UNIVERSITA' POLITECNICA DELLE MARCHE

FACOLTA' DI INGEGNERIA

Master's Degree in Biomedical Engineering

Data processing techniques for well-being and health monitoring of ageing people with mild cognitive impairment

Supervisor:

Prof. **Lorenzo Scalise**

Candidate:

Claudia Franceschetti

Co-supervisors:

Prof. **Sara Casaccia**

Dott. Ing. **Nicole Morresi**

Academic Year 2021/2022

Università Politecnica delle Marche

Department of Mechanical Engineering and Mathematical Sciences

Via Brecce Bianche — 60131 - Ancona, Italy

Contents

CHAPTER 1: INTRODUCTION **1**

1.1 Background and motivation 1

1.2 State of the Art 2

1.3 RESILIEN-T Project 7

1.4 Thesis Objectives 9

CHAPTER 2: MATERIALS AND METHODS **10**

2.1 Project Set-Up and Data Collection 10

2.2 Analysis of the Population 14

2.3 Data Analysis 19

2.3.1 Pre-Processing *20*

2.3.2 Dataset Creation *21*

2.3.3 Impact of the RESILIEN-T System *25*

2.3.4 Statistical analysis *26*

2.3.5 Supervised Machine Learning Analysis *27*

2.3.6 Unsupervised Machine Learning Analysis *37*

CHAPTER 3: RESULTS AND DISCUSSION **42**

3.1 Results about Impact of the System 42

3.2 Statistical analysis results 45

3.3 SML Results 48

3.4 UML Results 52

CHAPTER 4: CONCLUSIONS **54**

BIBLIOGRAPHY .... **V**

List of Figures

Figure 1: RESILIEN-T Architecture 8
Figure 2: Compaan Tablet 8
Figure 3: iHealth Wave Wearable device 9
Figure 4: Example of application of classification algorithm, based on μ ± σ, on user 2 Steps data 13
Figure 5: Distribution of study population for each partner (Netherlands, Italy, Switzerland and Canada) 14
Figure 6: Steps data of the user 2: outlier highlighted by the red square. 21
Figure 7: Steps trend in the whole period of trial of user 2, moving average (μ) on 2-weeks window and μ ± σ lines 23
Figure 8: basic idea of SVM classifier 29
Figure 9: Example of k-NN classification. 31
Figure 10: Idea behind the boosting algorithm, consisting of the parallel learners and weighted dataset 31
Figure 11: Illustration for the concept of bootstrap aggregating 32
Figure 12: Simplified representation of a Neural Network 33
Figure 13: Steps trend of user 2 with a 2-weeks window for calculation of μ 36
Figure 14: Steps trend of user 2 with a 1-week window for calculation of μ 36
Figure 15: Iterative K centroids positioning 39
Figure 16: Schematic representation of Hierarchical Clustering method 40
Figure 17: Impact of the system on user 1 42
Figure 18: Impact of the system on user 2 42
Figure 19: Impact of the system on user 3 43
Figure 20: Impact of the system on user 4 44
Figure 21: Impact of the system on user 5 44
Figure 22: Impact of the system on user 6 45
Figure 23: Trend of Wellness score and Steps data of user 5 47
Figure 24: Trend of Wellness score and DiffStep data of user 5 47
Figure 25: Comparison between mean accuracies of 3 features and 5 features dataset for each algorithm 49
Figure 26: Comparison between mean accuracies of original, 5x and 10x augmented datasets for each algorithm 51

List of Tables

Table 1: Summary of contents available in the APP 11

Table 2: Surveys questions and possible answers 12

Table 3: Information about Italy population 16

Table 4: Information about Switzerland population 16

Table 5: results of the Statistical Analysis of User 1. 17

Table 6: Information about Netherlands population 17

Table 7: Statistical information about trial duration for each partner population 18

Table 8: Percentage of data completeness, divided in 4 ranges, for each partner population 18

Table 9: Information about trials of Italy users considered in the analysis. 20

Table 10: Extract of the final dataset 24

Table 11: Examples of coaching messages based on steps trend 25

Table 12: SML Dataset with 3 features 35

Table 13: SML Dataset with 5 features 35

Table 14: Extract of UML Dataset 38

Table 15: Pearson coefficients 46

Table 16: Spearman coefficients 46

Table 17: Results of supervised analysis on the 3 features dataset 48

Table 18: Results of supervised analysis on the 5 features dataset 48

Table 19: SML Accuracies on the original dataset 50

Table 20: SML Accuracies on the 5 times augmented dataset 50

Table 21: SML Accuracies on the 10 times augmented dataset 50

Table 22: Silhouette Scores for K-means clustering analysis on 5 times and 10 times augmented dataset 52

Table 23: Silhouette Scores for Hierarchical clustering analysis on 5 times and 10 times augmented dataset 52

Chapter 1: Introduction

1.1 Background and motivation

In recent years, the average life expectancy is considerably increased, creating a rapid growth of the aging population. The increase in life expectancy and the decrease in birth rate are predicted to generate soon a substantial elderly population. Worldwide, the number of people aged 65 years old or over was 727 million in 2020. In the next three decades, this number is estimated to more than double, achieving over 1.5 billion in 2050, leading to a consistent increase from 9.3% to 16% of the world's population [1, 2].

Despite the development of medicine and the consequent success of reduced mortality, a large part of the world's population is still affected by age-related health problems such as dementia, Alzheimer's, cardiovascular diseases, diabetes, chronic age-related diseases, limitations in physical activities and so on [3].

Mild Cognitive Impairment (MCI) is one of these age-related conditions, in which an individual has mild but measurable changes in cognitive abilities that are greater than expected. However, even though these changes are noticeable, they do not affect the individual's ability to carry out everyday activities. Approximately, from 15% to 20% of people aged 65 or older have MCI. People with MCI, especially MCI involving memory problems, are more likely to develop Alzheimer's disease or other dementias than people without MCI [4,5].

Therefore, it is evident the parallel growth of life expectancy on one side and the demand and costs for healthcare services on the other one.

Prevention is the key: there is good evidence that a healthy lifestyle and cognitive stimulation can reduce both the risk of developing dementia and its progression [6,7].

According to predictions, however, this type of care provided by informal caregivers, i.e., friends and family members, will decrease in the future. For this reason, studies encourage society to concentrate on improving the lifestyle of the elderly, helping them to remain independent for a longer period, by using innovative tools [8].

In this scenario, in fact, ICT solutions could represent an effective tool to encourage people with Cognitive Impairment (PwCI) and with any other frailty condition to actively participate in the care decision-making process, allow them to interact directly with healthcare providers, express their personal health concerns and take initiatives to improve their own health.

In particular, the advances and diffusion of mobile technologies (i.e., smartphones and connected objects), artificial intelligence and robots are paving the way for the development of virtual coaches or e-coaches, which are able to support, complement, and possibly replace human coaches in health interventions.

This kind of solution would have the potential to make the healthcare system not only effective but more efficient: decreasing the burden of caregivers, who are less and less available, as well as healthcare costs, which are increasingly expected to rise.

The work presented in this thesis is aimed at proposing a system for health monitoring through well-being measurement of elderly subjects affected by MCI with data collected in the context of the use of a coaching system promoting their empowerment.

1.2 State of the Art

A new research trend is developing in this direction: proposing methods to identify and measure the activity and well-being of elderly users who are healthy or suffering from cognitive impairment.

The focus of these studies is not limited to user’s monitoring but also concerns the development of systems that can promote users' well-being through physical, cognitive and social stimulation. This kind of technological solution is generally called “virtual coach” or “e-coach”.

In **[9]** authors propose a smartphone wellbeing application which monitors user behaviour along three health dimensions: sleep, physical activity, and social interaction.

The aim of the system is to improve behavioural patterns via feedback rendered as an ambient display on the smartphone's wallpaper. Feedback mechanism would allow users to easily understand the consequences of their actions, enabling them to make appropriate changes in their behaviour and more informed choices going forward. Findings from the study show that users react positively to their overall experience and even show improvements in their ability to link everyday actions to wellbeing outcomes.

In [8] authors present the EMPATHIC (Empathic, Expressive, Advanced Virtual Coach to Improve Independent Healthy- Life-Years of the Elderly) project which aims to contribute to technological progress in the area of assistive ICT systems by researching, innovating and validating new interaction paradigms and platforms for future generations of personalized virtual coaches (VC) to promote healthy and independent aging. EMPATHIC-VC is presented as a non-obtrusive, emotionally-expressive virtual coach whose aim is to engage senior users in enjoying a healthier lifestyle concerning diet, physical activity, and social interactions. The objective is to assist and guide elderly people in actively minimizing their risk of potentially chronic diseases, contributing to their ability to maintain a pleasant and autonomous life, helping in turn their carers.

EMPATHIC-VC is a conversational agent based on Natural Language Understanding, Natural Language Generator and emotion detection from speech and text. The aim is to improve life quality of the user by means of coaching sessions in different topics. In contrast to other approaches such as task-oriented dialogue systems and chit-chat implementations, the agent should display a pro-active attitude, driving the conversation to reach a number of diverse coaching goals, requiring ability for adaptation to individual user profiles and skills, preferences and emotional states [10].

The paper [8] focuses on some sessions where the seniors carried out interactions with a Wizard of Oz driven, simulated system. A coaching strategy based on the GROW model [11] is used throughout these sessions so as to guide interactions and engage the elderly with the goals of the project. This way, the senior users are, on the one hand, given the chance to interact with what they thought was a final system and, on the other hand, able to provide very valuable information as to its potential future developments. In addition, it allows for the collection of an audio-visual data corpus which is currently used to train the machine learning models underpinning the different modules of the entire EMPATHIC system.

So, the main focus of the analysis of the elderly behavior while interacting with the simulated virtual coach relies on their affective state, which might work as an indicator of the success of the Virtual Coach (VC).

Results show that users have positive feelings with regard to the interaction with the system. In terms of categories, the most frequent label was *calm*, suggesting also that the dialogue system seems to be perceived user friendly.

In a systematic review of 2020 [12] Virtual Coaches are defined as “disruptive technologies in the healthcare sector” as they might provide cost-effective solutions for increasing human wellbeing in different domains such as physical, nutritional, cognitive, social, and emotional.

Such digital systems assume various forms, from classic apps, to more advanced conversational agents or robots. The main activity considered to be part of the e-coach intervention is linked to supporting the user to reach predefined goals, motivating him/her and possibly providing personalization of the intervention. In this context, models of health behavior change theories are used to understand how a set of psychological constructs can jointly explain how individuals can be motivated to change an established behavioral pattern. The behavioral pattern is based on the interest of improving or maintaining overall long-term wellbeing, that is the final purpose of Virtual Coaches technology. Main intervention techniques and features that could affect health outcomes, usability and adherence are for example reduction of activity options by setting short-term goals, to eventually reach long-term goals; personalization of goals with intermediate goal thresholds adapted by the system according to user’s preferences and states; use of validity-tested devices. Coaching domains that appear to be more treated are related to physical activity, nutrition, social contacts, cognitive activity and emotional state [12].

Regarding the system architectures, system implementation is divided into three components: monitoring, processing, and intervention medium. First of all, the choice of hardware and software used for monitoring the user's progress with the coaching system is essential to the design of an intervention medium.

Concerning the hardware, the most commonly reported cases are those of pedometers [13-19] or accelerometer-based activity sensors [20, 21] used to track physical activity. Many studies explore the use of robots equipped with a variety of sensors for motion or voice recognition [22-29], or environmental sensors to detect humidity, gas, and smoke conditions [30]. In other cases, devices used to gather information about monitoring of elderly people are bracelets [31,32] and tablets [33,34].

In terms of software for self-reporting, the most common examples are mobile/tablet applications [14,16] and web or PC applications [13, 16, 18, 20, 21]. In some instances, where a robot is used, it's possible to communicate information verbally [22].

Second component of system implementation is related to processing of the acquired data in terms of techniques applied to recognize user activities, e.g., step counter [13,14,17] or self-monitoring [18]. In [9] for example, data collected from accelerometer, microphone, and phone usage are elaborated to derive the score for three different dimensions: physical activity, sleep, social activity. Some studies use detection algorithms regarding speech [22], gesture and emotion recognition as methods for processing data [23].

Regarding the last component of system implementation, Web or mobile applications are most widely used as an intervention medium for older users: coaching solutions are given through applications installed mainly on smartphone but also on tablet, laptop or PC [16, 18, 19, 20, 21, 24].

Hence, the literature shows us a wide variety of scenarios, whose common goal, however, is to offer a solution to maintain or increase the well-being of elderly and frail users by the use of new assistive technologies.

In the context of offering this type of assistive solutions, it emerges therefore the need to develop effective, reliable and affordable methods for measuring well-being and behavioral trends of users. Several studies have been carried out in this direction.

In [35] it's presented a measurement protocol to compute and distinguish abnormal from normal behavior of older people with early to middle stage dementia living alone at home using training artificial intelligence (AI) algorithms (K-means, Agglomerative and Spectral Clustering Algorithms). In order to gather knowledge about human health and well-being, the study aims at measuring and tracking the activities of daily living (ADLs) of older people in their home environment, by collecting data from an AI-based sensor network composed of Passive Infrared (PIR) motion sensors and door sensors. The obtained results confirmed that PIR and door sensors installed in individuals' homes can provide beneficial information about the status of the resident, since K-means and Agglomerative Clustering Algorithms are able to group participants with similar normal behaviours distinguishing them from those with an abnormal one.

Also in [36] a similar approach is used: data are acquired through a minimally invasive home sensor network, realised by movement and light sensors. Then, unsupervised machine learning algorithms (UML) (K-means) are used to cluster these daily data in different pattern of activation of light and movement sensors while supervised machine learning (SML) (decision tree algorithms) is implemented to give interpretation to the result of clustering processes of UML. The study demonstrates that the combination of UML and SML is effective to evaluate seniors' activity in the built environment appropriately equipped with PIR and light sensors by separating the daily number of activations of sensors in different time slots among the different domestic environments.

Similarly, in [37] a methodology based on SML analysis of data acquired by a domotic sensor network is proposed to measure the well-being of elderly people in private home environments. Two ML algorithms are compared, Random Forest (RF) and Regression Tree (RT), such that to verify whether the users' well-being is encoded in behavioural patterns obtained from the domotic data. In fact, changes in the behavioural patterns measured with domotic sensors could be associated with the variation of the user's well-being.

The novelty introduced in this study is in the use of daily self-evaluation surveys as a reference for the training of ML algorithms: the user responses represent the output of

ML models while the domotic data gathered in the home environment represent the input. In current literature, human behaviour is studied in terms of sequence of Activities of Daily Living (ADLs): each ADL is individuated as a sequence of sensor activation patterns that characterise human behaviour but do not provide any information about users' self-perception and well-being. In this study, instead, authors propose a methodology to quantitatively measure the well-being of older users living at home and the only reference system to evaluate the well-being of the older users is provided by a daily survey. From the survey, the authors extracted numerical indices representing the self-evaluation of users' physical and mental states. Eventually, the prediction of human well-being derived from the trained algorithm, just using the domotic sensor data, may allow the identification of any cognitive or physical decline and provide specific services to improve the life-quality of the users at home. Final results suggest that RF and RT algorithms applied to domotic data provide a robust methodology for predicting the well-being of a user living in an apartment equipped with environmental sensors in a non-intrusive manner.

1.3 RESILIEN-T Project

In this research scenario, it is possible to place also the RESILIEN-T project, funded by the European 'Ambient Assisted Living' (AAL) programme which has the objective to promote the development of smart technologies applied to pro-active health care, independent living, and active aging at home.

RESILIEN-T project is addressed to elderly people suffering from MCI. Thanks to the development of a coaching solution, affected people are properly stimulated to reinforce their self-monitoring ability, with the purpose of slowing down their cognitive and behavioural decline and increase their resilience.

The system architecture (Fig. 1) is represented by a modular, integrated, and open platform offering different services. The main components are:

- a tablet (Compaan), with a specific user-friendly interface, on which the RESILIEN-T mobile application is installed to deliver coaching services related to feeding, physical activity, cognitive exercises, and social interaction;
- a wearable device (Fig. 2) fully integrated with the system (iHealth Wave) to collect mobility, physical activity, and other physiological data from the user;
- a remote cloud-based platform to collect all the data generated from the user's interaction with the tablet and from the Bluetooth synchronization of the smartwatch with the tablet, but also acting as a repository for the contents delivered to the user through the app.

Figure 1: RESILIEN-T Architecture

Figure 2: Compaan Tablet

Figure 3: iHealth Wave Wearable device

1.4 Thesis Objectives

This dissertation is born inside the framework of the RESILIEN-T project. In the first part of this work, the analysis of the quality of the collected dataset is performed: information about the amount of drop-out in the population, average time duration of each trial, and the percentage of data loss in the trial duration are extracted.

Then, starting from the steps data acquired by the smartwatch, the impact of RESILIEN-T systems on the behaviour of the daily steps was analysed observing variations of physical activity trend in correspondence of days in which users received motivational messages through the RESILIEN-T application on the tablet.

Thereafter, data collected from wearable sensor and RESILIEN-T app are analysed through traditional methods (i.e., Statistical Analysis) and innovative approaches (i.e., ML and AI) with the final aim to identify if it's possible to measure the users' Well-being, referring to the answers to daily questionnaires proposed through the RESILIEN-T application, using the dataset defined from the daily Steps data. Thus, well-being is intended in terms of health and mental status of older users based on their personal feeling/perception (questionnaires) and daily physical activities (smartwatch data).

The objective is therefore to understand whether it is possible to measure well-being in a totally non-invasive manner, without the need to install domotic sensor networks, starting from data collected by elderly users with MCI who live independently at home and who are supported by a coaching system to keep them active and stimulate them by offering different services.

Chapter 2: Materials and Methods

2.1 Project Set-Up and Data Collection

The RESILIEN-T project involved four European partners, INRCA, Careyn, GoldenAge, University of Toronto, and was carried out from 2019 to 2022.

The experiment with end-users has been conducted in Italy, Canada, Switzerland and in the Netherlands involving an overall amount of 96 total elderly people suffering from MCI, with a variable duration for each user according to participants' compliance.

Users were provided with the Compaan tablet, developed and tested for the elderly, on which, once having received their own accounts, they interface with the RESILIEN-T application to access coaching and keep track of information about their own activities. Users could choose the personality that best described their character and could select two topics that interest them most out of four available: nutrition, exercise, cognitive training and social interaction. This gave them possibility to personalize the tone and the content of the received coaching advice. In fact, based on these choices, a database of content was selected to provide appropriate coaching messages, to be shown once or twice a day. Table 1 shows some examples of coaching advice for each content, provided through the daily messages.

Topic	Motivation	Contents
Nutrition	Reduce consumption of salt, carbohydrates, sugar, packaged products. Take in the right amount of water, protein, vitamins, fruit and vegetables.	Textual information 

<table>
<tr><td>Physical activity</td><td>Balance exercises, regular exercise, stretching, aerobics, walking.</td><td rowspan="3">Images / Video / Quiz

Websites
</td></tr>
<tr><td>Cognitive training</td><td>Brain-training activities, memory, singing, having a hobby, breaking the routine, language training, riddles.</td></tr>
<tr><td>Social interaction</td><td>Maintaining social contacts, writing telephoning video chatting, rules for good conversation, volunteering, social work, group activities and courses.</td></tr>
</table>

Table 1: Summary of contents available in the APP

Users were also provided with a smartwatch (Wave model from iHealth) that constituted one of the elements of the monitoring and coaching activity. Measurement set-up was represented by the wearable device, being able to collect data about:

- Daily Number of Steps
- Calories burned (kcal)
- Distance travelled (km)
- Sleep efficiency (%)
- Time in Bed period (hh:mm)

The tablet provided to the users, instead, allowed the collection of data about answers to daily questionnaires.

Normally the well-being of users can be evaluated through a series of questionnaires, that represent a reliable measurement of human well-being, as reported in literature [38]. In order to monitor the physical and mental status of users and possible variations of their condition, the study protocol included the completion of daily surveys by the users.

In particular, in the morning users were asked how they perceived their health to investigate the self-evaluation of well-being. Then, in the evening, they were asked to give feedback on their appreciation for the activity suggested in the RESILIEN-T daily message. The answers were given through a specific scoring system: 1 stands for 'good', 2 stands for 'could do better', 3 stands for 'bad'. The data collected were related to the answers to these questionnaires (Wellness score and Feedback score). Questionnaire's answers were stored in an Excel file: each row of the file reports the date of fulfilment, the user's id number, and his/her relative answers. Each answer is mapped to the range 1 to 3 so that the best answer is represented by the value 1, the middle answer by the value 2, while the worst answer is equivalent to a value of 3. The survey questions provided to users on a daily basis and the possible answers are shown in the Table 2.

Wellness questionnaire	**Possible answers**
How do you feel this morning?	*Good* (score 1); *Ok* (score 2); *Bad* (score 3)

Feedback questionnaire	**Possible answers**
Did you enjoy today's activity	*Good* (score 1); *Could do better* (score 2); *Bad* (score 3)

Table 2: Surveys questions and possible answers

Moreover, participants could choose whether or not to use the wearable device. In order to collect data from the smartwatch on the cloud platform, the users had to synchronise the device with the tablet via Bluetooth.

Due to system errors during the data upload phase on the tablet, data related to calories and distance were incorrect while data about sleep information were not present since users did not wear the device during the night. Only steps data were correctly acquired and so, used in the analysis presented in this work.

Thanks to the monitoring of physical activity of users in terms of daily steps, RESILIEN-T system included a coaching service to motivate and stimulate users to keep physically active according to their activity trend.

A specific algorithm has been designed and implemented on the system to define if daily measured steps were within, above or below the average trend. The average activity trend was considered calculating the moving average (μ) of the number of Steps applied to a period of 2-weeks with a 1-day sliding window. Also the standard deviation associated to the same time window was calculated (σ). The criteria used for the classification was based on the following rule:

- days within the range μ ± σ were classified within the category 'Normality';
- days with a number of steps above the range μ + σ were classified as 'Positive Abnormality';
- days with a number of steps below the range μ - σ were classified as 'Negative Abnormality'.

Figure 4 shows an example of application of the classification algorithm on the steps data of user 2, highlighting with different colours the trend of moving average and the ranges defined by the moving standard deviation.

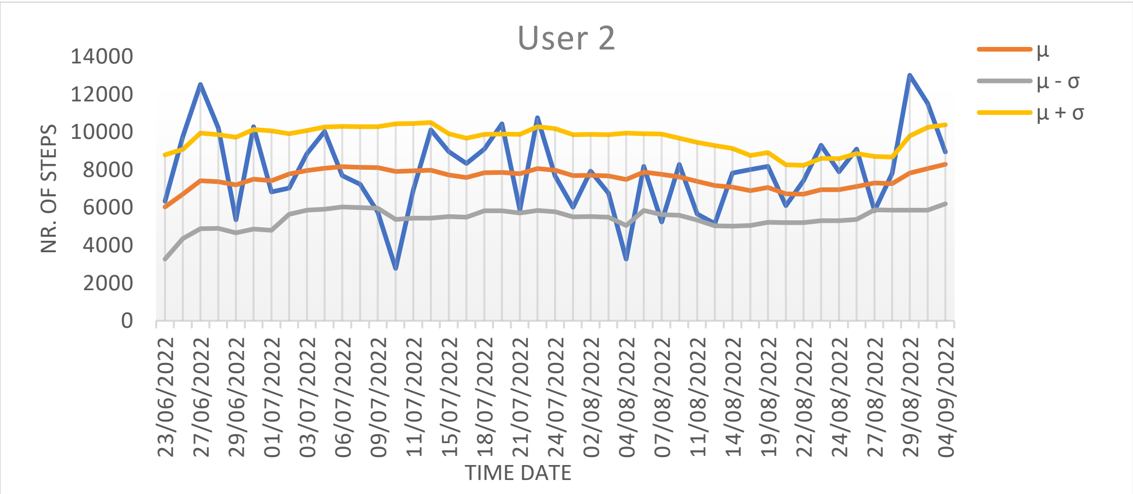

Figure 4: Example of application of classification algorithm, based on μ ± σ, on user 2 Steps data

The classification results were used as decision criteria between 3 main types of messages that include suggestions in case of 'normality', congratulations in case of 'positive abnormality', or alert messages in case of 'negative abnormality'. The different texts of messages sent on the Tablet were user specific, reporting the name of the elderlies wearing the smartwatch.

2.2 Analysis of the Population

The study involved an overall population of 96 elderly people, distributed with different percentage for each partner, as shown in the pie chart of Figure 5.

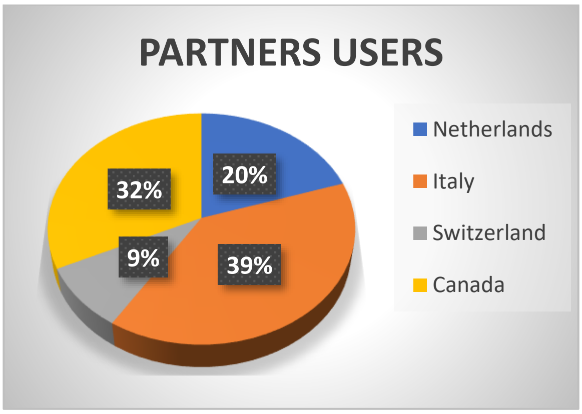

Figure 5: Distribution of study population for each partner (Netherlands, Italy, Switzerland and Canada)

Italy enrolled the majority of the users involved (39%) and immediately after we find Canada. A 20% of users came from The Netherlands while just a 9% was enrolled by Switzerland.

A preliminary analysis of the population was done to investigate average amount of data collected for each partner, presence of drop-out and gaps in the acquisition period. In fact, the questionnaire file of the users did not contain all the days of the analysis period since the users could skip some days between one survey and the following one.

The analysis of collected data was done considering the answers to the survey of the evening, that one related to the feedback respect to the activity proposed by RESILIEN-T application.

Table 3,4,5 and 6 show information, respectively for Italy, Switzerland, Canada and Netherlands users, about:

- date of the first acquisition for each user;
- date of the last acquisition for each user;
- number of total days from the first to the last acquisition for each user;
- number of effective days of acquisition in the entire period of the trial for each user;
- % of acquired days with respect to the total period of the trial for each user, calculated with the following equation (Equation 1):

$$\%\ of\, Acquired\ days = \frac{Nr.of\ \ Acquired\ days}{Nr.of\ \ Total\ days} \cdot 100 \tag{1}$$

- % of missing days of acquisition (gaps) with respect to the total period of the trial for each user, calculated with Equation 2:

$$\%\ of\, Missing\ days\ of\, Acquisition = \frac{Nr.of\ \ Tot.days - Nr.of\ \ Acquired\ days}{Nr.of\ \ Tot.days} \cdot 100 \tag{2}$$

Italy Users	Start date	End date	Nr. of total days	Nr. of acquired days	% of acquired days	% of missing days
User 1	08/02/2022	24/05/2022	106	26	25%	75%
User 2	09/06/2022	07/09/2022	91	89	98%	2%
User 3	17/03/2022	12/07/2022	118	50	42%	58%
User 4	20/01/2022	10/03/2022	50	37	74%	26%
User 5	20/01/2022	28/04/2022	99	96	97%	3%
User 6	11/03/2022	25/06/2022	107	66	62%	38%
User 7	18/03/2022	19/06/2022	94	39	41%	59%
User 8	04/02/2022	30/05/2022	116	78	67%	33%
User 9	17/02/2022	24/04/2022	67	48	72%	28%
User 10	18/01/2022	10/05/2022	113	67	59%	41%
User 11	16/06/2022	22/08/2022	68	53	78%	22%
User 12	12/05/2022	01/09/2022	113	108	96%	4%
User 13	21/02/2022	25/05/2022	94	90	96%	4%
User 14	19/05/2022	29/08/2022	103	101	98%	2%
User 15	24/01/2022	05/06/2022	133	85	64%	36%
User 16	31/01/2022	10/05/2022	100	94	94%	6%
User 17	12/05/2022	30/08/2022	111	102	92%	8%
User 18	03/05/2022	22/08/2022	112	101	90%	10%
User 19	03/02/2022	02/05/2022	89	69	78%	22%
User 20	18/02/2022	10/05/2022	82	76	93%	7%
User 21	22/03/2022	26/06/2022	97	93	96%	4%
User 22	18/01/2022	21/04/2022	94	91	97%	3%
User 23	10/03/2022	09/06/2022	92	78	85%	15%
User 24	13/05/2022	24/08/2022	104	88	85%	15%
User 25	29/03/2022	12/07/2022	106	104	98%	2%
User 26	18/02/2022	14/05/2022	86	28	33%	67%
User 27	01/02/2022	04/05/2022	93	88	95%	5%
User 28	08/03/2022	24/08/2022	170	114	67%	33%
User 29	07/02/2022	22/05/2022	105	95	90%	10%
User 30	07/04/2022	24/04/2022	18	13	72%	28%
User 31	04/02/2022	17/05/2022	103	101	98%	2%
User 32	18/01/2022	31/01/2022	14	14	100%	0%
User 33	02/02/2022	02/05/2022	90	88	98%	2%
User 34	13/05/2022	28/08/2022	108	71	66%	34%
User 35	21/04/2022	28/07/2022	99	84	85%	15%
User 36	10/06/2022	04/09/2022	87	85	98%	2%
User 37	25/01/2022	08/06/2022	135	26	19%	81%

Table 3: Information about Italy population

Switzerland Users	Start date	End date	Nr. of total days	Nr. of acquired days	% of acquired days	% of missing days
User 1	11/03/2022	15/05/2022	66	10	15%	85%
User 2	19/04/2022	18/07/2022	91	87	96%	4%
User 3	16/02/2022	06/05/2022	80	20	25%	75%
User 4	12/02/2022	22/06/2022	131	83	63%	37%
User 5	16/02/2022	06/07/2022	141	131	93%	7%
User 6	16/02/2022	03/07/2022	138	91	66%	34%
User 7	16/02/2022	21/07/2022	156	151	97%	3%
User 8	08/02/2022	22/06/2022	135	106	79%	21%
User 9	17/02/2022	07/09/2022	203	55	27%	73%

Table 4: Information about Switzerland population

Canada Users	Start date	End date	Nr. of total days	Nr. of acquired days	% of acquired days	% of missing days
User 1	22/04/2022	21/07/2022	90	83	92%	8%
User 2	27/04/2022	28/07/2022	92	76	83%	17%
User 3	09/06/2022	30/08/2022	82	27	33%	67%
User 4	02/04/2022	27/04/2022	25	15	60%	40%
User 5	14/03/2022	11/05/2022	58	9	16%	84%
User 6	25/03/2022	07/07/2022	104	82	79%	21%
User 7	13/04/2022	26/04/2022	13	9	69%	31%
User 8	17/04/2022	05/07/2022	79	14	18%	82%
User 9	24/05/2022	13/08/2022	81	31	38%	62%
User 10	17/05/2022	15/08/2022	90	31	34%	66%
User 11	11/05/2022	15/08/2022	96	83	86%	14%
User 12	03/04/2022	06/07/2022	94	93	99%	1%
User 13	17/03/2022	27/06/2022	102	91	89%	11%
User 14	10/03/2022	17/06/2022	99	55	56%	44%
User 15	22/04/2022	24/07/2022	93	80	86%	14%
User 16	10/03/2022	22/05/2022	73	27	37%	63%
User 17	29/04/2022	26/07/2022	88	14	16%	84%
User 18	28/03/2022	10/05/2022	43	7	16%	84%
User 19	15/05/2022	18/08/2022	95	38	40%	60%
User 20	20/06/2022	22/08/2022	63	15	24%	76%
User 21	27/04/2022	26/07/2022	90	47	52%	48%
User 22	07/05/2022	21/08/2022	106	65	61%	39%
User 23	12/05/2022	01/08/2022	81	44	54%	46%
User 24	10/03/2022	06/07/2022	118	104	88%	12%
User 25	12/04/2022	15/07/2022	94	92	98%	2%
User 26	04/06/2022	21/06/2022	17	15	88%	12%
User 27	02/05/2022	03/07/2022	62	49	79%	21%
User 28	27/03/2022	03/07/2022	98	81	83%	17%
User 29	11/03/2022	22/05/2022	72	30	42%	58%
User 30	07/03/2022	16/06/2022	101	48	48%	52%
User 31	20/04/2022	06/07/2022	77	23	30%	70%

Table 5: results of the Statistical Analysis of User 1.

Netherlands Users	Start date	End date	Nr. of total days	Nr. of acquired days	% of acquired days	% of missing days
User 1	28/01/2022	02/05/2022	94	61	65%	35%
User 2	18/01/2022	29/07/2022	192	37	19%	81%
User 3	20/05/2022	03/06/2022	14	8	57%	43%
User 4	12/02/2022	01/04/2022	48	11	23%	77%
User 5	02/02/2022	11/07/2022	159	91	57%	43%
User 6	21/05/2022	19/08/2022	90	84	93%	7%
User 7	22/04/2022	18/05/2022	26	12	46%	54%
User 8	26/05/2022	18/08/2022	84	53	63%	37%
User 9	30/01/2022	12/02/2022	13	10	77%	23%
User 10	04/02/2022	07/05/2022	92	12	13%	87%
User 11	29/04/2022	17/06/2022	49	35	71%	29%
User 12	28/01/2022	14/02/2022	17	13	76%	24%
User 13	21/05/2022	01/09/2022	103	101	98%	2%
User 14	27/05/2022	05/09/2022	101	63	62%	38%
User 15	22/05/2022	19/07/2022	58	56	97%	3%
User 16	30/01/2022	07/02/2022	8	**8**	100%	0%
User 17	01/06/2022	01/08/2022	61	9	15%	85%
User 18	10/03/2022	16/03/2022	6	6	100%	0%
User 19	22/05/2022	03/09/2022	104	62	60%	40%

Table 6: Information about Netherlands population

From those data, statistical information was extracted about maximum, minimum and mean values, with standard deviation, of number of effectively acquired days for each partner population (Table 7). Moreover, an investigation about degree of data completeness was performed: Table 8 reports the percentage of data completeness, for each partner population, in terms of ratio between effective days of acquisition and entire trial duration. Level of completeness has been represented considering 4 ranges: from 0 to 25%; from 25% to 50%; from 50% to 75% and from 75% to 100%.

PARTNER	MAXIMUM TRIAL DURATION (DAYS)	MINIMUM TRIAL DURATION (DAYS)	MEAN TRIAL DURATION (DAYS)	STD TRIAL DURATION (DAYS)
Italy	114	13	74	28
Switzerland	151	10	82	47
Canada	104	7	48	31
Netherlands	101	6	39	32

Table 7: Statistical information about trial duration for each partner population

PARTNER	Data completeness 0-25%	Data completeness 25%-50%	Data completeness 50%-75%	Data completeness 75%-100%
Italy	5%	8%	24%	62%
Switzerland	11%	22%	22%	44%
Canada	16%	26%	19%	39%
Netherlands	21%	5%	37%	37%

Table 8: Percentage of data completeness, divided in 4 ranges, for each partner population

From the observation of Tables 7 and 8 it emerged that populations with largest number of mean days of acquisition were from Switzerland and Italy, with 82 and 74 days respectively. Nevertheless, Italy had a quite smaller standard deviation (28 days vs. 47 days) and a way larger population (39% vs. 9% of the entire sample) with respect to Switzerland one, showing a more homogenous duration of the trials among participants.

Moreover, Italy appeared to have the largest percentage of population in the highest range of data completeness (62% of its population), while just 13% of Italy users had a percentage of data completeness smaller than 50%. Thus, Italy population showed the lower presence of gaps inside the acquisition phase and the lower number of drop-out.

However, in general all the partners show quite good adherence to the RESILIEN-T coaching system: at least 58% of each partner population had a percentage of data completeness higher than 50%.

2.3 Data Analysis

Data concerning answers to morning and evening surveys were collected from the entire population of 96 elderly people suffering from MCI. However, just a few percentage of participants decided to use the wearable device. This decreases dramatically the amount of data available for the analysis performed. In fact, the aim of this work was to investigate a methodology that can predict the physical and mental well-being of an elderly subject with early cognitive decline using the data acquired from a smartwatch, a daily wellness self-evaluation survey and a daily questionnaire about satisfaction with the proposed coaching activities.

The first approach used was to verify if prediction of the user's well-being was possible by training Supervised ML (SML) algorithms using the morning questionnaire's answers as reference data. The idea behind this analysis is that the changes in the behavioural patterns measured with wearable sensors could be associated with the variation of the user's well-being.

A second approach, instead, was the application of Unsupervised ML (UML), trying to perform clustering of the dataset to identify different behavioural pattern based on daily steps data.

For this reason, this analysis was focused on the users who wore the smartwatch and answered to the surveys in the same days, in order to have a complete dataset. Among all users, only a subset of Italy population used the wearable device. In particular, a total amount of 31 Italian participants wore the smartwatch in some occasions, but of this group just **8 users** had at least 14 days of acquisitions for both feedback and wellness questionnaires data and steps data and just **6 users** collected for at least 30 days feedback and wellness questionnaires data and the steps data.

In order to use a more robust approach, analysis mainly focused on **6 participants** with at least **30 days** of complete dataset. Table 9 shows information about first and last day of acquisition, number of effective days of acquisition and percentage of data completeness in terms of ratio between effective days of acquisition and entire trial duration. Almost all the users have collected slightly more than 30 days of acquisition, apart from user 2 who has a complete dataset of 57 days. User 2 was also the one with the highest percentage of data completeness (66%), the only one with more than a half, meaning that he or she interacted with the system and wore the smartwatch in a more continuous manner throughout the trial.

ITALY USERS	START DATE	END DATE	MONITORED PERIOD LENGTH (DAYS)	% OF DATA COMPLETENESS
USER 1	10/06/2022	07/09/2022	30	34%
USER 2	09/06/2022	04/09/2022	57	66%
USER 3	12/05/2022	27/08/2022	39	36%
USER 4	03/02/2022	19/05/2022	30	29%
USER 5	10/03/2022	14/05/2022	31	48%
USER 6	31/01/2022	10/05/2022	30	30%

Table 9: Information about trials of Italy users considered in the analysis.

2.3.1 Pre-Processing

After the data collection, the first step of the is the removal of all the possible outliers, defined as samples that are exceptionally far from the mainstream of the data. The data collected are pre-processed to remove all the possible days that contain values identified as outliers. In particular, we focused on Step data, that were the only reliable information collected from the smartwatch. A daily number of Steps equal to 0 was considered as an outlier and deleted, since users were expected to move or walk every day. In fact, steps count equal to zero could mean that the smartwatch was not worn by the user or that there was a smartwatch malfunctioning, thus providing useless information. Figure 6 shows the example of dataset from user 2 containing one outlier.

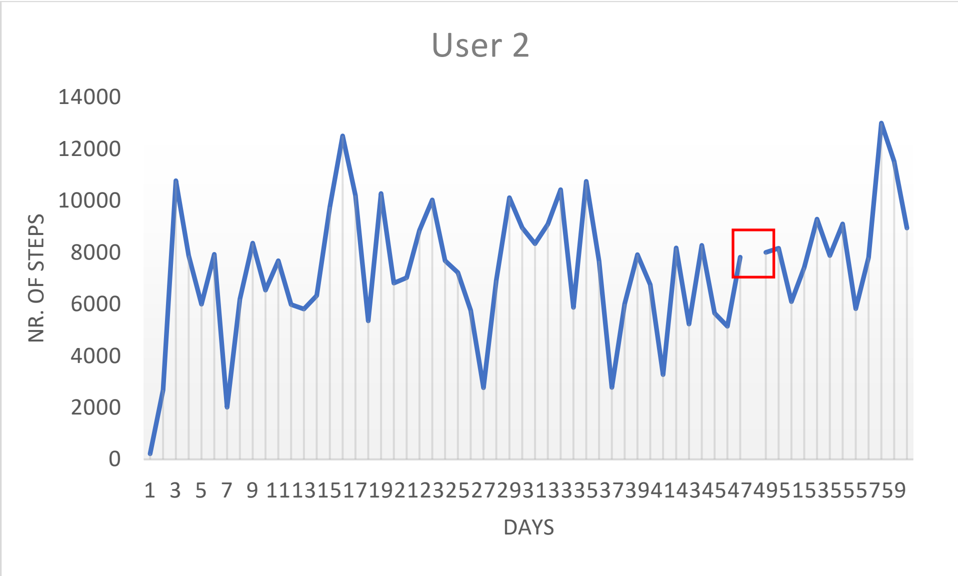

Figure 6: Steps data of the user 2: outlier highlighted by the red square.

Moreover, in the dataset some days appeared to have more than one answer to the same questionnaire, probably due to an error in the system synchronization. These days were removed from the dataset since it was not possible to discriminate which was the correct answer related to that specific day.

Finally, days for which the complete database was not available (feedback score, wellness score and steps count) were obviously not included in the analysis, since not all the needed information to perform a well-being prediction were present.

2.3.2 Dataset Creation

Once pre-processed, dataset of the 6 final users containing date of fulfilment, ID number, feedback score and wellness score have been elaborated to extract new features to be included in the dataset for the ML analysis, to predict the wellbeing of the user using data collected by the Resilient-t system.

In particular, 3 new features have been extracted from steps data to expand the database: Diffstep, Step Index and Step Score, which are described below.

DiffStep: This feature was calculated as the difference between the actual number of steps and the number of steps of the previous acquisition day, according to Equation 3:

$$DiffStep_i = Step_i - Step_{i-1} \quad (3)$$

With this procedure, we were forced to exclude the first row of each user dataset since no value of DiffStep was associated to the first day of acquisition.

Step Index: This feature was calculated based on the classification of ***DiffStep*** value, considering 5 different ranges:

- If $DiffStep_i < -5000$: *Step Index* = -2
- If $-5000 < DiffStep_i < -1000$: *Step Index* = -1
- If $-1000 < DiffStep_i < 1000$: *Step Index* = 0
- If $1000 < DiffStep_i < 5000$: *Step Index* = 1
- If $DiffStep_i > 5000$: *Step Index* = 2

Step Score: This feature was calculated applying an algorithm presented in [39], showing a measurement procedure for daily Motion and sleep classification in aging people using smartwatch data:

- Moving average (μ) and standard deviation (σ) of steps data on a 2-weeks window were calculated;
- For each sample, comparison between step value ($Step_i$) and $\mu \pm \sigma$ of previous 2-weeks was done;
- If $Step_i > \mu + \sigma$, step count was classified as a "Positive abnormality": *Step Score* = 1
- If $Step_i < \mu - \sigma$, step count was classified as a "Negative abnormality": *Step Score* = 3
- If $\mu - \sigma < Step_i < \mu + \sigma$, step count was within range of "Normality": *Step Score* = 2.

Scoring mechanism was chosen to reflect that one of the feedback and wellness surveys, defined in the study design by partners of the project, so that 1 represent the best condition, 3 was associated to the worst case and 2 reflected a normality condition.

Applying this algorithm, first 13 days of acquisition were lost in the final dataset since we could not calculate Step Score for the first 13 samples.

Figure 7 shows an example of steps trend in the whole period of trial of user 2. In evidence there are the moving average (μ) trend, calculated on a 2-weeks window, and upper and lower thresholds of $\mu \pm \sigma$. Step counts above the upper threshold were classified as positive abnormality and corresponded to a Step Score equal to 1 while below the lower threshold there are the data points associated to condition of negative abnormality and so a Step Score equal to 2. Within the range of $\mu \pm \sigma$ a normal condition was considered, corresponding to a Step Score equal to 2.

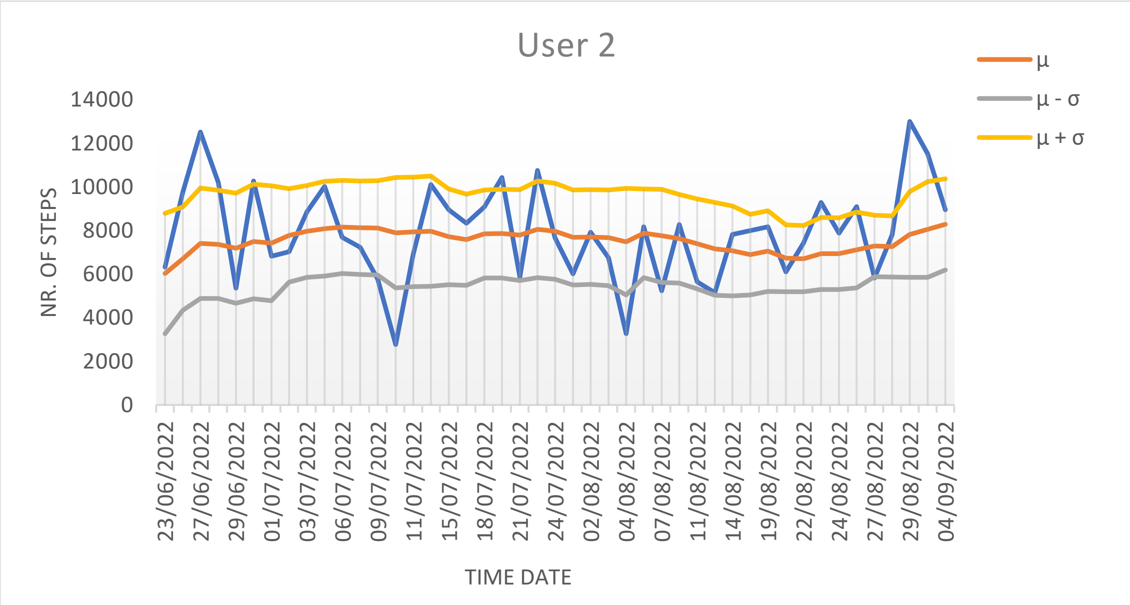

Figure 7: Steps trend in the whole period of trial of user 2, moving average (μ) on 2-weeks window and μ ± σ lines

After having calculated the three new features for each user database and having deleted the missing or repeated days, final dataset containing data of all the users have been created.

Table 10 reports an extract of the final dataset. First column with user number have been added to the previously mentioned features in order to be able to distinguish rows corresponding to each user and thus, being able to perform both inter-subject (considering the whole dataset with all the subjects together) and intra-subject (differentiating datasets of each subject) analysis.

USER NUMBER	STEPS	DIFF STEP	STEP INDEX	STEP SCORE	FEEDBACK SCORE	WELLNESS SCORE
1	12375	7206	2	2	2	2
1	13015	640	0	2	2	2
1	7764	-5251	-2	2	2	2
1	13037	5273	2	2	2	3
1	1804	-11233	-2	3	2	2
1	2829	1025	1	3	2	2
1	4072	1243	1	2	2	2
1	4624	552	0	2	2	2
…	…	…	…	…	…	…

Table 10: Extract of the final dataset

The focus of the analysis was the wellness score, as indicator of the well-being of the subjects. Different attempts have been done to investigate if some correlation between the wellness score and the other parameters of the dataset existed, in order to verify if it was possible to predict the well-being of users with mild cognitive decline starting from a dataset based on number of daily steps, features extracted from it and answers to survey about coaching activities satisfaction.

Thus, the final dataset has been analysed applying 3 different techniques:

- Traditional statistical methods;
- Supervised machine learning;
- Unsupervised machine learning.

2.3.3 Impact of the RESILIEN-T System

A preliminary insight about of the impact of the coaching system on user's physical activity was done, before proceeding with the analysis of the dataset to investigate possible identification of Well-being measurement procedures.

As explained in paragraph 2.1, RESILIEN-T system included a daily steps monitoring system, on the basis of which users wearing the smartwatch received a coaching message every day on the Tablet. Depending on whether the number of daily steps was higher, lower or in the range of the average trend of the past 14 days, the user was sent a congratulations, alert or suggestion message respectively. In fact, praise messages mechanism is presented in Literature as one of the main intervention techniques that could improve health outcomes and adherence to the program [12].

Table 11 shows example of coaching messages sent the users wearing the smartwatch.

BEHAVIOUR CLASSIFICATION	COACHING MESSAGGE
Positive abnormality	You had a great exercise, excellent! This is a good activity for you.
Normal condition	Your physical activity is good. Exercise can increase the well-being. Plan a walk!
Negative abnormality	Exercise such as walking can increase your well-being. Plan a walk!

Table 11: Examples of coaching messages based on steps trend

The aim of this analysis was to observe if receiving an alert message had a positive impact on the number of steps of the day following the message, with respect to the previous day and with respect to the average trend of daily steps of the last 2 weeks. The percentage variation of Steps respect to the day before (% Diff. d_{i+1} w.r.t. d_i) and to the mean value of the previous 2 weeks (% Diff. d_{i+1} w.r.t. μ_{14gg}) was done applying the following formulas, where $Step_i$ corresponds to the Nr. of steps of the day in which the message was sent, $Step_{i+1}$ represents the Nr. of steps measured the day after sending the message, and $\overline{Step}_{14gg}$ is the mean Nr. of steps of the last 2 weeks with respect to the i^{th} day (Equation 4 and 5).

$$\% \text{ Diff. } d_{i+1} \text{ w.r.t. } d_i = \frac{Step_{i+1} - Step_i}{Step_i} \cdot 100 \quad (4)$$

$$\% \text{ Diff. } d_{i+1} \text{ w.r.t. } \mu_{14gg} = \frac{Step_{i+1} - \overline{Step}_{14gg}}{\overline{Step}_{14gg}} \cdot 100 \quad (5)$$

2.3.4 Statistical analysis

A first preliminary analysis was performed using traditional statistical methods to investigate if a correlation between Wellness Score and number of Steps, DiffStep and Step index existed, considering separate datasets for each user.

Two different approaches were applied:

- Firstly, a linear correlation was searched calculating the Pearson coefficient between Wellness Score and number of Steps, DiffStep and Step index. Pearson's correlation coefficient (ρ) is a measure of linear correlation between two sets of data and is calculated as the ratio between the covariance of two variables and the product of their standard deviations. Given a pair of random variables (X, Y), the formula for ρ is:

  $$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (6)$$

  The correlation coefficient ranges from −1 to 1. An absolute value of exactly 1 implies that a linear equation describes the relationship between X and Y perfectly. A value of +1 implies that all data points lie on a line for which Y increases as X increases, vice versa for a value of -1 Y increases as X decreases. A value of 0 implies that there is no linear dependency between the variables. Thus, the sign of ρ shows the direction of the correlation: negative ρ means that the variables are inversely related while positive ρ means that the variables are directly related and the strength of the correlation increases both from 0 to +1, and 0 to −1 [40].

- Secondly, Spearman coefficients were calculated between Wellness Score and number of Steps, DiffStep and Step index to investigate any non-linear correlations. Spearman's rank correlation coefficient (r_s) is a nonparametric measure of rank correlation and it assesses how well the relationship between two variables can be described using a monotonic function. While Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). Given two sets of data X and Y, the Spearman correlation coefficient r_s is defined as the Pearson correlation coefficient (ρ) between the rank variables $(\mathrm{R}(X), \mathrm{R}(Y))$ as in the following formula:

$$r_s = \rho_{\mathrm{R}(X),\mathrm{R}(Y)} = \frac{\mathrm{cov}(\mathrm{R}(X), \mathrm{R}(Y))}{\sigma_{\mathrm{R}(X)}\sigma_{\mathrm{R}(Y)}} \quad (7)$$

  The sign of the Spearman correlation indicates the direction of association between X (the independent variable) and Y (the dependent variable). A positive Spearman correlation coefficient corresponds to an increasing monotonic trend between X and Y. A negative Spearman correlation coefficient corresponds to a decreasing monotonic trend between X and Y. When X and Y are perfectly monotonically related, absolute value of r_s becomes 1 [41].

After the computation of Pearson and Spearman coefficients, the trend of Wellness Score and number of Steps, DiffStep and Step index have been done for each subject, in order to visualize and compare the trends. This kind of graph, in fact, are more appropriate for displaying the correlation between two variables that have very different orders of magnitude, such as Wellness Score and number of Steps or DiffStep.

2.3.5 Supervised Machine Learning Analysis

Supervised Machine Learning analysis was performed to investigate a methodology that could predict the well-being of an elderly subject using the Steps data acquired by the smartwatch and a daily survey about the self-perception of the health status of the subject. The prediction of the user's well-being was possible by training Supervised Machine Learning algorithms using the wellness questionnaire's answers as reference data.

In fact, functioning of Supervised machine learning algorithms is based on learning a function that maps an input variable (a set of features) to an output variable (label), based on a set of input-output pairs given as examples. So, the machine learning algorithm is trained on a dataset of labelled data, that are called predictors. It works under supervision since it is provided with the actual outcome for each of the training inputs. As the training process goes on, the algorithm improves its capacity to identify the relationships between the two variables. By supplying the supervised algorithm with more and more instances, it becomes able to learn more properly and predict an output more accurately. The success of the classification can be measured by testing the created model with a separate set of examples for which the true classifications are known but are hidden to the classifier. Thus, two sets of input-output pairs are generally provided to the algorithm: training set, and test set. The training set is used to build the classifier model, while the test set is used to measure the accuracy of the classifier to evaluate the performance of the method, i.e., it is a measure on how well it generalizes to unseen instances.

Supervised learning could be distinguished in two main categories: classification predictive modelling problems and regression predictive modelling problems. Classification is the task of predicting a discrete class label while regression is the task of predicting a continuous quantity.

In this work, we faced a classification predictive problem, since the target variable to be predicted was the wellness score, a discrete integer parameter that can assume values equal to 1, 2 and 3. However, even if the target variable could assume three different values, the final dataset, obtained after pre-processing and feature extraction, contained just values equal to 1 (high level of well-being) and 2 (medium level of well-being). Thus, the final ML analysis was represented by a binary classification problem.

Five different supervised algorithms were used in our analysis, in order to compare their performance and identify which is the best one for our application.

Classifiers used were:

- Linear Support Vector Machine (SVM)
- K nearest neighbour (KNN)
- Boosted Trees
- Bagged Trees
- Narrow Neural Network

Support Vector Machine (SVM) is a supervised machine learning algorithm used to solve problems of data classification and regression. Developed by Vladimir Vapnik with colleagues in 1995 [42], the idea behind this model is to perform an optimal data transformation that determine boundaries between data points based on predefined classes, labels, or outputs. Technically, the primary objective of the SVM algorithm is to identify the hyperplane, in a high- or infinite-dimensional space, that has the largest distance to the nearest training-data point of any class, so the maximum margin, and thus that distinguishably segregates the data points of different classes with the lowest generalization error. So, the hyperplane is localized in such a manner that the largest margin separates the classes under consideration, as shown in Figure 8.

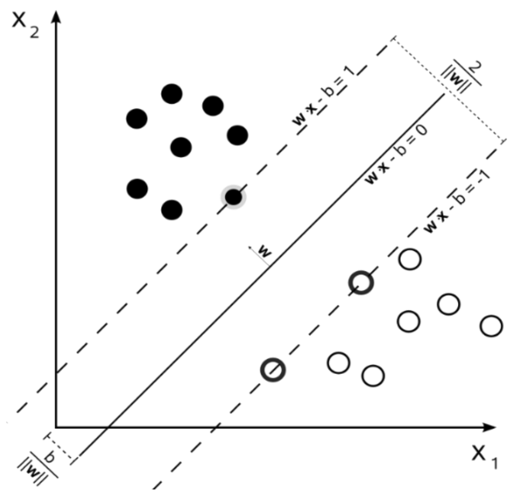

Figure 8: basic idea of SVM classifier

Given a set of training examples, each belonging to one of two categories, a SVM algorithm creates a model that predicts if a new example falls to one class or the other, making it a non-probabilistic binary linear classifier. SVM maps training examples to points in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

A linear SVM, used in this work, refers to the simplest type of SVM, used for classifying linearly separable data. This implies that a dataset can be segregated into categories or classes with a single straight line.

K nearest neighbour (KNN) is a non-parametric supervised learning method first developed by Evelyn Fix and Joseph Hodges in 1951 [43], and later expanded by Thomas Cover[44], which assumes that similar things exist in close proximity. Specifically, it assumes the similarity between the new instances and available data and put the new data into the category that is most similar to the available categories. Similarity is sometimes called distance, proximity or closeness. KNN is used for both classification and regression, even if it is mostly adopted for classification problems. The input consists of the k closest training examples in a dataset, while the output is a class membership. An object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour.

Figure 9 shows an example of KNN classification: the test sample (green dot) should be classified either to blue squares or to red triangles. If k = 3 (solid line circle) it is assigned to the red triangles because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the blue squares (3 squares vs. 2 triangles inside the outer circle).

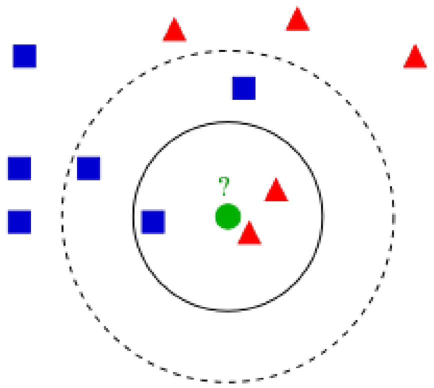

Figure 9: Example of k-NN classification.

Boosted trees represent a ML ensemble meta-algorithm, i.e., a method using multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Boosting is based on the idea that a set of weak learners can create a single strong learner, proposed firstly by Kearns and Valiant [45,46] and then theorised by Robert Schapire in a 1990 paper [47].

Boosting algorithms consist of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. When they are added, they are weighted in a way that is related to the weak learners' accuracy. After a weak learner is added, the data weights are readjusted, known as "re-weighting". Misclassified input data gain a higher weight and examples that are classified correctly lose weight. Thus, future weak learners focus more on the examples that previous weak learners misclassified. Figure 10 shows the idea behind the boosting algorithms.

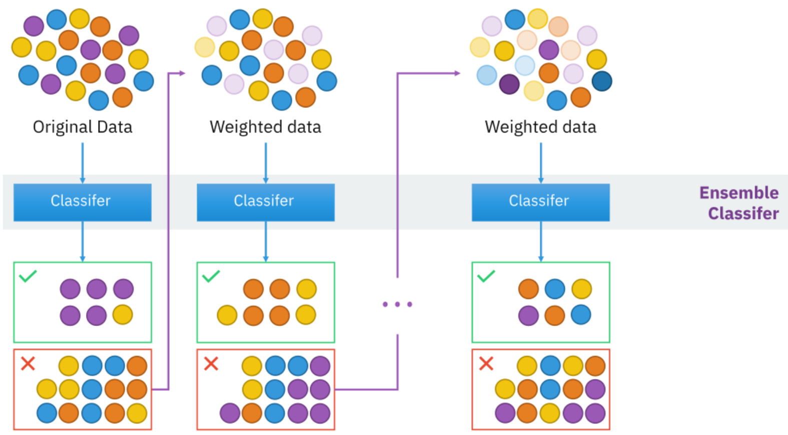

Figure 10: Idea behind the boosting algorithm, consisting of the parallel learners and weighted dataset

Bagged trees (from **B**ootstrap **agg**regat**ed**) represent another machine learning ensemble meta-algorithm used in statistical classification and regression, designed to improve the stability and reduce the variance of decision tree methods. Bootstrap aggregation can be related to the posterior predictive distribution [48].

The technique can be described as follow, as represented in Figure 11. Starting from a standard training set *D* of size *n*, bagging generates *m* new training sets D_i, each of size *n'*, by sampling from *D* uniformly and with replacement. This way, some observations may be repeated in each D_i. If $n'=n$, then for large *n* the set D_i is expected to have the fraction (1 - 1/e) (≈63.2%) of the unique examples of *D*, the rest being duplicates [49]. This kind of sample is known as a bootstrap sample. Sampling with replacement ensures each bootstrap is independent from its peers, as it does not depend on previous chosen samples when sampling. Then, *m* models are fitted using the above *m* bootstrap samples and combined by averaging the output (for regression) or voting (for classification).

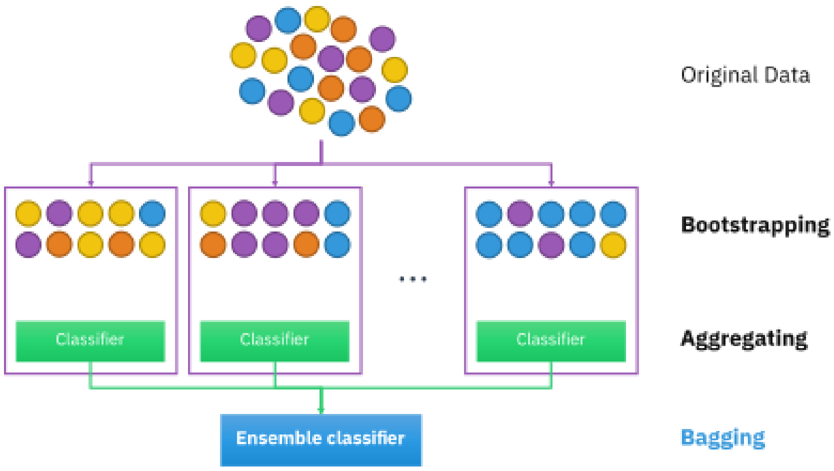

Figure 11: Illustration for the concept of bootstrap aggregating

Neural Networks (NNs) are a class of artificial intelligence algorithms that emerged in the 1980s from developments in cognitive and computer science research. This network is composed of many artificial neurons that are mutually connected. The connections are called parameters and learned knowledge from a data set is then represented by these model parameters. This feature makes a NN model similar to a human brain [50]. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron receives signals then processes them and can signal neurons connected to it. The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold.

Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times. A **narrow neural network** is basically a NN model with a small number of layers.

Figure 12 shows a simplified representation of an interconnected group of nodes, where each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another one.

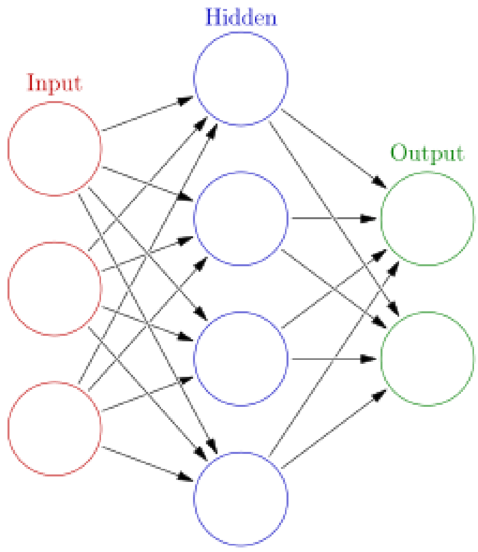

Figure 12: Simplified representation of a Neural Network

A neural network can be trained on a set of input and output (target or label). The weights inside the neural network interact with the input and produce an output. As the network is trained, the weights are updated such that it tries to match the output with the target value. Basically, neural network learns the mapping between input and output.

A NN model typically does not make any prior assumptions about data distribution before learning. This greatly promotes the usability of NNs in various applications. [50]

Once the different models of SML are trained, getting the outputs in form of a label, the next phase is to evaluate their effectiveness using a performance metric. In this study, **accuracy** was computed to evaluate the performance of each algorithm. The classification Accuracy represents the ratio between correct prediction and total number of input samples, so it is the percentage of instances correctly classified. The formula of the classification accuracy is:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \cdot 100 \qquad (8)$$

The procedure used to validate the classification result was the Leave-One-Subject-Out Cross Validation (LOOCV). It represents a cross validation approach that utilizes each individual person as a “test” set. It is a specific type of k-fold cross validation, where the number of folds, k, is equal to the number of participants in your dataset. Thus, training data were represented by all the dataset minus one subject, that turned at each session (number of sessions=number of subjects). Test data, instead, were given by the dataset of the subject that each turn was excluded from the training session. After iterating through and building a model for each fold and then testing it on a unique person for each fold, three different evaluation metrics were obtained, one from each fold. To assess the accuracy of the entire model, mean and standard deviation of these accuracy metrics was calculated for each supervised algorithm.

LOOCV is the most robust way to test a model that contains data on a participant level. However, it is also the most computationally expensive. Thus, it is recommended to validate models built on smaller datasets, also because in this case a standard test/train split may introduce significant bias into the model. Since our dataset was quite limited and divided by participants, LOOCV approach was implemented to obtain a more robust metric result.

The analysis was performed in two different phases.

In a first stage, dataset with the 6 users, derived as explained in paragraph 2.3, was considered. However, user 1 was excluded since his or her well-being questionnaire data were too homogeneous and therefore not useful for the training of the ML algorithms.

Firstly, in order to validate the significance of the new features extracted from Number of steps as explained in paragraph 2.3.2, a comparison between the 5 users' dataset with just 3 features and with the 5 total features was done. In particular, in the very first analysis just Number of Steps, Feedback Score and Step Score (directly derived from Number of Steps) were included in the dataset. Table 12 and 13 show an example of the two different datasets, with 3 and 5 features respectively.

FEATURES				LABEL
USER NUMBER	STEPS	STEP SCORE	FEEDBACK SCORE	WELLNESS SCORE
2	5764	3	1	1
2	2772	3	3	1
…	…	…	…	…
3	2104	3	2	1
3	11673	1	2	2
…	…	…	…	…

Table 12: SML Dataset with 3 features

FEATURES						LABEL
USER NUMBER	STEPS	DIFF STEP	STEP INDEX	STEP SCORE	FEEDBACK SCORE	WELLNESS SCORE
2	5764	-1464	-1	3	1	1
2	2772	-2992	-1	3	3	1
…	…	…	…	…	…	…
3	2104	-6876	-2	3	2	1
3	11673	9569	2	1	2	2
…	…	…	…	…	…	…

Table 13: SML Dataset with 5 features

In the second phase of supervised ML analysis, a new dataset was derived. The objective was to increase the amount of data, since ML algorithms are designed to work with very large dataset in order to increase the learning capacity and thus being able to predict an output more accurately.

First of all, in order to investigate if a higher number of users could be involved in the analysis, Step Score feature was recalculated reducing the width of the time window considered for the computation of the moving average from 2 weeks to 1 week. Figure 13 and 14 show an example of comparison between steps trend in the whole period of trial of user 2 calculating μ with 2-weeks and 1-week window, respectively.

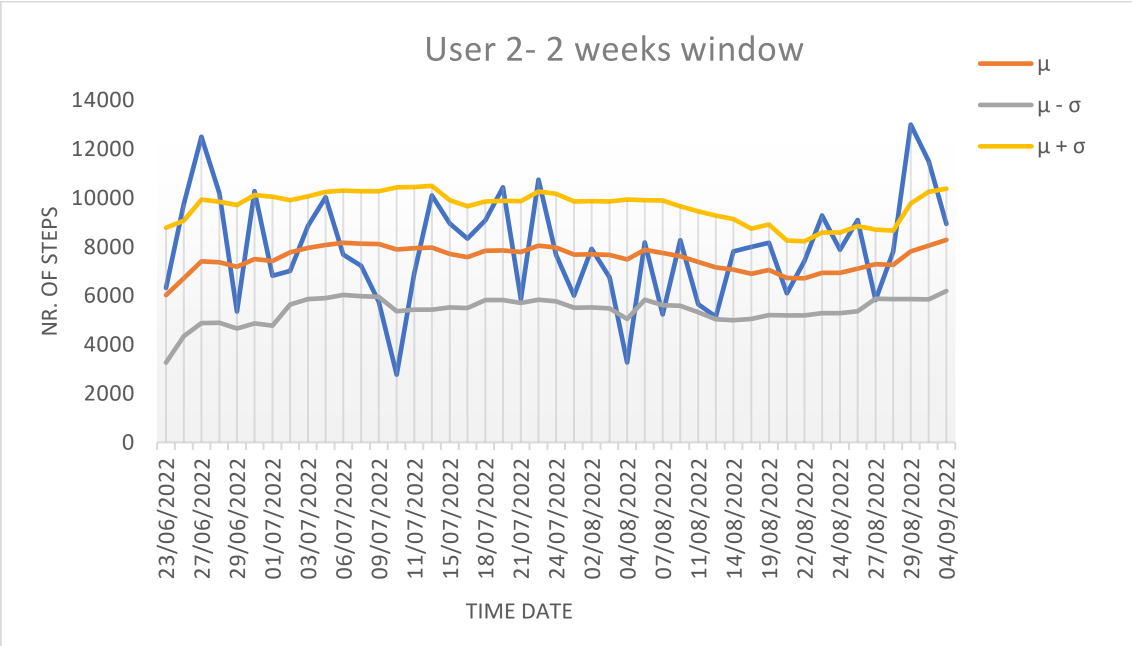

Figure 13: Steps trend of user 2 with a 2-weeks window for calculation of μ

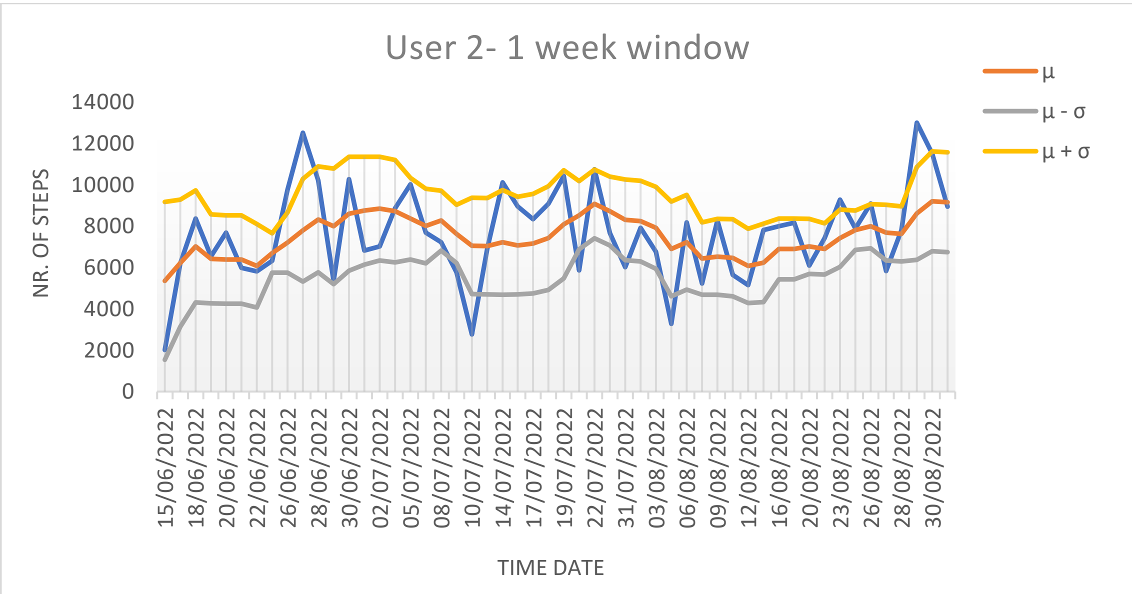

Figure 14: Steps trend of user 2 with a 1-week window for calculation of μ

This way, we passed from 6 users to 8 users available for the analysis. However, user 7 and 8 had to be excluded because their well-being questionnaire data were too homogeneous and thus not useful for the training.

Once performed the supervised ML analysis with this new dataset of 8 users, a data augmentation procedure was applied. In particular, to further increase the dataset, synthetic data were obtained using the approach presented in [51], described in [52] as a conventional data augmentation algorithm and very similar to that one of [53]. For each feature data, samples associated to Wellness Score equal to 1 and Wellness Score equal to 2 were divided into 2 smaller datasets.

Then, starting from those smaller datasets, simulated samples were randomly taken from a gaussian distribution range centred in the mean value of each smaller dataset samples and limited by ± std. Following the approach of [53], the new values obtained from the data augmentation process were added to the original database with a 5-times and 10-times augmentation.

In the end, Supervised ML analysis was performed on the original dataset of 6 users, the 5 times augmented dataset and the 10 times augmented dataset.

2.3.6 Unsupervised Machine Learning Analysis

Unsupervised machine learning algorithms differ from supervised techniques, where models learn to map the input to the target output, since in this case unlabelled data are given to the learning algorithm, leaving it on its own to find hidden structure and patterns in its input. Relationships between data points are gathered by the algorithm in an abstract manner, without any inputs from human beings: neither the target labels nor the internal structure are known by the unsupervised algorithm.

Since values for the output data are unknown, Unsupervised ML methods cannot be directly employed to regression or classification problems. Once the unlabelled dataset is provided to the model, the algorithm analyses the data and eventually finds a classification criterion. It is the model itself that finds the labels to be assigned to the examples [54]. In particular, the task of grouping or dividing datasets with common attributes in order to infer algorithmic relationships is called Clustering.

In this work, an unsupervised ML approach was applied in order to investigate if a measurement method for elderly users' well-being could be derived from the cluster analysis of a dataset based on daily steps and answers to activities satisfaction questionnaire. Answers of wellness questionnaires were not included in the dataset, since unsupervised algorithms must be provided with unlabelled data.

Unsupervised algorithms need large dataset work well, this part of the analysis was performed with the 5 times and 10 times augmented datasets of the six final users, obtained as explained in paragraph 2.3.4. Table 14 shows an example of the total dataset with 6 users used for the unsupervised ML analysis.

FEATURES					
USER NUMBER	STEPS	DIFF STEP	STEP INDEX	STEP SCORE	FEEDBACK SCORE
2	5764	-1464	-1	3	1
2	2772	-2992	-1	3	3
…	…	…	…	…	…
3	2104	-6876	-2	3	2
3	11673	9569	2	1	2
…	…	…	…	…	…

Table 14: Extract of UML Dataset

Both inter-subject (considering the whole dataset with all the subjects together) and intra-subject (differentiating datasets of each subject) analysis were performed. Firstly, each user dataset was analysed separately in order to let the algorithm recognize possible patterns specifically for each user. Then, the entire dataset was provided to the unsupervised algorithms, in order to investigate if any pattern recognition was possible with an inter-subject approach.

Since in the final dataset we had two possible class labels (corresponding to "high" and "medium" well-being), clustering problem was set to divide the dataset in two final clusters. Two different unsupervised algorithms were applied:

- K-means clustering
- Agglomerative Hierarchical clustering

K-means clustering is one of the simplest and most popular unsupervised ML algorithms for clustering operations [55]. It belongs to the category of Exclusive Clustering algorithm. Exclusive clustering is a form of grouping that stipulates a data point can exist only in one cluster.

K-means technique is based on the attempt to find k clusters in a dataset, by minimizing the sum of squared distances between each data point and its assigned cluster center. This algorithm works by choosing k random points as initial cluster centers, then assigning each data point to the closest cluster center. The cluster centers are then updated by taking the mean of the data points assigned to it, and this process is repeated until no data point changes its cluster assignment or a maximum number of iterations is reached.

Figure 15 shows the 'k' centroids changing their location step by step, until no more changes are done.

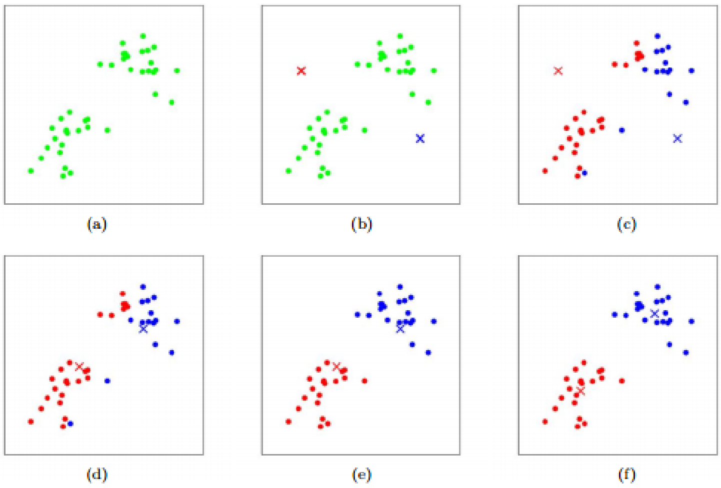

Figure 15: Iterative K centroids positioning

Hierarchical clustering is a method of cluster analysis that seeks to build a hierarchy of clusters. In the agglomerative version, it represents a "bottom-up" approach: the algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects, moving up the hierarchy. The result is a tree-based representation of the objects, named dendrogram [56].

In order to decide which clusters should be combined, a measure of dissimilarity between sets of observations is required. This is achieved by use of an appropriate distance d, such as the Euclidean distance, between single observations of the data set, and a linkage criterion, which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets.

In our analysis, Euclidean distance was used as distance metric and the Average linkage was chosen as linkage criterion, for which the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster.

Figure 16 shows a schematic representation of the Hierarchical Clustering method.

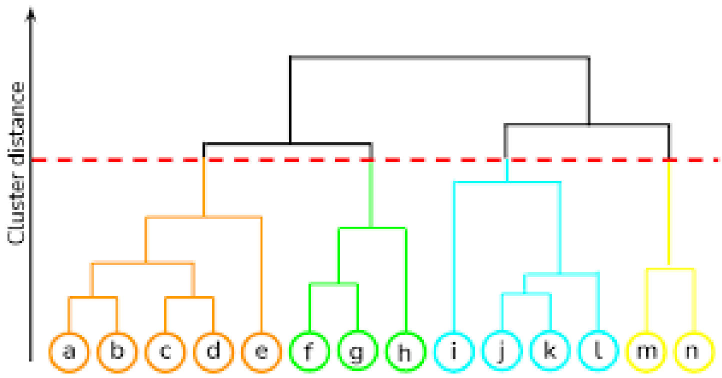

Figure 16: Schematic representation of Hierarchical Clustering method

Once obtained clustering results, the Silhouette score was used to evaluate the quality of clusters. It represents a measure of how well data points are clustered with other data points that are similar to each other [57].

In order to calculate the Silhouette score for each data point, it is necessary to compute the following distances for each observation belonging to all the clusters:

- Mean distance between the element and all other data points in the same cluster. This distance can also be called a mean intra-cluster distance and it is denoted by ***a***.

- Mean distance between the element and all other data points of the next nearest cluster. This distance can also be called a mean nearest-cluster distance and it is denoted by ***b***.

For each data point i, the Silhouette score is calculated using the following formula:

$$S_i = \frac{b_i - a_i}{\max[a_i, b_i]} \tag{9}$$

For each user dataset, the final Silhouette score is then calculated as the mean of all the values.

The score is bounded between -1, corresponding to a wide spread clustering, and +1, which indicates a highly dense clustering. Scores around zero indicate overlapping clusters. Thus, the score is higher when clusters are dense and well separated from each other while a negative score between indicates that the samples might have got assigned to the wrong clusters.

Chapter 3: Results and Discussion

3.1 Results about Impact of the System

Results about impact of RESILIEN-T motivational messages on daily steps are presented in the following graphs for each user.

The first 13 days of acquisition, representing the baseline monitoring period, are highlighted by a red rectangle.

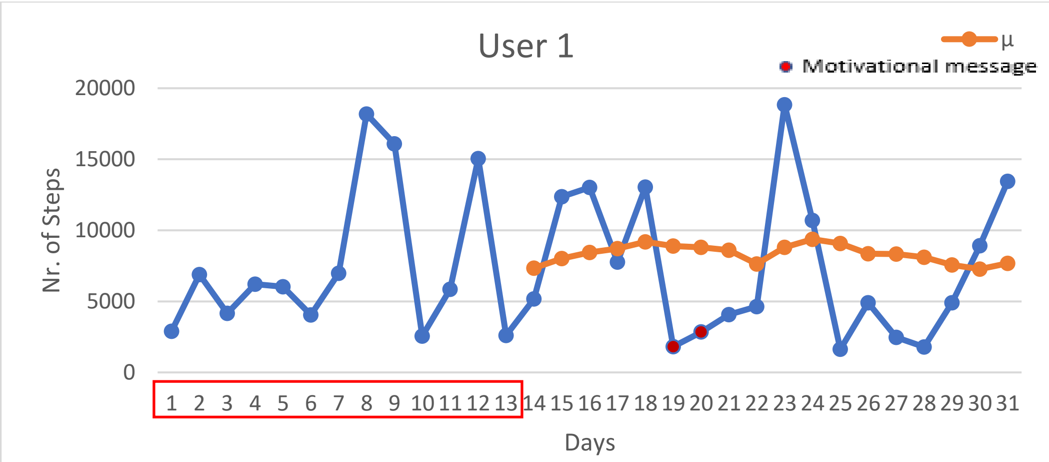

Figure 17: Impact of the system on user 1

As shown in Figure 1, user 1 received two consecutive messages. After the first one, daily steps increase of 57% while after the second one a further increment of 44% is registered, compared to the previous days. However, in both cases the increase is not so high to exceed the average trend of the previous 2 weeks.

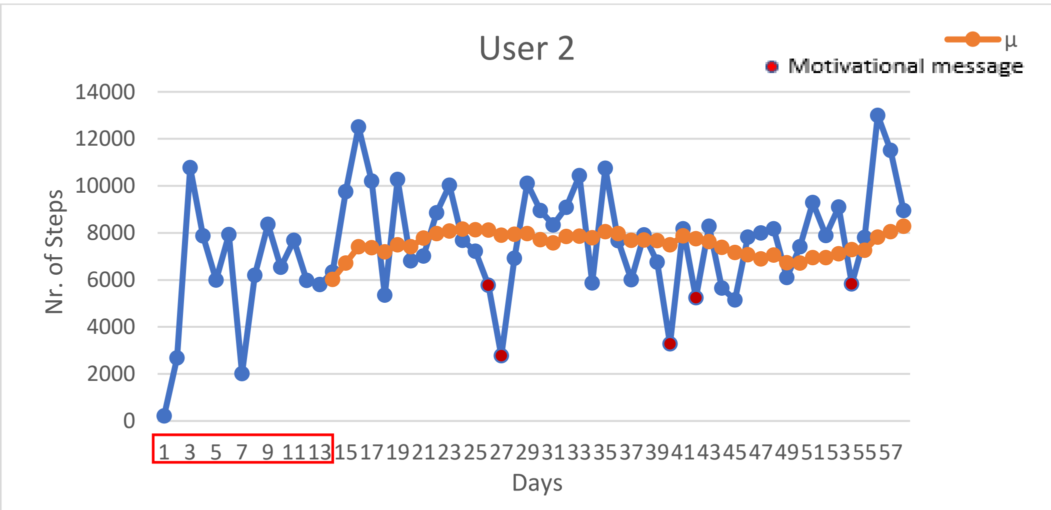

Figure 18: Impact of the system on user 2

Figure 18 shows that user 2 received five motivational messages. The first two are consecutive and in particular after the first one number of steps decrease of 52% while after the second one a consistent increase of more than double is shown, compared to the previous day. After 13 days, other two close messages are sent, after which an increase of 9% and 7% with respect to the average trend is reached, respectively.

The last message is received after 12 days and leads to a further increase of 7% with respect to the average trend.

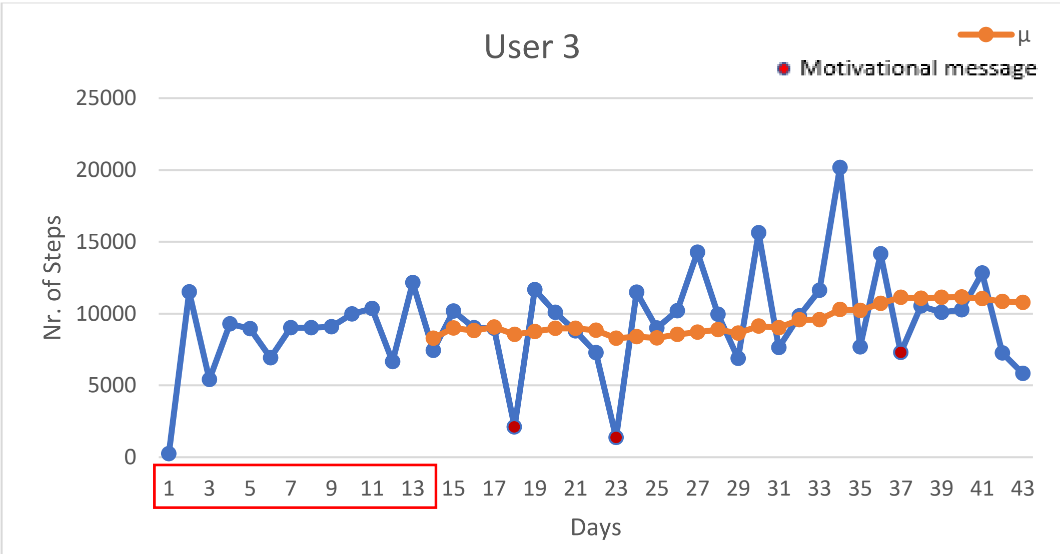

Figure 19: Impact of the system on user 3

User 3 received three messages, leading to an increase of 36%, 39% and 8% with respect to the average trend, respectively (Figure 19).

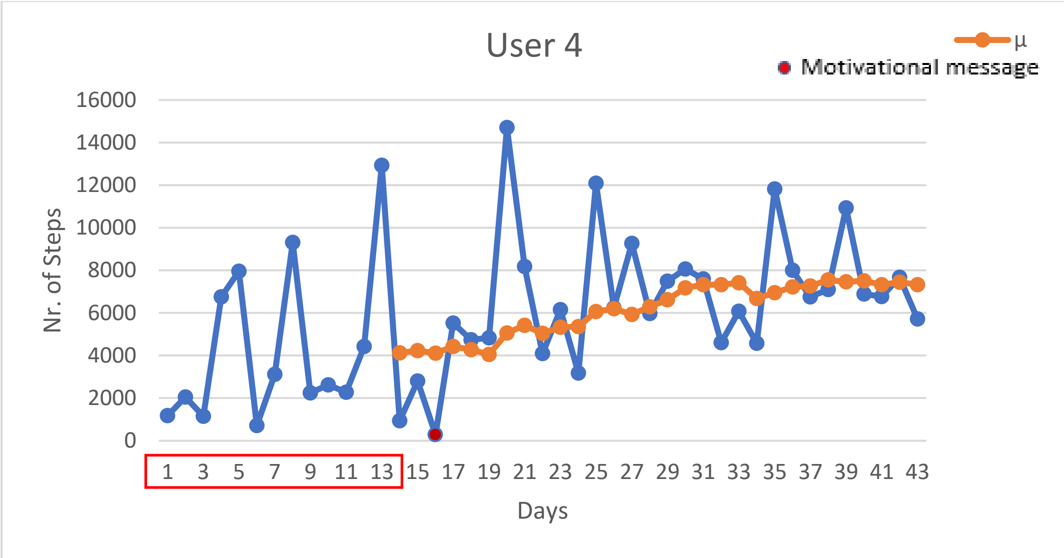

Figure 20: Impact of the system on user 4

As shown in Figure 20, only one message was sent to user 4, following an increase of 35% with respect to the mean of the previous 2 weeks.

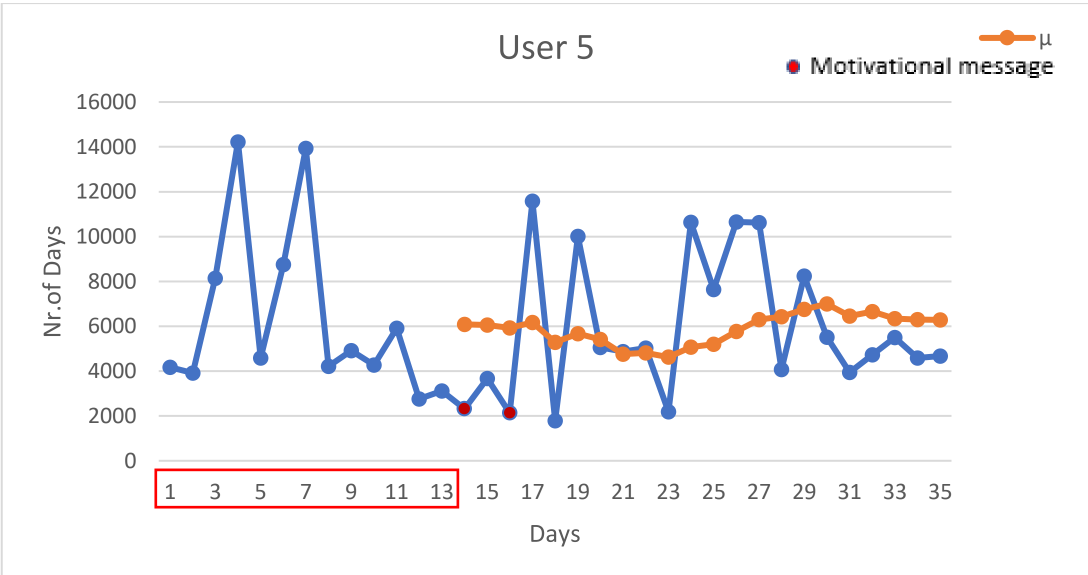

Figure 21: Impact of the system on user 5

User 5 received two close messages: after the first one an increase of 58% with respect to the day before is registered, not enough to realign to the average trend that instead is overcome after the second message, with a consistent increase of 95% (Figure 21).

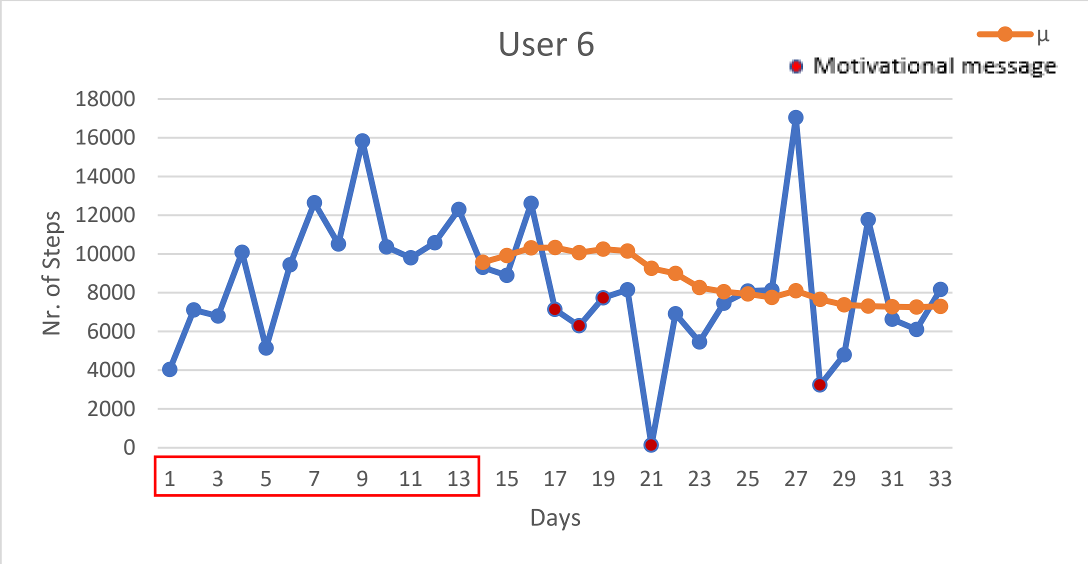

Figure 22: Impact of the system on user 6

User 6 received five messages, the first three of which were consecutive. Except for the second, all of them lead to an increase in the dataset compared to the previous day. However, increments are not sufficient to reach the average trend. In general, after one or two consecutive or close motivational messages users tend to improve their physical activity, exceeding the average trend of daily steps.

3.2 Statistical analysis results

The results of the preliminary analysis performed with the calculation of Pearson and Spearman coefficients are reported, in percentage, in Table 15 and 16, respectively.

	WELLNESS SCORE - STEP INDEX	WELLNESS SCORE - STEPS	WELLNESS SCORE - DIFFSTEP
USER 1	33%	16%	12%
USER 2	NaN	NaN	NaN
USER 3	21%	34%	28%
USER 4	-14%	-49%	3%
USER 5	-2%	-27%	6%
USER 6	15%	34%	16%

Table 15: Pearson coefficients

	WELLNESS SCORE - STEP INDEX	WELLNESS SCORE - STEPS	WELLNESS SCORE - DIFFSTEP
USER 1	38%	31%	31%
USER 2	NaN	NaN	NaN
USER 3	24%	38%	30%
USER 4	-12%	-45%	6%
USER 5	6%	-34%	23%
USER 6	13%	36%	20%

Table 16: Spearman coefficients

For each user, Pearson and Spearman coefficients values do not present relevant differences.

In general, coefficients values do not reveal any significant correlation (linear and non-linear) neither between Wellness score and Step Index, nor between Wellness score and DiffStep, nor between Wellness score and Steps data.

The highest correlation value is reported for User 4, for which Pearson equal to -49% and Spearman equal to -45%. It is a negative correlation, as we expect that increasing number of steps, well-being improves and thus wellness score decreases).

Then, example of trends from user 5 of Wellness score and Steps and of Wellness score and DiffStep are shown in Figure 23 and 24, respectively.

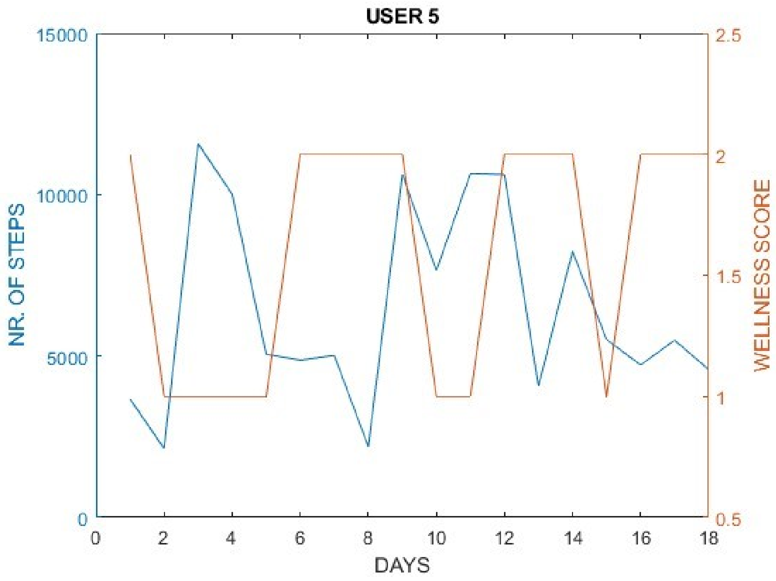

Figure 23: Trend of Wellness score and Steps data of user 5

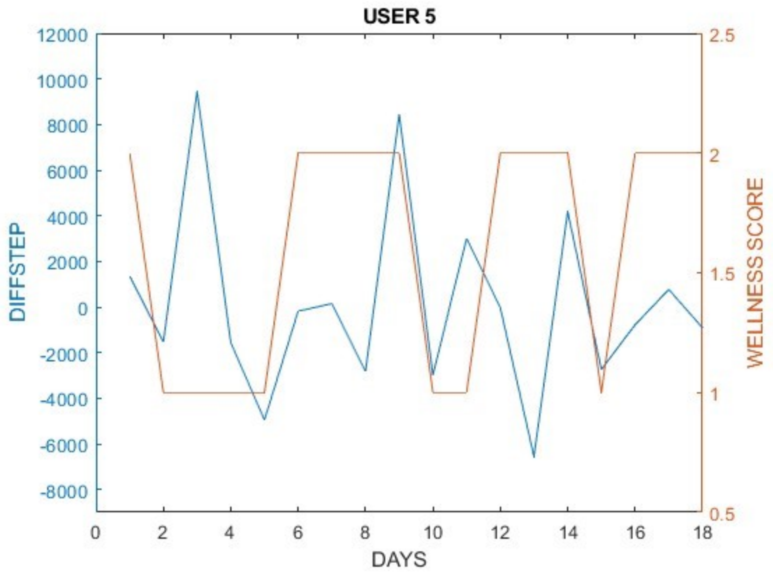

Figure 24: Trend of Wellness score and DiffStep data of user 5

From the trends of user 5 it's possible to notice that in some cases, positive peaks of number of steps value correspond to a decrease in the wellness score, corresponding to a higher well-being. However, for the entire dataset in general it's not recognizable a significant correlation pattern among the analysed variables, in accordance with the results of Pearson and Spearman coefficients, showing that traditional statistical methods are not applicable to this kind of datasets.

3.3 SML Results

As discussed in paragraph 2.3.5, the aim of Supervised Machine Learning analysis was to investigate a methodology that could predict the well-being of an elderly subject, using the wellness questionnaire's answers as target data. Results of this analysis are presented in this paragraph.

The first analysis, including 5 users, was performed on 3 features and 5 features dataset. The accuracy obtained for each algorithm in each session and the mean value with standard deviation of all the sessions for each algorithm are presented in table 17, for the 3 features dataset, and in table 18, for the 5 features dataset.

3 FEATURES	Session 1 Accuracy	Session 2 Accuracy	Session 3 Accuracy	Session 4 Accuracy	Session 5 Accuracy		
Algorithms:						**Mean value**	**Std**
Narrow Neural Network	65,90%	15,40%	19,20%	**72,20%**	**76,50%**	49,84%	26,81%
Boosted Trees	63,60%	26,90%	15,40%	**66,70%**	**82,40%**	51,00%	25,45%
Bagged Trees	63,60%	26,90%	19,20%	**72,20%**	**76,50%**	51,68%	23,87%
Fine KNN	65,90%	19,20%	34,60%	66,70%	**88,20%**	54,92	24,72%
Linear SVM	45,50%	15,40%	15,40%	38,90%	47,10%	32,46%	14,20%

Table 17: Results of supervised analysis on the 3 features dataset

5 FEATURES	Session 1 Accuracy	Session 2 Accuracy	Session 3 Accuracy	Session 4 Accuracy	Session 5 Accuracy		
Algorithms:						**Mean value**	**Std**
Narrow Neural Network	61,40%	19,20%	64,70%	**88,90%**	**82,40%**	63,32%	24,37%
Boosted Trees	61,40%	15,40%	41,20%	**83,30%**	**76,50%**	55,56%	24,74%
Bagged Trees	61,40%	19,20%	52,90%	**83,30%**	**82,40%**	59,84%	23,50%
Fine KNN	72,70%	34,60%	52,90%	**83,30%**	**76,50%**	64,0	17,84%
Linear SVM	45,50%	15,40%	23,50%	38,90%	58,80%	36,42%	15,48%

Table 18: Results of supervised analysis on the 5 features dataset

As shown in the tables, best accuracy was obtained by KNN algorithm for both datasets, with mean value of about 55% with a standard deviation of about 25% in the case of 3 features and of 64% with a standard deviation of about 18% in the case of 5 features.

Thus, a more robust result was obtained for the dataset with a higher number of features since mean accuracy increases by almost 10% while standard deviation decreases of about 7%.

Actually, this is verified for each algorithm: Figure 25 shows that mean accuracy values improve passing from the 3 features dataset to the 5 features dataset, revealing that larger dataset provide better results.

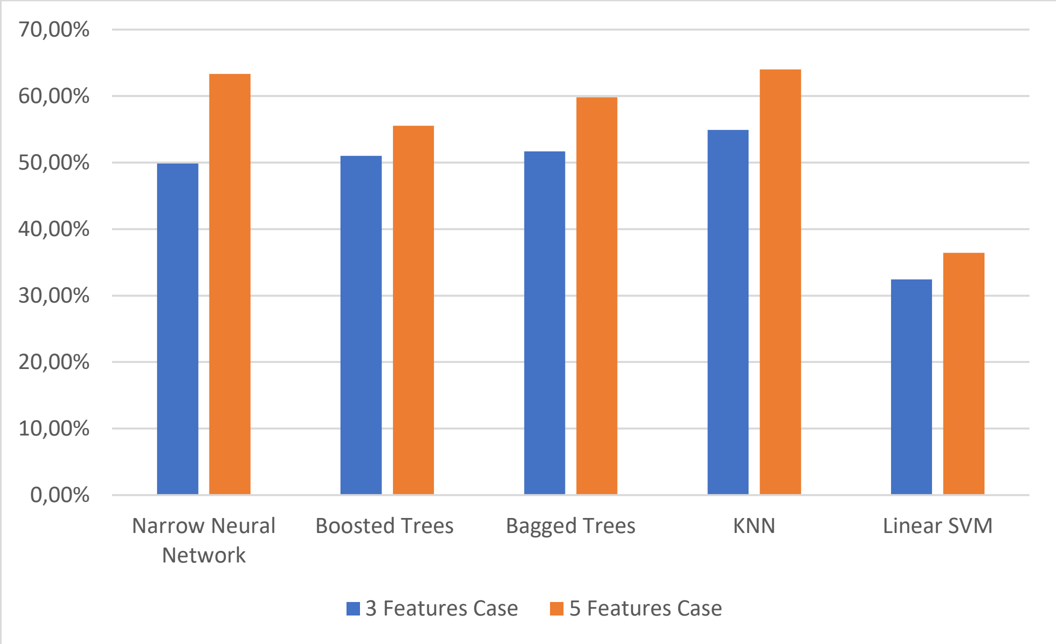

Figure 25: Comparison between mean accuracies of 3 features and 5 features dataset for each algorithm

In general, best accuracies are obtained in session 4 and session 5 for each algorithm and for both the datasets, reaching values of accuracy higher than 80%.

In particular, for the 3 features dataset, best results are obtained in session 5, where the test subject is user 6. The highest accuracy of 88.2% is reached again by KNN algorithm.

While for the 5 features dataset, best results are obtained in session 4, where the test subject is user 5. Here, the best performing algorithm is Narrow Neural Network with an accuracy of 88.9%.

As regards the second analysis, performed on the larger dataset of 6 users firstly original and then 5 times and 10 times augmented, results are shown in Tables 19, 20 and 21.

Original dataset	Session 1 Accuracy	Session 2 Accuracy	Session 3 Accuracy	Session 4 Accuracy	Session 5 Accuracy	Session 6 Acccuracy		
Algorithms:							**Mean value**	**Std**
Narrow Neural Network	56,0%	39,4%	61,9%	72,0%	64,0%	63,6%	**59,5%**	**10,1%**
Boosted Trees	52,0%	27,3%	38,1%	68,0%	56,0%	27,3%	44,8%	15,1%
Bagged Trees	50,0%	33,3%	33,3%	60,0%	60,0%	72,7%	51,6%	14,5%
KNN	54,0%	27,3%	52,4%	60,0%	64,0%	63,6%	53,6%	12,5%
Linear SVM	8,0%	18,2%	23,8%	48,0%	56,0%	45,5%	33,3%	17,5%

Table 19: SML Accuracies on the original dataset

Augmented dataset 5x	Session 1 Accuracy	Session 2 Accuracy	Session 3 Accuracy	Session 4 Accuracy	Session 5 Accuracy	Session 6 Acccuracy		
Algorithms:							**Mean value**	**Std**
Narrow Neural Network	44,0%	27,3%	47,6%	60,0%	56,0%	72,7%	51,3%	14,1%
Boosted Trees	44,0%	27,3%	42,9%	56,0%	68,0%	45,5%	47,3%	12,5%
Bagged Trees	52,0%	33,3%	38,1%	64,0%	68,0%	54,5%	51,7%	12,6%
KNN	50,0%	33,3%	57,1%	60,0%	60,0%	63,6%	**54,0%**	**10,2%**
Linear SVM	8,0%	18,2%	28,6%	44,0%	56,0%	36,4%	31,9%	15,9%

Table 20: SML Accuracies on the 5 times augmented dataset

Augmented dataset 10x	Session 1 Accuracy	Session 2 Accuracy	Session 3 Accuracy	Session 4 Accuracy	Session 5 Accuracy	Session 6 Acccuracy		
Algorithms:							**Mean value**	**Std**
Narrow Neural Network	46,2%	35,2%	40,5%	56,0%	53,6%	51,8%	47,2%	7,4%
Boosted Trees	42,8%	34,8%	36,7%	53,6%	48,4%	48,2%	44,1%	6,7%
Bagged Trees	47,6%	41,5%	39,5%	54,0%	56,0%	47,3%	47,7%	6,0%
KNN	45,0%	41,8%	47,6%	54,0%	52,4%	55,5%	**49,4%**	**5,0%**
Linear SVM	8,0%	18,2%	17,6%	46,8%	57,2%	29,1%	29,5%	17,3%

Table 21: SML Accuracies on the 10 times augmented dataset

Concerning the original dataset, highest value of mean accuracy was obtained by Narrow Neural Network with almost 60% (with a standard deviation of about 10%), while for both 5-times and 10-times augmented dataset best results are obtained by KNN algorithm. However, it’s possible to notice that increasing the amount of synthetic data, mean accuracy decreases passing from 54% (with standard deviation at 10.2%) augmenting 5 times the dataset to 49.4% (with standard deviation at 5%) augmenting 10 times the dataset.

Figure 26 shows the comparison among average accuracies and standard deviations for each algorithm, considering the original and the 5-times and 10-times augmented datasets.

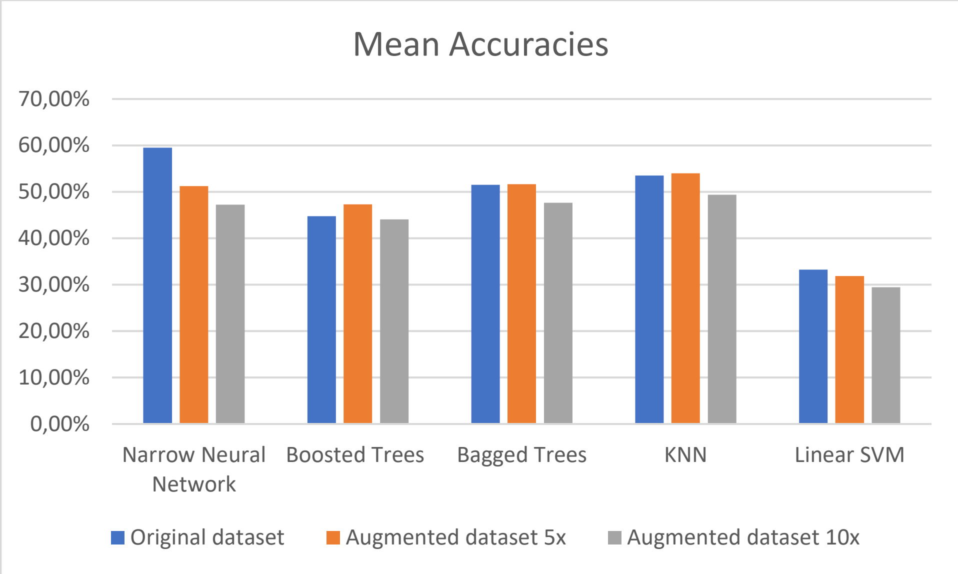

Figure 26: Comparison between mean accuracies of original, 5x and 10x augmented datasets for each algorithm

The dataset augmentation is significant not strictly for improving the accuracy of our system but rather for the better representativeness of the dataset because a larger dataset should better represent the state of reality.

However, applying this augmentation method, the accuracies do not improve so much and in fact at a 10-times data augmentation, results get significantly worse. So probably if we push to a 20-times or even 50-times augmentation we will get much worse. This means that the original dataset cannot support a too large increase in the data.

However, it can be seen that at a 5-times increase in the dataset, the average accuracy slightly improves in the case of Boosted Trees, Bagged Trees and KNN algorithms.

Thus, probably, doubling or tripling the dataset could allow to obtain better results.

This reveals that this kind of dataset can be augmented with the method used, but not too much. Maybe we can a perform a 2-times or 3-times dataset augmentation without making the accuracy worse: by keeping basically the same accuracy or slightly increase it, we can work on a larger dataset and this allows us to better represent the analysed phenomenon.

The most likely motivation lies in the fact that the original data is mostly categorical and the resolution of the system is not so high because the scoring system has been designed to have only three levels of evaluation, in order to make it easier and faster to the users, increasing usability and acceptability. Moreover, the actual final classification is binary, having just class labels corresponding to high and medium level of wellbeing, so there might be a less marked difference in the corresponding feature values and thus more difficult to be identified.

3.4 UML Results

Both inter-subject and intra-subject unsupervised analysis were performed on the 5 times and 10 times augmented datasets, to investigate if any pattern recognition was possible.

Results of Silhouette scores related to K-means clustering are reported in Table 22 while results for Hierarchical clustering are reported in Table 23.

K-means clustering	**Dataset 5x Silhouette Score**	**Dataset 10x Silhouette Score**
User 1	0.54	0.54
User 2	0.53	0.55
User 3	0.53	0.57
User 4	0.48	0.50
User 5	0.52	0.51
User 6	0.56	0.49
All users	0.48	0.48

Table 22: Silhouette Scores for K-means clustering analysis on 5 times and 10 times augmented dataset

Hierarchical clustering	**Dataset 5x Silhouette Score**	**Dataset 10x Silhouette Score**
User 1	0.59	0.51
User 2	0.50	0.63
User 3	0.56	0.49
User 4	0.52	0.53
User 5	0.51	0.80
User 6	0.54	0.51
All users	0.48	0.48

Table 23: Silhouette Scores for Hierarchical clustering analysis on 5 times and 10 times augmented dataset

As shown in the tables, there is not a significative trend in the values obtained passing from a 5-times to a 10-times augmented dataset. Increasing the dataset augmentation, for both the clustering algorithms the silhouette score slightly increases for some users while for others it slightly decreases. Values obtained are very similar one to each other, ranging around 0.5. This reveals a quite good clustering result but no significative improvements are shown increasing the dataset with more synthetic data.

Moreover, trying to identify two distinct patterns with an inter-subject approach leads to a slightly worse result (Silhouette score equal to 0.48 for each case), revealing that best approach is that one of an intra-subject analysis, that allows to better recognize separated clusters in the dataset.

Chapter 4: Conclusions

This dissertation is born inside the framework of the RESILIEN-T project, which, providing coaching solutions to elderly people suffering from MCI and monitoring their daily number of steps, aims to slowing down their cognitive and behavioural decline and improve their quality of life.

The preliminary objective of this work was to analyze the population involved in the project and to investigate if effective improvements in the habits related to the physical activity of users were verified.

Results show that only 6 of the total 96 participants have collected a complete dataset, including smartwatch data. Comparing the trend of the daily number of steps with the average value of the 2 previous weeks, it emerged that in general after one or two consecutive motivational messages users tend to improve their physical activity, exceeding the average trend of daily steps. These results support literature regarding the powerful utility of technology and motivational messages to stimulate a change in behavioural pattern of seniors.

Data collected from Resilien-t system were further analysed: three new features were extracted to create the final dataset for the SML and UML analysis, which aimed at measuring users' well-being from daily number of steps and answers to RESILIEN-T questionnaires data.

Firstly, five supervised ML algorithms have been trained and validated with a LOOCV approach, with the aim of predicting users' well-being using wellness score as label. In the first dataset of 5 users, obtained considering a two weeks sliding window, comparison between accuracies of the 3 features dataset and the 5 features dataset reveal that increasing number of features, accuracy improves. Most probably, by further increasing the number of features the results will improve even more.

The best mean accuracy is obtained by KNN algorithm, with a value of 64%. However, highest values in absolute, for each algorithm and for both the datasets, are obtained in session 4 and session 5, exceeding 80%. In particular, for the 3 features dataset, best results are obtained in session 5, where the test subject is user 6. The highest accuracy of 88.2% is reached again by KNN algorithm. While for the 5 features dataset, best results are obtained in session 4, where the test subject is user 5. Here, the best performing algorithm is Narrow Neural Network with an accuracy of 88.9%. This demonstrates that performance of the algorithm may vary a lot according to the test set used. Thus, using LOOCV is the most robust way to test a model built on this kind of dataset, resulting in a reliable and unbiased estimate of its performance and allowing to go insight the possible great variations from one test subject to another.

Then, a second supervised ML analysis was performed on a larger dataset of 6 users, obtained considering a 1-week sliding window, and its 5 times and 10 times augmentation.

Results show that, applying this augmentation method, the accuracies do not improve so much and in fact at a 10-times data augmentation, results get significantly worse. Thus, this kind of dataset cannot support a too large increase in the data. Anyway, increasing the dataset by collecting a higher amount of data means longer trials and higher risk of drop-out, more users to recruit and more difficulties in managing the pilot.

Thus, the best compromise may be reached by doubling or tripling the data, allowing to work with a larger and best representative dataset, without decreasing or even slightly improving performance of the model. A 5-times dataset augmentation can already lead to a decrease of accuracy for some algorithms.

As regards unsupervised ML analysis, results about silhouette score appear to be quite homogeneous for the 5-times and 10-times augmented datasets, ranging at a quite good value of 0.5. Unsupervised analysis confirms that increasing the dataset with the method used does not lead to a significant improvement in the performance of the model.

On the other hand, performance of the model is not extremely robust since some limitations are present in the study.

The iHealth Wave smartwatch used by participants does not have a very high accuracy, being a low-priced commercial device, and its measurement uncertainty can increase the error in the calculation of the features derived from number of steps and also in the identification of correlation among variables.

Moreover, collection of only steps data due to system errors during data recording has prevented the use of measures that would have enabled the computation of additional features with different information content, e.g., concerning the sleep. Further studies may implement a richer dataset, considering a higher number of features providing a wider information spectrum.

Finally, trials with elderly subjects have not been conducted in a controlled environment, so acquisition of data was not very robust. On the other hand, assisting the user in the completion of daily questionnaires in order to obtain cleaner data would make impossible to collect large databases. Thus, some trade-offs should be accepted by carrying on this types of activities, which quite inevitably tend to be influenced by participants subjective components.

Nevertheless, considering all these limitations, acceptable results have been obtained from our analysis, especially in the first supervised analysis (KNN average accuracy of 64%) and in the silhouette score of unsupervised analysis (being on average higher than 0.5) and these may be improved in further studies, taking into account all the considerations explained above.

Several advantages, in fact, are potentially present in proposing a well-being measurement protocol based on this kind of dataset: It's performed in a totally non-invasive manner, without the need to install domotic sensor networks, starting from data easily collected by elderly users themselves, even suffering from slight cognitive decline, who, living independently at home, have the chance to be continuously monitored and empowered.

Bibliography

1)**World Health Organization: Ageing and health.** https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. [Online] 2018.

2)*World Population Prospects 2019.* **United Nations, Department of Economic and Social Affairs, Population Division.** 2019.

3)**World Health Organization. Disability and health.** https://www.who.int/en/news-room/fact-sheets/detail/disability-and-health. [Online] 2020.

4)Kantarci K, Weigand SD, Przybelski SA, Shiung MM, Whitwell JL, Negash S, et al. Risk of dementia in MCI: Combined effect of cerebrovascular disease, volumetric MRI, and 1H MRS. Neurology 2009;72(17):1519-25.

5)Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia: Meta-analysis of 41 robust nception cohort studies. Acta Psychiatr Scand 2009:119:252-65

6)Barnes DE, Yaffe K.The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol. 2011 Sep;10(9):819-28

7)Francesca Mangialasche, Miia Kivipelto, Alina Solomon and Laura Fratiglioni. Dementia prevention: current epidemiological evidence and future perspective. Alzheimer's Research & Therapy 2012 4:6

8) Justo, R., Ben Letaifa, L., Palmero, C. et al. Analysis of interaction between elderly people and a simulated virtual coach. J Ambient Intell Human Comput 11 , 6125-6140 (2020). https://doi.org/10.1007/s12652-020-01983-3

9) Lane, N.D., Lin, M., Mohammod, M. *et al.* BeWell: Sensing Sleep, Physical Activities and Social Interactions to Promote Wellbeing. *Mobile Netw Appl* **19**, 345–359 (2014). https://doi.org/10.1007/s11036-013-0484-5

10) Montenegro, C.; López Zorrilla, A.; Mikel Olaso, J.; Santana, R.; Justo, R.; Lozano, J.A.; Torres, M.I. A Dialogue-Act Taxonomy for a Virtual Coach Designed to Improve the Life of Elderly. *Multimodal Technol. Interact.* **2019**, *3*, 52. https://doi.org/10.3390/mti3030052

11) Whitmore J (2010) Coaching for performance: growing human potential and purpose: the principles and practice of coaching and leadership. People skills for professionals. Nicholas Brealey Publishing. https://books.google.es/books?id=eTZiP_8dqIYC

12) M. E. Kamali et al., "Virtual Coaches for Older Adults' Wellbeing: A Systematic Review," in IEEE Access, vol. 8, pp. 101884-101902, 2020, doi: 10.1109/ACCESS.2020.2996404.

13) I. M. Albaina, T. Visser, C. A. Van Der Mast, and M. H. Vastenburg, "Flowie: A persuasive virtual coach to motivate elderly individuals to walk," *3rd International Conference on Pervasive Computing Technologies for Healthcare*, 2009: IEEE, pp. 1-7.

14) T. W. Bickmore, L. Caruso, K. Clough-Gorr and T. Heeren . 'It's just like you talk to a friend'relational agents for older adults. Interacting with Computers, vol. 17 no,6, pp. 711-735, 2005.

15) S. Hanke et al., "CogniWin–a virtual assistance system for older adults at work," in International Conference on Human Aspects of IT for theAged Population, Springer, pp. 257-268, 2015

16) C. M. Castro Sweet et al., "Outcomes of a digital health program with human coaching for diabetes risk reduction in a Medicare population"; *Journal of aging and health,* vol. 30, no. 5, pp. 692-710,2018.

17) A. C. King et al., "Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics," *Contemporary clinical trials,* vol. 61, pp. 115-125, 2017.

18) N. O'Brien et al., "Integrating evidence from systematic reviews, qualitative research, and expert knowledge using co-design techniques to develop a web-based intervention for people in the retirement transition," *Journal of medical Internet research,* vol. 18, no. 8, p. e210, 2016.

19) T. Ellis, N. K. Latham, T. R. DeAngelis, C. A. Thomas, M. Saint-Hilaire, and T. W. Bickmore, "Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease," American journal of physical medicine & rehabilitation/Association of Academic Physiatrists, vol. 92, no. 6, p.472, 2013.

20) K. Broekhuizen et al., "An internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial," *Journal of medical Internet research,* vol. 18, no. 4, p. e74, 2016.

21) C. A. Wijsman et al., "Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial," *Journal of medical Internet research,* vol. 15, no. 11, p. e233, 2013.

22) R. Khosla and M.-T. Chu, "Embodying care in Matilda: an effective communication robot for emotional wellbeing of older people in Australian residential care facilities," *ACM Transactions on Management Information Systems (TMIS),* vol. 4, no. 4, pp. 1-33, 2013.

23) G. Cortellessa et al., "ROBIN, a telepresence robot to support older users monitoring and social inclusion: development and evaluation," *Telemedicine and e-Health,* vol. 24, no. 2, pp. 145-154, 2018.

24) D. Kitakoshi, T. Okano, and M. Suzuki, "An empirical study on evaluating basic characteristics and adaptability to users of a preventive care system with learning communication robots," *Soft computing,* vol. 21, no. 2, pp. 331-351, 2017.

25) J. Fan, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A robotic coach architecture for multi-user human-robot interaction (ramu) with the elderly and cognitively impaired," *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN),* IEEE, pp. 445-450, 2016

26) Y. Matsuo, S. Miki, T. Takeda, and N. Kubota, “Self-efficacy estimation for health promotion support with robot partner,” 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE, pp. 758-762, in 2015

27) R. Khosla, M.-T. Chu, and K. Nguyen, "Enhancing emotional wellbeing of elderly using assistive social robots in Australia," *in 2013 International Conference on Biometrics and Kansei Engineering, IEEE,* pp. 41-46, 2013

28) R. Khosla, M.-T. Chu, and K. Nguyen, "Affective robot enabled capacity and quality improvement of nursing home aged care services in australia," *IEEE 37th Annual Computer Software and Applications Conference Workshops,* IEEE, pp. 409-414, in 2013

29) Y. Oida, M. Kanoh, M. Inagaki, Y. Konagaya, and K. Kimura, "Development of a robot-assisted activity program for elderly people incorporating reading aloud and arithmetic calculation," *in Asian Perspectives and Evidence on Health Promotion and Education: Springer,* pp. 67-77, 2011

30) P. J. Gonçalves, P. Torres, and P. Lopes, "Robiho–a robot companion for elderly people's homes," *in Applied Mechanics and Materials, vol. 282: Trans Tech Publ,* pp. 158-161, 2013

31) C. Carlsson, A. Sell, C. Walden, P. Walden, S. Lundqvist, and L. Marcusson, "Go Vendla Go! Creating a Digital Coach for the Young Elderly," in European, Mediterranean, and Middle Eastern Conference on Information Systems, Springer, pp. 204-209, 2017

32) C. Carlsson and P. Walden, "Digital Wellness Services and Sustainable Wellness Routines," *in European, Mediterranean, and Middle Eastern Conference on Information Systems, Springer, pp. 337-352, 2017*

33) M. Báez, F. Ibarra, I. K. Far, M. Ferron, and F. Casati, "Online group exercises for older adults of different physical abilities," in 2016 International Conference on Collaboration Technologies and Systems (CTS), IEEE, pp. 524-533, 2016

34) A. Brandenburgh, W. van Breda, W. van der Ham, M. Klein, L. Moeskops, and P. Roelofsma, "Increasing physical and social activity through virtual coaching in an ambient environment," in International Conference on Active Media Technology, Springer, pp. 487-500, 2014

35) A. Farsi, S. Casaccia and G. M. Revel, "Assessment of normal and abnormal behaviour of people with dementia in living environment through non-invasive sensors and unsupervised AI," 2022 IEEE International Workshop on Metrology for Living Environment (MetroLivEn), Cosenza, Italy, 2022, pp. 71-75, doi: 10.1109/MetroLivEnv54405.2022.9826949.

36) N. Morresi, S. Casaccia, L. Scalise and G. M. Revel, "Machine learning algorithms for the activity monitoring of elders by home sensor network," 2022 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT), Trento, Italy, 2022, pp. 338-342, doi: 10.1109/MetroInd4.0IoT54413.2022.9831726.

37) S. Casaccia et al., "Measurement of Users' Well-Being Through Domotic Sensors and Machine Learning Algorithms," in IEEE Sensors Journal, vol. 20, no. 14, pp. 8029-8038, 15 July15, 2020, doi: 10.1109/JSEN.2020.2981209.

38) N. Kettlewell, R. W. Morris, N. Ho, D. A. Cobb-Clark, S. Cripps, e N. Glozier, «The differential impact of major life events on cognitive and affective wellbeing», SSM - Population Health, vol. 10, pag. 100533, Apr. 2020, doi: 10.1016/j.ssmph.2019.100533.

39) Casaccia, S., Morresi, N., Revel, G.M., Scalise, L. (2022). Measurement Procedure for Daily Motion and Sleep Classification in Aging People Using Smartwatch Data. In: Conti, M., Orcioni, S. (eds) Social Innovation in Long-Term Care Through Digitalization. WS-LTC 2021. Lecture Notes in Bioengineering. Springer, Cham. https://doi.org/10.1007/978-3-031-16855-0_1

40) Haldun Akoglu, User's guide to correlation coefficients, Turkish Journal of Emergency Medicine, Volume 18, Issue 3, 2018, Pages 91-93, ISSN 2452-2473, https://doi.org/10.1016/j.tjem.2018.08.001.

41) E. C. FIELLER, H. O. HARTLEY, E. S. PEARSON, TESTS FOR RANK CORRELATION COEFFICIENTS. I, Biometrika, Volume 44, Issue 3-4, December 1957, Pages 470–481, https://doi.org/10.1093/biomet/44.3-4.470

*42) Cortes, C., Vapnik, V. Support-vector networks. Mach Learn **20**, 273–297 (1995). https://doi.org/10.1007/BF00994018*

43) Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. International Statistical Review / Revue Internationale de Statistique, 57(3), 238–247. https://doi.org/10.2307/1403797

44) T. Cover and P. Hart, "Nearest neighbor pattern classification," in IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, January 1967, doi: 10.1109/TIT.1967.1053964.

45) M. Kearns, "Thoughts on Hypothesis Boosting," Unpublished, Machine Learning Class Project, 1988

46) Michael Kearns; Leslie Valiant (1989). Cryptographic limitations on learning Boolean formulae and finite automata. Symposium on Theory of Computing. Vol. 21. ACM. pp. 433–444. doi:10.1145/73007.73049. ISBN 978-0897913072. S2CID 536357

47) Schapire, R.E. The strength of weak learnability. Mach Learn ***5****, 197–227 (1990). https://doi.org/10.1007/BF00116037*

48) "The bootstrap predictive distribution is considered to be an approximation of the Bayesian predictive distribution". Bayesian bootstrap prediction, Tadayoshi Fushiki, http://dx.doi.org/10.1016/j.jspi.2009.06.007

49) Aslam, Javed & Popa, Raluca & Rivest, Ronald. (2007). On Estimating the Size and Confidence of a Statistical Audit. Proceedings of the USENIX Workshop on Accurate Electronic Voting Technology. 8-8.

50) Yang, Z.R.; Yang, Z. (2014). Comprehensive Biomedical Physics. Karolinska Institute, Stockholm, Sweden: Elsevier. p. 1. ISBN 978-0-444-53633-4. Archived from the original on 28 July 2022. Retrieved 28 July 2022

51) A. Farsi, S. Casaccia and G. M. Revel, "Assessment of normal and abnormal behaviour of people with dementia in living environment through non-invasive sensors and unsupervised AI," 2022 IEEE International Workshop on Metrology for Living Environment (MetroLivEn), Cosenza, Italy, 2022, pp. 71-75, doi: 10.1109/MetroLivEnv54405.2022.9826949.

52) Song K, Park T-J, Chang J-H. Novel Data Augmentation Employing Multivariate Gaussian Distribution for Neural Network-Based Blood Pressure Estimation. Applied Sciences. 2021; 11(9):3923. https://doi.org/10.3390/app11093923

53) Emre Avuçlu, A new data augmentation method to use in machine learning algorithms using statistical measurements, Measurement, Volume 180, 2021, 109577, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2021.109577.

54) Berry, M. W., Mohamed, A., & Yap, B. W. (Eds.). (2019). Supervised and unsupervised learning for data science. Springer Nature.

55) Some Methods for classification and Analysis of Multivariate Observations. ***J. B. MacQueen.*** *1967, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability.*

56) Nielsen, Frank (2016). "8. Hierarchical Clustering". Introduction to HPC with MPI for Data Science. Springer. pp. 195–211. ISBN 978-3-319-21903-5.

57) Peter J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, Volume 20, 1987, Pages 53-65, ISSN 0377-0427, https://doi.org/10.1016/0377-0427(87)90125-7.