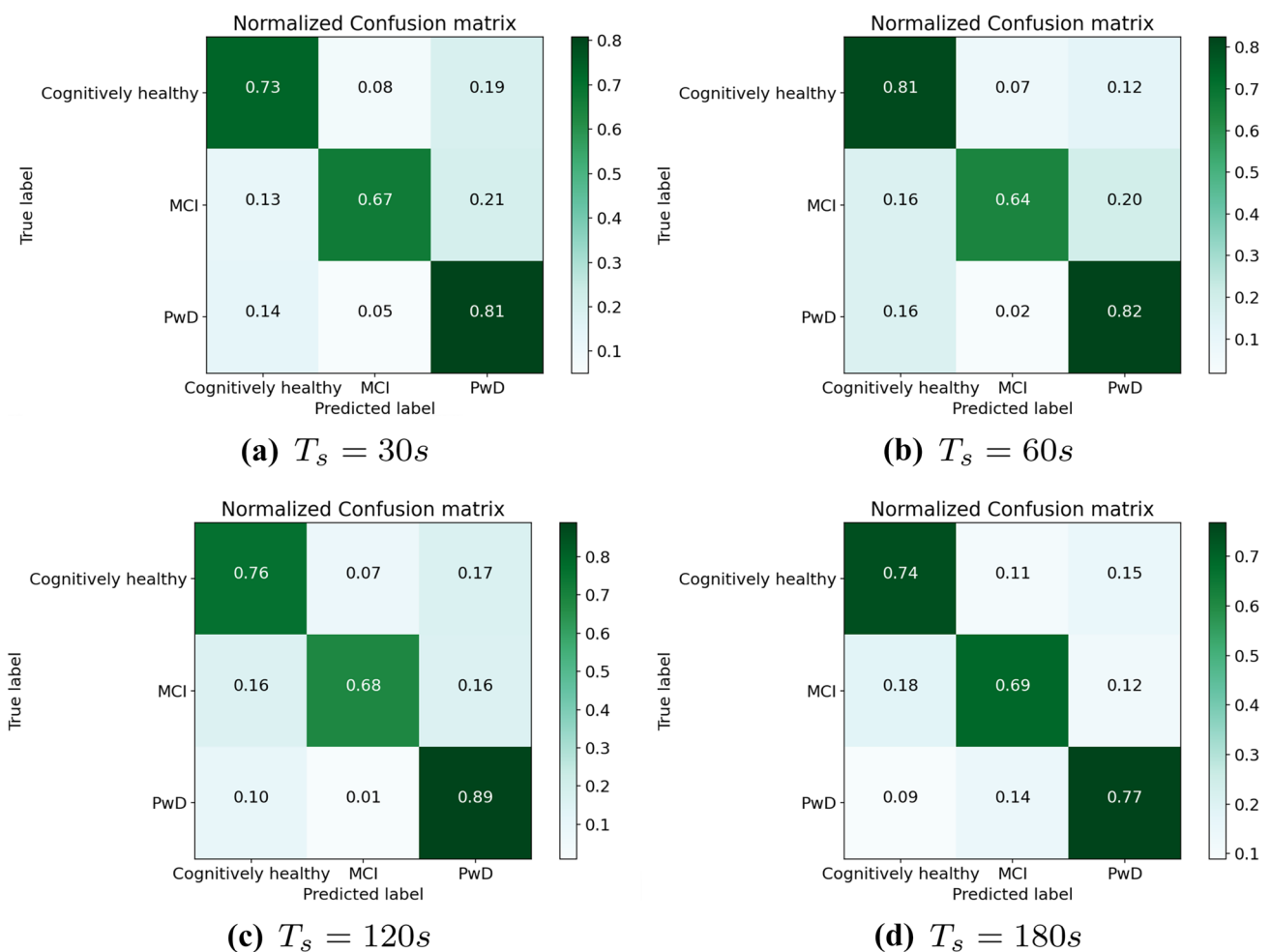


Fig. 9 Trajectory images classification: confusion matrices of our TRAJ+SPEED model and MLP DNN

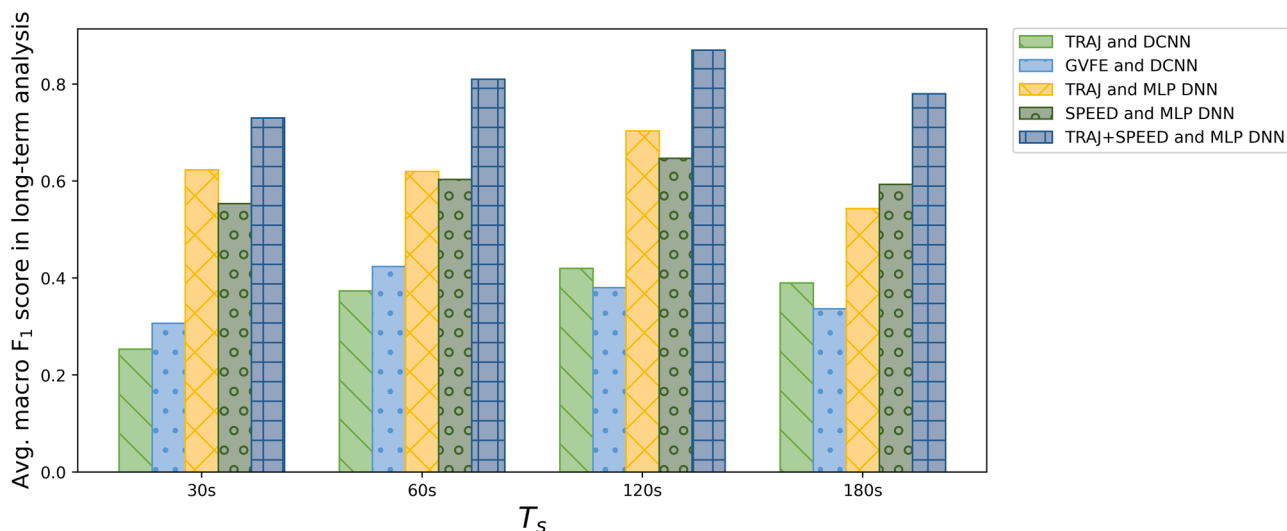


Fig. 10 Long-term analysis: macro- F_1 score for the different techniques

experiments using both TRAJ and SPEED trajectory images as input to our proposed MLP DNN. Indeed, since TRAJ and SPEED images represent different features of trajectories, their combined use may increase recognition rates. This setup corresponds to the TraMiner architecture shown in Figs. 2 and 4.

The achieved results are presented in Table 4. Like in the previous experiments, the best results in terms of macro- F_1 score are obtained using $T_s = 120s$. Also in this experiment, the worse results are obtained using threshold values of 30s and 180s. Considering these results, as also shown in Fig. 8, it is evident that the combined use of TRAJ and SPEED features significantly improves the recognition performance with respect to the use of the single feature extraction methods. Indeed, the best achieved F_1 score is larger than 0.71, while the single feature extraction methods achieve F_1 scores close to 0.57.

The detailed results can be inspected through the confusion matrices reported in Fig. 9. By observing the confusion matrices, it is evident that the total number of samples changes depending on the chosen value of T_s . For instance, with $T_s = 30s$ we have 5, 195 trajectories, while with $T_s = 180s$, we have only 540 trajectories. Indeed, in general, the lower the threshold for trajectory segmentation, the larger the number of generated trajectories. Hence, by using larger values of T_s we obtain a smaller number of samples, but we can encode more information in the single trajectory images, since trajectories are generally longer. On the negative side, too large values of T_s may determine very involved images, that may confuse the DNN. On the contrary, too small values of T_s may determine very short trajectories, that do not encode enough information for the

DNN. According to our experiments, the value $T_s = 120s$ provides a good trade-off in this sense.

However, the achieved results, which are computed on the classification of single trajectories in isolation, are not sufficient for providing a reliable hypothesis about the cognitive status of the individual. For this reason, in the following experiments we evaluate the performance of the module for LONG-TERM TRAJECTORY ANALYSIS, which considers the whole history of trajectories acquired during a certain period of time.

Table 5 Long-term analysis: results obtained using two input trajectory images classification (TRAJ + SPEED features) with our MLP DNN

Class	Measure	TRAJ+SPEED and MLP DNN			
		30s	60s	120	180s
Cognitively healthy	F_1 score	0.86	0.93	0.92	0.88
	Precision	0.85	0.94	0.95	0.84
	Recall	0.87	0.93	0.89	0.92
MCI	F_1 score	0.78	0.82	0.87	0.87
	Precision	0.69	0.74	0.8	0.83
	Recall	0.9	0.93	0.93	0.9
PwD	F_1 score	0.57	0.67	0.83	0.61
	Precision	0.79	0.84	0.89	0.79
	Recall	0.44	0.55	0.77	0.5
Avg.	F_1 score	0.73	0.81	0.87	0.78
	Precision	0.77	0.84	0.88	0.82
	Recall	0.73	0.80	0.86	0.77

Results of Long-Term Trajectory Analysis

In these experiments, we apply the algorithm for long-term analysis described in Section 5.2 with all the different techniques for trajectory image classification evaluated in Sections 6.5 and 6.6. Figure 10 provides an overview of the achieved results.

As expected, the results achieved using DCNN with the TRAJ or GVFE feature extraction methods are rather poor. Indeed, those techniques achieve the lowest recognition rates for trajectory image classification. In particular, the TRAJ method with DCNN achieves an F_1 score slightly larger than 0.4 with $T_s = 120s$. Its F_1 score is lower with the other values of the threshold. The best result of the GVFE method with DCNN is similar, but obtained with $T_s = 60s$. Overall, these two techniques do not provide significant results for cognitive assessment, even in the long term.

Results are significantly better using our MLP DNN with the TRAJ or SPEED feature extraction methods. For both techniques, the best results are obtained using $T_s = 120s$.

In particular, the TRAJ method achieves a macro- F_1 score of 0.7, while the SPEED method achieves 0.65 macro- F_1 score. With these techniques, the increase of recognition performance introduced by the long-term evaluation algorithm is evident. Indeed, both techniques obtain a lower macro- F_1 score for trajectory image classification, which is close to 0.57. However, the results obtained with these techniques are still insufficient for providing reliable hypothesis of diagnosis about cognitive assessment.

The best results in this pool of experiments are achieved using two input trajectory images classification (TRAJ + SPEED features) with our MLP DNN. Indeed, the best results are achieved with $T_s = 120s$, with a macro- F_1 score of 0.873. The detailed results are shown in Table 5. As it can be observed, the best results with $T_s = 120s$ are achieved for the class of cognitively healthy subjects (F_1 score = 0.92), which is the most frequent one. The class of MCI subjects obtains F_1 score = 0.87, while the class of PwD people achieves F_1 score = 0.83.

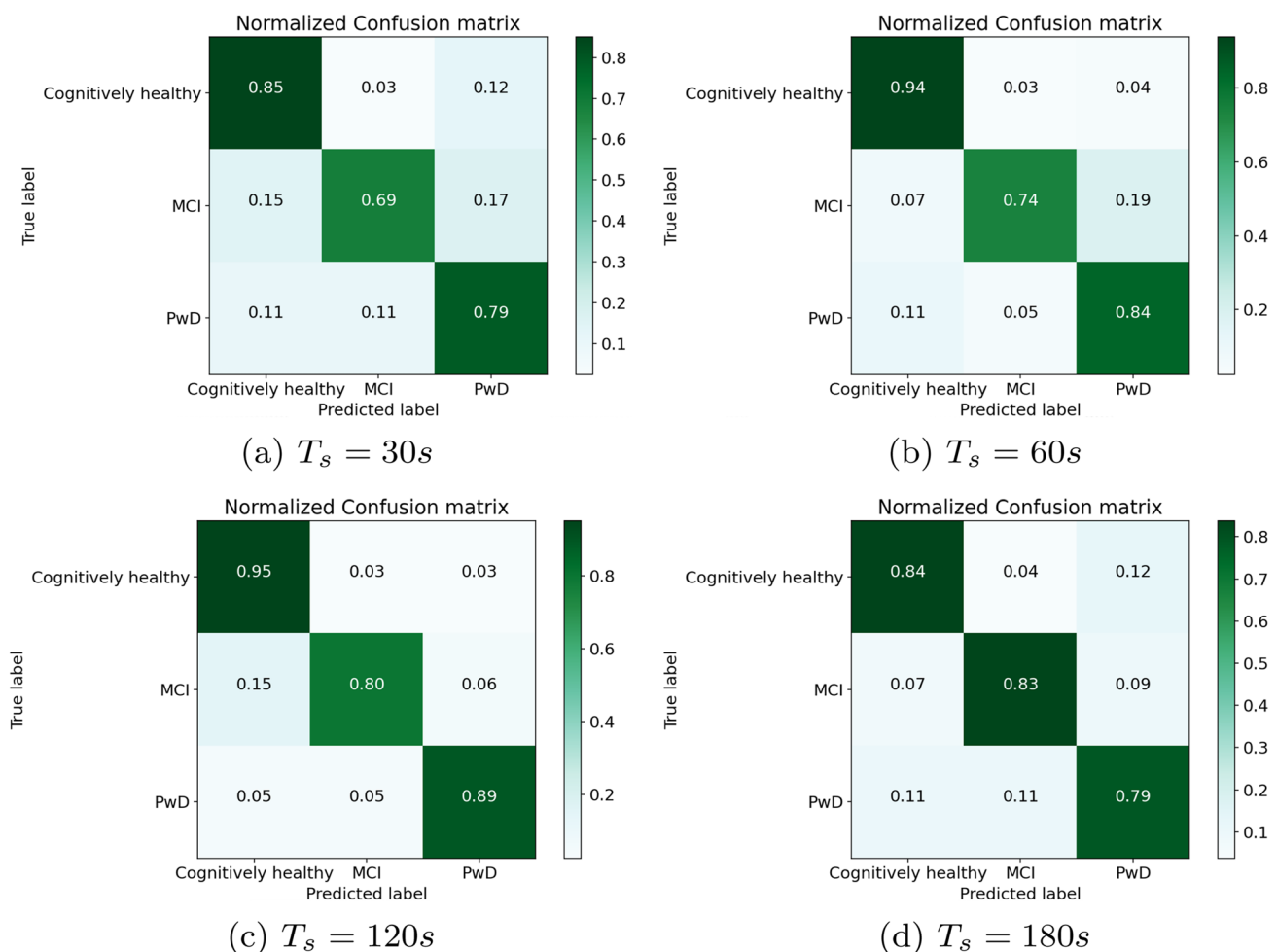


Fig. 11 Long-term analysis: confusion matrices obtained using two input trajectory images classification (TRAJ + SPEED features) with our MLP DNN

Fig. 12 A screenshot of the TraMiner dashboard

By closely inspecting the results in the confusion matrices shown in Fig. 11, we can observe that, with $T_s = 120s$, 17 PwD subjects out of 19 are correctly recognized. Among the other ones, one subject is classified as a person with MCI, and one as a cognitively healthy person. Hence, the false negative rate of PwD is very low. Regarding false-positive predictions of dementia, we observe that three persons with MCI out of 54 are classified as PwD. Out of 80 cognitively healthy subjects, only two are classified as PwD. Hence, the false positive rate of PwD is also low. Regarding the 54 persons with MCI, 43 are correctly recognized, while 8 are classified as cognitively healthy, and 3 of them as PwD. We consider these results positive, since MCI is an intermediate state between cognitive health and dementia, which is difficult to diagnose, especially with automatic tools. Among the 80 cognitively healthy seniors, we achieved only 4 false positives. Indeed, two of them were classified as persons with MCI, and two of them as PwD.

Dashboard for Clinicians

In order to allow clinicians inspecting the predictions of our system, we have developed a user-friendly dashboard, using the Google Data Studio framework. The dashboard allows inspecting the predictions obtained through the various

techniques experimented in our work, achieved using the threshold value $T_s = 120s$. The dashboard can be freely accessed on the Web⁵.

A screenshot of the dashboard is illustrated in Fig. 12. The user can select the patient through a drop-down list. The actual diagnosis for the current patient is shown in the left-hand side of the dashboard. On the right-hand side, the user can inspect the predicted diagnosis based on the five different methods: 'GVFE and DCNN', 'SPEED and MLP DNN', 'TRAJ and DCNN', 'TRAJ and MLP DNN', and 'TRAJ+SPEED and MLP DNN'. We remind that the latter is the actual method implemented by TraMiner, while the other ones are shown only as a reference.

By selecting a patient, the lower part of the dashboard shows the history of all input trajectories. For each trajectory, the user can visualize the extracted visual features, according to the three experimented methods: 'TRAJ', 'SPEED', and 'GVFE'. Those images are available in a table, and can be opened in a separate window through a hyperlink. A sample of three images in the dashboard for one trajectory is shown in Fig. 13.

⁵ <https://sites.unica.it/domusafe/traminer/>

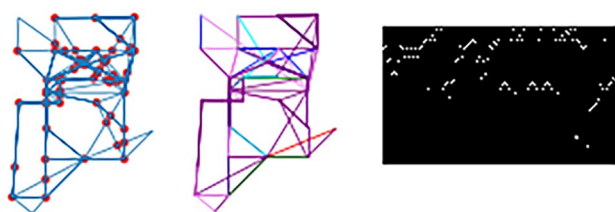


Fig. 13 A sample of images for one trajectory in the TraMiner dashboard. From left to right, the images are obtained using the TRAJ, SPEED, and GVFE feature extraction methods, respectively

Discussion and Limitations

Our experimental evaluation shows that the use of our combined visual features (TRAJ and SPEED), coupled with the use of the MLP DNN, outperforms a state-of-the-art method based on a different visual feature extraction technique and on a more complex DNN. The advantage of our solution consists in the encoding of additional features, such as speed, intersections, and low-level anomaly indicators, that are not captured by existing solutions relying on images. Indeed, those locomotion features are known in the literature to be reliable indicators of cognitive diseases.

In general, our system achieves better results with more frequent classes. Hence, we expect that recognition rates may improve using a training set composed of larger sets of individuals with MCI and PwD. The achieved macro- F_1 score suggests that TraMiner may be a useful support for the clinicians to provide a clinical evaluation of the cognitive health status of the elderly. However, this hypothesis should be confirmed by a large trial with the support of clinicians and deployment of our system in real-world conditions.

The experiments show that the recognition performance of TraMiner strongly improves by considering the whole history of trajectories. In this respect, we recall that, despite the dataset having been acquired from more than 150 individuals, each person was monitored only for few hours in one day. Such short observation period may be insufficient to reliably predict the cognitive status of all individuals. Hence, we expect to achieve more accurate predictions by considering a longer history of observations. However, this intuition needs to be verified by additional experiments on a large trial.

A limitation of our vision-based method is the difficulty of manually analyzing the reasons that determined the actual hypothesis of diagnosis by TraMiner. Indeed, it is not straightforward for a human observer to distinguish the trajectory images produced by our feature extraction techniques for the different classes of seniors. In order to provide explainable AI capabilities, TraMiner should be complemented with other methods for cognitive assessment, possibly considering other

behavioral models of cognitive decline based on overt [57] or subtle anomalies [58], which are easier to interpret by a domain expert. In order to recognize those behavioral anomalies, TraMiner should be extended with additional sensors and algorithms to recognize activities at a fine-grained level.

Given the nature of the dataset used in our experiments, all the patients' data were acquired in the same smart-home context. Whether the learned model is portable to a different context is still an open research question. We believe that advanced transfer learning methods specifically designed for image classification may enable training data portability [59], but this aspect should be confirmed by additional experiments with other datasets. Even without the use of transfer learning methods to enhance data portability, our system could be applied to important domains. In particular, a residence for elderly people may consist of several similar apartments. The DNN model could be trained based on the trajectories walked by those inhabitants for which a cognitive status diagnosis is known. That model could be used for the cognitive assessment of the other inhabitants.

Conclusion and Future Work

In this paper, we tackled the challenging issue of recognizing symptoms of cognitive decline based on the analysis of indoor movements. To this aim, we proposed a technique to extract visual features from indoor trajectories, and a two-input deep learning model for classification. Experiments with a real-world dataset collected from cognitively healthy seniors, people with MCI, and PwD, showed that our system achieves good accuracy for long-term cognitive assessment and outperforms state-of-the-art techniques.

Several directions remain open for future research. Visual feature extraction could be improved by adding some other useful characteristics related to low-level abnormal motion indicators. Our neural network model could be refined to more extensively exploit visual features. The optimal value of the temporal threshold T_s may depend on the kind of activity currently performed by the individual, and by the home shape. Hence, we will investigate techniques to fine-tune T_s considering the current context of the inhabitant. Finally, the system could be provided with explainable AI capabilities by adopting additional models of abnormal behaviors.

Acknowledgements The authors would like to thank the anonymous reviewers for their insightful comments and suggestions for improving the technical content and presentation of this paper. This work was partially funded by the POR FESR Sardegna 2014-2020 project "MISTER: Match Information System and Technologies for the Evaluation of the Performance" and by project ADAM (CUP F74I19000900007, Convenzione tra Fondazione di Sardegna e Atenei Sardi, annualità 2018).

Funding Open Access funding provided by Università degli Studi di Cagliari

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Ethical Standard All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bloom DE, Luca DL. The global demography of aging: facts, explanations, future. In *Handbook of the economics of population aging*. Elsevier. 2016. vol. 1, p. 3–56.
- Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. COVID-19: how doctors and healthcare systems are tackling coronavirus worldwide. *Br Med J*. 2020;368:
- Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary costs of dementia in the united states. *N Engl J Med*. 2013;368(14):1326–34.
- Schaat S, Koldrack P, Yordanova K, Kirste T, Teipel S. Real-time detection of spatial disorientation in persons with mild cognitive impairment and dementia. *Gerontology*. 2020;66(1):85–94.
- Ng J, Kong H. Not all who wander are lost: Smart tracker for people with dementia. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems 2016*. p. 2241–2248.
- Lin Q, Zhang D, Huang X, Ni H, Zhou X. Detecting wandering behavior based on GPS traces for elders with dementia. In *12th International Conference on Control Automation Robotics & Vision*. IEEE, 2012. p. 672–677.
- Khodabandehloo E, Riboni D. Collaborative trajectory mining in smart-homes to support early diagnosis of cognitive decline. *IEEE Transactions on Emerging Topics in Computing* 2020.
- Kearns WD, Fozard JL, Nams VO, Craighead JD. Wireless telesurveillance system for detecting dementia. *Gerontechnology*. 2011. p. 90.
- Kumar A, Lau CT, Ma M, Chan S, Kearns W. Trend analysis in the trajectory of the dementia patients. In *Proc. of ICSEC: IEEE*; 2017. p. 1–5.
- Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol*. 2005;62(7):1160–3.
- Zolfaghari S, Khodabandehloo E, Riboni D. Towards vision-based analysis of indoor trajectories for cognitive assessment. In *Proceedings of the 2020 IEEE International Conference on Smart Computing (SMARTCOMP) IEEE*, 2020. p. 290–295.
- Algase DL, Moore DH, Vandeweerd C, Gavin-Dreschnack D. Mapping the maze of terms and definitions in dementia-related wandering. *Aging Ment Health*. 2007;11(6):686–98.
- Martino-Saltzman D, Blasch BB, Morris RD, McNeal LW. Travel behavior of nursing home residents perceived as wanderers and nonwanderers. *The Gerontologist*. 1991;31(5):666–72.
- Sheridan PL, Solomont J, Kowall N, Hausdorff JM. Influence of executive function on locomotor function: divided attention increases gait variability in Alzheimer's disease. *J Am Geriatr Soc*. 2003;51(11):1633–7.
- Dabiri S, Heaslip K. Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation research part C: emerging technologies*. 2018;86:360–71.
- Li X. Using complexity measures of movement for automatically detecting movement types of unknown GPS trajectories. *Am J Geogr Inf Syst*. 2014;3(2):63–74.
- Wang Z, Xin J, Wang Z, Gu H, Zhao Y, Qian W. Computer-aided dementia diagnosis based on hierarchical extreme learning machine. *Cogn Comput*. 2020. p. 1–15.
- Castillo JC, Castro-González Á, Fernández-Caballero A, Latorre JM, Pastor JM, Fernández-Sotos A, Salichs MA. Software architecture for smart emotion recognition and regulation of the ageing adult. *Cogn Comput*. 2016;8(2):357–67.
- Rodríguez AC, Roda C, Montero F, González P, Navarro E. An interactive fuzzy inference system for teletherapy of older people. *Cogn Comput*. 2016;8(2):318–35.
- Pires P, Mendes L, Mendes J, Rodrigues R, Pereira A. Integrated e-healthcare system for elderly support. *Cogn Comput*. 2016;8(2):368–84.
- Riboni D, Bettini C, Civitarese G, Janjua ZH, Helaoui R. SmartFABER: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment. *Artif Intell Med*. 2016;67:57–74.
- Seelye AM, Schmitter-Edgecombe M, Cook DJ, Crandall A. Naturalistic assessment of everyday activities and prompting technologies in mild cognitive impairment. *J Int Neuropsychol Soc*. 2013;19(4):442–52.
- Dawadi P, Cook DJ, Schmitter-Edgecombe M. Automated cognitive health assessment using smart home monitoring of complex tasks. *IEEE Trans Syst Man Cybern Syst*. 2013;43(6):1302–13.
- Mc Ardle R, Del Din S, Galna B, Thomas A, Rochester L. Differentiating dementia disease subtypes with gait analysis: feasibility of wearable sensors? *Gait & Posture*. 2020;76:372–6.
- Goerss D, Hein A, Bader S, Halek M, Kernebeck S, Kutschke A, Heine C, Krueger F, Kirste T, Teipel S. Automated sensor-based detection of challenging behaviors in advanced stages of dementia in nursing homes. *Alzheimers Dement*. 2019;.
- Kirste T, Hoffmeyer A, Koldrack P, Bauer A, Schubert S, Schröder S, Teipel S. Detecting the effect of Alzheimer's disease on everyday motion behavior. *J Alzheimers Dis*. 2014;38(1):121–32.
- Kekade S, Hsieh C-H, Islam MM, Atique S, Khalfan AM, Li Y-C, Abdul SS. The usefulness and actual use of wearable devices among the elderly population. *Comput Methods Prog Biomed*. 2018;153:137–59.

28. Vuong N, Chan S, Lau CT, Lau K. Feasibility study of a real-time wandering detection algorithm for dementia patients. In Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare 2011. p. 1–4.
29. Lin Q, Zhao W, Wang W. Detecting dementia-related wandering locomotion of elders by leveraging active infrared sensors. *Journal of Computer and Communications*. 2018;6(05):94.
30. Dodge H, Mattek N, Austin D, Hayes T, Kaye J. In-home walking speeds and variability trajectories associated with mild cognitive impairment. *Neurology*. 2012;78(24):1946–52.
31. Barra S, Carta SM, Corrigan A, Podda AS, Recupero DR. Deep learning and time series-to-image encoding for financial forecasting. *IEEE/CAA Journal of Automatica Sinica*. 2020;7(3):683–92.
32. Qin Z, Zhang Y, Meng S, Qin Z, Choo K-KR. Imaging and fusing time series for wearable sensor-based human activity recognition. *Information Fusion*. 2020;53:80–7.
33. Kiangala KS, Wang Z. An effective predictive maintenance framework for conveyor motors using dual time-series imaging and convolutional neural network in an industry 4.0 environment. *IEEE Access*, 8:121033–121049, 2020.
34. Wang P, Li W, Li C, Hou Y. Action recognition based on joint trajectory maps with convolutional neural networks. *Knowl-Based Syst*. 2018;158:43–53.
35. Kieu T, Yang B, Guo C, Jensen CS. Distinguishing trajectories from different drivers using incompletely labeled trajectories. In Proceedings of the 27th ACM international conference on information and knowledge management 2018. p. 863–872.
36. Endo Y, Toda H, Nishida K, Ikeda J. Classifying spatial trajectories using representation learning. *International Journal of Data Science and Analytics*. 2016;2(3–4):107–17.
37. Gochoo M, Tan T-H, Velusamy V, Liu S-H, Bayanduuren D, Huang S-C. Device-free non-privacy invasive classification of elderly travel patterns in a smart house using PIR sensors and DCNN. *IEEE Sensors J*. 2017;18(1):390–400.
38. Riboni D, Murru F. Unsupervised recognition of multi-resident activities in smart-homes. *IEEE Access*. 2020;8:201985–94.
39. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint 2012*. [arXiv:1207.0580](https://arxiv.org/abs/1207.0580)
40. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929–58.
41. Das R, Chaudhuri S. On the separability of classes with the cross-entropy loss function. *arXiv preprint 2019*. [arXiv:1909.06930](https://arxiv.org/abs/1909.06930)
42. Bengio Y. Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, p. 437–478. Springer, 2012.
43. Cook DJ, Schmitter-Edgecombe M, Crandall A, Sanders C, Thomas B. Collecting and disseminating smart home sensor data in the CASAS project. In Proceedings of the CHI Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research, 2009.
44. Cook DJ, Crandall AS, Thomas BL, Krishnan NC. CASAS: A smart home in a box. *Computer*. 2013;46(7):62–9.
45. Burgess PW. Strategy application disorder: the role of the frontal lobes in human multitasking. *Psychol Res*. 2000;63(3–4):279–88.
46. Rashidi P, Cook DJ, Holder LB, Schmitter-Edgecombe M. Discovering activities to recognize and track in a smart environment. *IEEE Trans. Knowl. Data Eng*. 2011;23(4):527–39.
47. Khodabandehloo E, Riboni D, Alimohammadi A. HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline. *Futur Gener Comput Syst*. 2021;116:168–89.
48. Frank E, Hall MA, Holmes G, Kirkby R, Pfahringer B. WEKA - A machine learning workbench for data mining. In *The Data Mining and Knowledge Discovery Handbook*, pages 1305–1314. Springer, 2005.
49. Rish I, et al. An empirical study of the Naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence 2001 volume 3*, p. 41–46.
50. le Cessie S, van Houwelingen J. Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1992;41(1):191–201.
51. Shawe-Taylor J, Cristianini N. *Support vector machines*, volume 2. Cambridge University Press Cambridge, 2000.
52. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. *Mach Learn*. 1991;6(1):37–66.
53. Cohen WW. Fast effective rule induction. In *Machine learning proceedings*. Elsevier. 1995. p. 115–123.
54. Quinlan JR. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
55. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern*. 1991;21(3):660–74.
56. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
57. Schwartz MF, Segal M, Veramonti T, Ferraro M, Buxbaum LJ. The naturalistic action test: A standardised assessment for everyday action impairment. *Neuropsychol Rehabil*. 2002;12(4):311–39.
58. Seligman SC, Giovannetti T, Sestito J, Libon DJ. A new approach to the characterization of subtle errors in everyday action: Implications for mild cognitive impairment. *Clin Neuropsychol*. 2014;28:97–115.
59. Deng C, Xue Y, Liu X, Li C, Tao D. Active transfer learning network: A unified deep joint spectral-spatial feature learning model for hyperspectral image classification. *IEEE Trans Geosci Remote Sens*. 2018;57(3):1741–54.