Project Title:	**Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions**
Contract No:	690140
Instrument:	Collaborative Project
Call identifier:	H2020-PHC-2014-2015
Topic:	PHC-21-2015: Advancing active and healthy ageing with ICT: Early risk detection and intervention
Start of project:	1 January 2016
Duration:	36 months

Deliverable No: D4.17

FrailSafe Decision Support System (vers. b)

Due date of deliverable:	M29 (1st June 2018)
Actual submission date:	31/05/2018
Version:	1.0
Lead Author:	Evangelia Zacharaki (UoP)
Lead partners:	Spyridon Kalogiannis, Thomas Papastergiou, Konstantinos Deltouzos, Dimitrios Vlachakis, Evangelia Zacharaki, Vasilis Megalooikonomou (UoP), Ilias Kalamaras, Konstantinos Votis (CERTH), Carlo Mancuso, Roberto Orselli (Smartex)

Horizon 2020
European Union funding
for Research & Innovation

Change History

Ver.	Date	Status	Author (Beneficiary)	Description
0.1	16/05/2018	draft	Evangelia Zacharaki, Konstantinos Deltouzos, Spyridon Kalogiannis, Thomas Papastergiou, Dimitrios Vlachakis, Vasilis Megalooikonomou (UoP), Ilias Kalamaras (CERTH), Carlo Mancuso, Roberto Orselli (Smartex)	First draft prepared according to the preliminary deliverable
0.2	23/05/2018	draft	Evangelia Zacharaki, Aimilia Papagiannaki, Thomas Papastergiou, Spyridon Kalogiannis (UoP)	Updates on the Clinical State Prediction engine
0.3	29/05/2018	draft	Konstantinos Deltouzos, Spyridon Kalogiannis (UoP)	Updates on the Risk Assessment
0.4	30/05/2018	draft	Ilias Kalamaras, Konstantinos Votis (CERTH), Roberto Orselli (Smartex)	Internal review comments
0.5	31/05/2018	draft	Konstantinos Deltouzos (UoP)	Updates based on internal review comments
1.0	31/05/2018	final	Evangelia Zacharaki (UoP)	Final version of the deliverable

EXECUTIVE SUMMARY

The aim of work package **WP4** is to develop methods for the offline and online management, fusion and analysis of multimodal and advanced technology data from social, behavioural, cognitive and physical activities, and physiological signals of frail older people and apply them to manage and analyse new data. Results from the analysis of existing and new data will be also used to create user-profiling virtual models of older people.

The main focus of the deliverable **D4.17** is to develop a clinical state prediction engine that simulates the behaviour of an existing patient model, taking into account up-to-date measurements of physiological factors. This is performed by integrating the results of the different sensors and self-reports (combination of all the acquired data) in order to provide the user with the appropriate feedback. These predictions are used either to alarm the patient in order to prevent adverse events, or to be included as supplemental input to the Decision Support System module. Furthermore, artificial intelligence methods are applied for knowledge discovery from data related to user activity and for physiological classification.

DOCUMENT INFORMATION

Contract Number:	H2020-PHC–690140	**Acronym:**	FRAILSAFE
Full title	Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions		
Project URL	http://FrailSafe-project.eu/		
EU Project officer	Mr. Jan Komarek		

Deliverable number:	4.17	**Title:**	FrailSafe Decision Support System (vers b)
Work package number:	4	**Title:**	Data Management and Analysis

Date of delivery	**Contractual**	01/06/2018 (M29)	**Actual**	31/05/2018
Status	Draft ☐		Final ☒	
Nature	Report ☒ Demonstrator ☐ Other ☐			
Dissemination Level	Public ☒ Consortium ☐			
Abstract (for dissemination)	The main focus of the deliverable D4.17 is to develop a clinical state prediction engine that simulates the behaviour of an existing patient model, taking into account up-to-date measurements of physiological factors. The predictions are used either to alarm the patient in order to prevent adverse events, or to be included as supplemental input to the Decision Support System.			
Keywords	clinical state prediction, decision support, generation of alerts, data fusion, probabilistic modelling, alert visualization			

Contributing authors (beneficiaries)	Evangelia Zacharaki, Konstantinos Deltouzos, Spyridon Kalogiannis, Thomas Papastergiou, Dimitrios Vlachakis, Vasilis Megalooikonomou (UoP), Ilias Kalamaras, Konstantinos Votis (CERTH), Carlo Mancuso, Roberto Orselli (Smartex)			
Responsible author(s)	Evangelia Zacharaki		**Email**	ezachar@upatras.gr
	Beneficiary	UoP	**Phone**	+30 2610 996 994

Table of contents

List of figures

List of Tables

LIST OF ABBREVIATIONS AND ACRONYMS

(*in alphabetic order*)

BMI	Body Mass Index
DSS	Decision Support System
eCRF	electronic Case Report Form
MIL	Multiple-Instance Learning
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
UI	User Interface
VPM	Virtual Patient Model
WWBS	Wearable WBan System
WWS	Wearable Wellness System

1 Introduction

Global population has lengthened its life expectancy tremendously over the last few decades. Subsequently, while humans tend to live longer, numerous chronic situations like cardiovascular diseases are found more frequently among them than in the past. In this hypothesis, it is safe to imagine each individual at its late age as a complex patient in need for the consultation of different specialists. Not only do the national health systems consume an excessive amount of resources to care for the average overaged patient, but also, on the other hand, the quality of life of the latter is impoverished. Therefore, prevention appears to be the absolute and crucial solution. Frailty, the one-word term for our body's state of physical and mental weakness could be explained by the parallel existence of several factors implicating each individual's health status [1, 2].

The first part of this deliverable targets the risk assessment platform, which the clinicians can use to monitor the older people and intervene accordingly in case of risks. For this purpose, a FrailSafe Decision Support System (DSS) engine is designed to monitor the older person's vital signs and forward alerts towards caregivers and clinicians. These alerts are generated when irregularities are observed in the VPM aggregated data or when a short-term event is detected by the online data analysis module. Also, in this part of the deliverable we discuss about the role of inflammation as a risk factor for frailty. In the draft deliverable there was an extensive review of recent studies which showed a clear association between inflammation, frailty, and age-related disease.

The second part of the deliverable focuses on the development of a clinical state prediction engine that aims to simulate the behaviour of an existing patient model, taking into account up-to-date measurements of physiological factors. Thus, the system is able to integrate the results of the different sensors and self-reports (combination of all the acquired data) in order to provide the user with the appropriate feedbacks, incentives, and actions. This way several, critical clinical parameters are predicted, such as prediction of frailty level change. These predictions can be used either to alarm the patient in order to prevent adverse outcomes, or to be included as supplemental input to the DSS module. Probabilistic techniques are applied on a provided set of parameters to give probabilistic prediction of specific indicators.

2 Risk assessment

Towards enhancing the risk assessment domain, the FrailSafe DSS engine is designed to monitor the older person's vital signs and forward alerts towards clinicians. These alerts are generated when irregularities are observed in the VPM aggregated data (e.g. on heart rate, respiration rate, and blood pressure) or when a short-term event is detected by the online data analysis components (e.g. fall or suicidal text detection). Thus, the DSS can serve as a risk assessment platform, which the clinicians can use to monitor the older people and intervene accordingly in case of risks.

2.1 VPM parameters update

The VPM parameters table which was composed in M6 for the scope of D4.5, captured the desired parameters that FrailSafe project intended to store for each participant. Since this table was composed in the beginning of the project, several assumptions were made about the recorded parameters from the devices and the data analysis results. Thus, a revision in the VPM parameters was needed because:

- some of the initial parameters are not actually measured during the project (e.g. we don't record the time that the participant spent indoors socializing with friends, or the mean heart rate value while the participant is sleeping),
- other parameters were described abstractly and a clarification was needed (e.g. "Personal Medical History", "Cognitive Questionnaires" or "Short-term Alerts"),
- and other parameters were not described initially but are actually collected in the FrailSafe project (e.g. psychological domain data, outdoor spatial mobility analysis using the GPS data, specific game data etc.).

We have divided all the revised VPM parameters into five distinct tables: (i) demographic data for participants, (ii) clinical data obtained during the clinical evaluations, (iii) data collected by FrailSafe devices, (iv) alerts and events generated by the system, and (v) recommendations and interventions (defined in D2.3). These tables are presented below:

Demographics

Source	Technical/Clinical Parameters	META Tags
eCRF	Participant id Gender Age Contact details (empty for participants of the FrailSafe study)	Type of data Source

Clinical Data

Source	Domain	Technical/Clinical Parameters	META Tags
eCRF	Medical domain (M)	Number of Comorbidities (M) Comorbidity's impact (M, P, s, ψ) Polymedication (M, p, c) Hospitalisations (M) Orthostatic hypotension (M, p) Visual impairment (M, S, p) Hearing impairment (m, S, c)	Type of data Source Domains
eCRF	General Condition Domain (M, ψ)	Unintentional weight loss (self-reported) (M, ψ) Self-reported exhaustion (M, ψ)	Type of data Source Domains
eCRF	Lifestyle domain (P, M, ψ, s)	Smoking (M, ψ, p, s) Alcohol (M, Ψ, S) Physical Activity (P, M, ψ, s)	Type of data Source Domains
eCRF	Physical Condition (P, m, c)	Balance (single foot standing) (P, m) Gait-related task speed (P, c) (Timed Get Up and Go test) Gait - speed 4 m (P, m) Lower limb strength (P, m) Grip strength –dynamometer (P, m) Low physical activity (P, M, s, ψ)	Type of data Source Domains
eCRF	Functional capacity domain (P, m, c, ψ)	Score in ADL and IADL scales Autonomy evaluation in activities of daily living	Type of data Source Domains
eCRF	Nutrition (M, Ψ, c, s)	BMI (M, Ψ, p, c, s) Body fat (M, Ψ, P, c, s) Waist circumference (M, Ψ, P, c, s) Lean body mass (M, P, ψ) Total MNA score (M, Ψ, p, c, s)	Type of data Source Domains
eCRF	Cognitive Domain (C, ψ, m, s)	MMSE scores (C, ψ, m) MoCA score (C, ψ, m) Subjective memory complaint (C, ψ, m, s) Natural language analysis (C, Ψ)	Type of data Source Domains
eCRF	Psychological Domain (Ψ, S, c)	GDS-15 (Ψ, S, c) Self-rated anxiety (Ψ, S, c) Natural Language Analysis (C, Ψ)	Type of data Source Domains
eCRF	Social Domain (S, Ψ, m)	Living conditions (S, Ψ, p, m) Leisure activities (S, Ψ, p, m) Membership of a club (S, Ψ, p, m) Number of visits and social interactions per week (S, Ψ, p, m) Number of telephone calls exchanged per week (S, ψ, m) Approximate time spent on phone per week (S, ψ, m) Approximate time spent on videoconference per week (S, ψ) Number of written messages sent by the participant per week (S, ψ, m, p)	Type of data Source Domains
eCRF	Environmental Domain (S, P, m)	Subjective suitability of the housing environment according to participant's evaluation (S, P, m) Subjective suitability of the housing environment according to investigator's evaluation (S, P, m) Number of steps to access house (S, P, m)	Type of data Source Domains
eCRF	Wellness (Ψ, S, M, P, c)	Quality of life self-rating (Ψ, S, M, P, c) Self-rated health status (M, Ψ) Self-assessed change since last year (M, ψ) Self-rated anxiety (Ψ, S, M, P, c) Self-rated pain (M, P, ψ)	Type of data Source Domains

FrailSafe device Data

Source	Technical/Clinical Parameters	META Tags
WWBS	Min/max/average daily values for: • Heart rate • Heart rate variability • R-R interval in ECG • Breathing rate • Breathing amplitude	Type of data Source Domains Location Room Posture
WWBS	Time spent per day in each posture (Walking, Sit/standing, Stairs, Lying, Transition)	Type of data Source Domains Location Room Posture
FORA blood pressure monitor	Average morning/evening values for: • Systolic pressure • Diastolic pressure	Type of data Source Domains
Mobil-o-graph	Values for: • Pulse Wave Velocity • Augmentation Pressure • Augmentation Index75 • Vascular Resistance • Cardiac Output • Stroke Volume • Cardiac Index	Type of data Source Domains
GPS	Daily values for: • total distance • total duration • total number of steps • radius covered • area covered • average walk speed • total walk time • total stop time • total vehicle time • walk time percentage • vehicle time percentage • stop time percentage • track number • track average distance • track average duration • track maximum distance • track maximum duration	Type of data Source Domains Location
Game 1: Force Analyzer	Daily values for: • Max force • Average max force • Average endurance • Max endurance • Average game duration • Max game duration	Type of data Source Domains
Game 2: Red Wings	Daily values for: • Max force • Average max force • Average endurance • Max endurance • Average score • Max score • Average game duration • Max game duration	Type of data Source Domains

Games 3: Railway	Average/max daily values for: • Score • Distance • Chest mobility • Arm mobility • Movement velocity	Type of data Source Domains
Game 4: Simon	Average/max daily values for: • Hits number • Fails number • Game duration • Sequence length	Type of data Source Domains
Game 5: Memory	Average/max daily values for: • Response time • Game duration • Hit percentage • Fail percentage	Type of data Source Domains
Game 6: Reflex	Average/max daily values for: • Reflects time • Game duration • Hit count • Failure count	Type of data Source Domains
Game 7: Virtual Supermarket	Average/max daily values for: • Game duration • Item time • Item number ratio • Item quantity ratio • Not requested item number ratio • Not requested item quantity ratio • Money ratio	Type of data Source Domains
Game 8: Gravity Ball	Daily values for: • Best time • Gravity deviation • Trajectory deviation	Type of data Source Domains
Game 9: Floating Archery Target	Average/max daily values for: • Accuracy • Hand response time • Head response time	Type of data Source Domains
Game 10: Memory AR	Average/max daily values for: • Visual accuracy • Visual reflex • Memory accuracy • Game duration • Head trajectory	Type of data Source Domains
Beacons	Time spent in each room during each day (Kitchen, Bedroom, Living room etc)	Type of data Source Domains Location Room

Events

Source	Technical/Clinical Parameters	META Tags
Fall detection app (smartphone)	Fall alert	Type of data Source Domains
Social media platform	Suicidal manifestation in text	Type of data Source Domains

WWBS	Heart Rate while Standing/sitting/Lying not in range 50-100 Heart Rate while Walking not in range 60-110 Breathing Rate not in range 12-28	Type of data Source Domains
Blood pressure monitor	Systolic Blood Pressure while Standing/sitting not in range 100-140 Diastolic Blood Pressure while Standing/sitting >90	Type of data Source Domains
eCRF	BMI<18 or >30 (M, Ψ, p, c, s) MMSE<24 (C, ψ, m) MoCA<26 (C, ψ, m) Alcohol consumption >28units/week for men and >21 for women (M, Ψ, S) ADL and/or IADL less than optimal (P, m, c, ψ) >4 medication taken (M, p, c) comorbidities with significant impact >0 (M, P, s, ψ)	Type of data Source Domains
eCRF	Any significant change in selected clinical scores should raise an alert	Type of data Source Domains

Interventions

Source	Domain	Guideline	Recommendation
eCRF	Cognitive domain	If MMSE score is under the normal values	A visit to the neurologist is recommended for providing a more comprehensive evaluation of your cognitive level
eCRF	Cognitive domain	If MMSE score is 24 (cut-off point)	Participation to a cognition enhancement programme is encouraged.
eCRF	Cognitive domain	If MMSE score is 25 and above	To increase your cognitive level in an entertaining way, playing FrailSafe's "Memory", "Simon" "Supermarket" or "Reflex" games frequently is recommended
eCRF	Lifestyle domain	If the person is smoking	• Smoking deteriorates health and weakens the immune system. Consider replacing it with exercise • There are plenty of programs available aiding people to quit smoking. Choosing one of them is encouraged *(maybe add a list per country?)*
eCRF	Lifestyle domain	If reported low or no physical activity or if GPS Logger doesn't show a satisfactory outdoor walking activity (in terms of distance and duration)	• Exercise can not only prevent but reverse frailty. Consider joining an aerobics and resistance exercise program • The following videos provide valuable exercises for home workout. Give them a try. • Walking for at least 2.5 hours per week can increase endurance and benefit your health • Play FrailSafe's "Redwings" game to strengthen your grip
eCRF	Lifestyle domain	If the person is consuming an excessive amount of alcohol (recommendation based on unit reported)	• Alcohol can cause severe damages to the liver. Consider lowering consumption. • There are groups helping people stop drinking alcohol • One glass of red wine per day is the only

			healthy alcohol consumption option
eCRF, FrailSafe devices	Medical domain	If Blood Pressure is consistently high	• Consider visiting your GP for regulating your blood pressure
eCRF	Medical domain	If more than 3 co-morbidities	• Consider medication list review according to pre-defined criteria (e.g. STOPP/START) • Consider modification of prescribed drugs according to the functional status of the patient for minimizing possible effects resulting to frailty
eCRF	Medical domain	If presence of Orthostatic Hypotension	• In order to avoid dizziness from orthostatic hypotension consider getting up from the chair, bed or other sitting posture slowly • If dizzy at any point, sit down, lower your head towards your legs and breathe slowly
eCRF	Psychological domain	If score is greater than 5	• Consider visiting a neurologist or psychologist to discuss about the issues in your life that make you sad • A smile can always make a day better • Meeting with friends is a helpful habit • For clinicians: Consider referral of the person to a psychiatrist, or neurologist for medical, psychological intervention
eCRF	Nutrition	If BMI too high	• Consider a healthy diet for achieving a balanced weight. Eat plenty of fruits, vegetables and protein • Aerobic exercise not only helps cardio function but also helps in losing weight
eCRF	Nutrition	If BMI too low	• A visit to the GP for investigating the low BMI is recommended • An enriched and nutritional diet is recommended. Consider visiting a nutritionist • Resistance exercise helps in increasing muscle mass and it is a perfect way of remaining physically healthy.
eCRF	Nutrition	If malnourished on MNA	• Visit a nutritionist • Resistance exercise increases muscle mass and improves appetite • Investigate cause of malnourishment (for clinicians) • Consider taking nutritional supplements
eCRF	Physical domain	If balance on single foot standing <5 seconds	• Try to hold on to a stable object and stand on single foot for 1 minute. If unsuccessful try it as often as possible until you achieve it. When it is achieved try slowly removing hands from the object.
eCRF	Physical domain	Slow gait speed	• Try walking for 2 minutes every day increasing the time as walking becomes easier

eCRF	Physical domain	Low lower limb strength	• Exercise can help increase limb strength. Consider a home workout
eCRF	Physical domain	Low grip strength	• Suggestion for reviewing medication (if medications taken might be the cause of low grip strength) • Resistance exercise is extremely helpful for strengthening grip thus reducing one of the main symptoms of frailty • Playing FrailSafe's "Red Wings" or "Force Analyzer" games two or three times a week can help in increasing grip strength
eCRF	Physical domain	If low muscle mass and poly-medication exists	• Send recommendation to the clinician to prescribe medication that increases muscle mass

The META tags below are used to describe the data in the DSS API are summarized below:

META tag name	**Possible value(s)**	**META tag name**	**Possible value(s)**
Type of data	• Devices • Clinical • Demographics • Alerts • Interventions	**Posture**	• Sitting/standing • Walking • Lying • Stairs • Transition
Location	• Indoors • Outdoors	**Room**	• Kitchen • Bathroom • Bedroom • Livingroom • Warehouse • Outdoor
Domain (main)	• Medical • General • Lifestyle • Physical • Functional • Nutrition • Social • Cognitive • Environmental • Wellness • Psychological	**Tags (reflecting impact of each item on each of the aspects of frailty)**	• Physical/functional: P dominant, p recessive • Medical: M dominant, m recessive • Social: S dominant, s recessive • Cognitive: C dominant, c recessive • Psychological: Ψ dominant, ψ recessive

2.2 Generation of alerts

The alerts which are stored in the DSS engine are generated by the offline and online data analysis modules as shown in Figure 1.

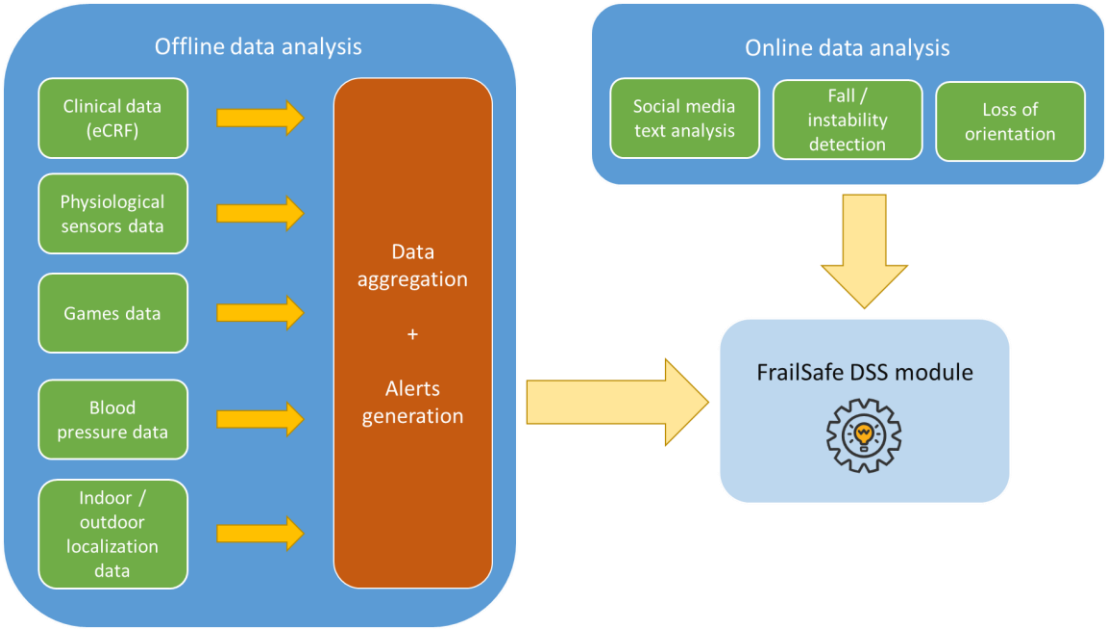


Figure 1: Data aggregation and alert generation from offline and online data analysis.

During the offline data processing, data are being collected from multiple sources and are subsequently summarized with the use of aggregation functions as described in D4.2. Through this procedure the values of the raw and the aggregated data are examined, and alerts are generated if these values are outside of the normal range, as this was defined by the clinicians. The acceptable ranges for the aggregated physiological parameters of older people are presented in Table 1.

Table 1: Normal ranges of measurements per activity.

Source	Measured Entity	Activity	Acceptable Range
WWBS	*Heart Rate*	*Standing/sitting/Lying*	*50-100*
WWBS	*Heart Rate*	*Walking*	*60-110*
WWBS	*Respiration Rate*	*All*	*12-28*
Blood pressure monitor	*Systolic Blood Pressure*	*Standing/sitting*	*100-140*
Blood pressure monitor	*Diastolic Blood Pressure*	*Standing/sitting*	*<90*

While heart rate and blood pressure normal ranges can be applied uniformly to all participants, the respiration rate is something that varies significantly in older people. Currently the above limits are selected based on [22], but in the future they will be personalized for each participant.

The examination of participants' physiological parameters according to these ranges was performed by multiple aspects:

- If the daily average of a participant heart rate or respiration rate falls out of the above ranges while performing an activity, then an alert is generated.
- If the heart rate or the respiration rate of a participant falls out of the normal ranges for a duration of few seconds, then we assume it is an outlier caused by an erroneous measurement. If the duration of abnormal values persists for longer time, then an alert is generated indicating the time period and the average value for the respective period. This type of alert is important, as one participant can e.g. experience normal daily average heart rate but his/her measurements for a specific period of the day could be abnormal.
- If the average daily values of the systolic/diastolic blood pressure fall out of the normal ranges, then an alert is generated for the specific day.
- If the value of the systolic blood pressure for a participant is measured to be under the threshold at any measurement taken, then an alert is generated as it can an indication of hypotension.

Besides the alerts generated from the abnormal values of the physiological data, there are additional alerts generated from the clinical evaluation data which are inserted into the eCRF system. The thresholds for these alerts were defined by the clinicians and are shown in Table 2.

Table 2: Normal ranges of clinical data.

Source	Measured Entity	Acceptable Range
eCRF	*BMI*	*18-30*
eCRF	*MMSE*	*≥ 24*
eCRF	*MoCA*	*≥ 26*
eCRF	*Alcohol consumption (units)*	*≤ 28 for men, ≤ 21 for women*
eCRF	*Medications taken*	*≤ 4*
eCRF	*Significant comorbidities*	*< 1*

The predefined thresholds set above for the alerts generated from the abnormal values of the physiological and clinical data, can be used for all the participants of the FrailSafe study. However, as these values cannot capture the individual characteristics of each participant, the clinician will be able to change and personalize the thresholds accordingly.

The online data analysis components run in real-time and in case they detect any emergency situations they trigger an alarm. The real-time alerts generated are:

a. Social media text alert: It is generated by the social media sensing platform, when it detects suicidal manifestation in a text posted by the participant in his/her social media accounts.
b. Fall detection / instability alert: It is generated by the fall detection android app, when it detects that the participant has fallen or that he/she shows an increased instability.
c. Loss of Orientation alert: It is generated by the respective application, when the participants experiences signs of disorientation in his/her outdoor movement.

2.3 Decision Support System API

The alerts generated by the offline and online data analysis, which are described in the previous section, are sent to the DSS module and are visualized by the DSS UI module. The data stored in the DSS module are also shared with the games platform in order to personalize the games for each user. For the purpose of communication between the DSS and the other submodules of the FrailSafe system, an API was created. At the time that this deliverable is written, the API can be accessed by directing the URL in this address:

http://172.16.2.50:5052/dss/

The access is limited to the submodules that are part of the FrailSafe cloud private subnet (as it is defined in the D1.4). The internal structure of the DSS module and the connection with external submodules are depicted in Figure 2.

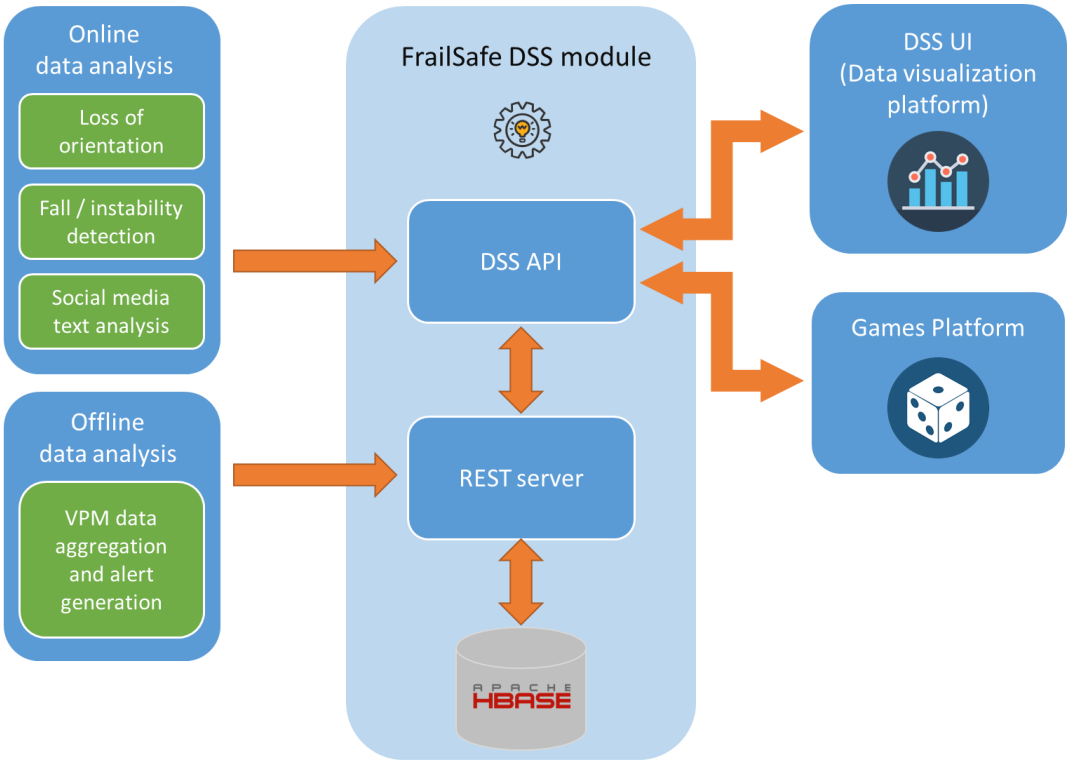


Figure 2: FrailSafe Decision Support System module.

The list of commands that the DSS API allows are the following:

- fetch_ids: Fetches a JSON containing all participant IDs for whom there are VPM data stored.
- fetch_demographics: Fetches a JSON containing all demographic data stored in the VPM.
- fetch_pid_demographics: Fetches a JSON containing demographic data stored in the VPM for one specific participant.
- fetch_all: Fetches a JSON containing all VPM data stored.
- fetch_pid_data/<pid>: Fetches a JSON containing all VPM data stored for one specific participant.
- fetch_pid_date_data/<pid>/<date>: Fetches a JSON containing all VPM data stored for one participant on a specific date.
- fetch_pid_latest_data/<pid>: Fetches a JSON containing the latest VPM parameter values stored for one specific participant.
- fetch_center_data/<cid>: Fetches a JSON containing all VPM data stored for each clinical center.
- fetch_all_interventions: Fetches a JSON containing all interventions stored.
- fetch_pid_interventions/<pid>: Fetches a JSON containing all interventions stored for one specific participant.

- fetch_pid_latest_interventions/<pid>: Fetches a JSON containing latest interventions stored for one specific participant.
- fetch_all_alerts: Fetches a JSON containing all alerts stored.
- fetch_pid_alerts/<pid>: Fetches a JSON containing all alerts stored for one participant.
- fetch_pid_date_alerts/<pid>/<date>: Fetches a JSON containing all alerts stored for one participant on a specific date.
- add_social_alerts/<pid>/<date>: Stores alerts coming from social media analysis.
- add_fall_alerts/<pid>/<date>: Stores alerts coming from fall detection / loss of stability android app.
- add_orientation_alerts/<pid>/<date>: Stores alerts coming from loss of orientation submodule.
- add_dss_parameter/<pid>/<date>: Stores the value of a parameter for the participant which the clinician has changed using the DSS UI.
- add_game_parameter/<pid>/<date>: Stores the value of a game parameter used by the games platform.

The API allows the storing of alerts generated either by the VPM data, or by the real-time data analysis applications (Fall Detection, Loss of Stability, Loss of Orientation and Suicidal Text Detection). These alerts can be classified in three categories, based on the duration of the event detected:

a. Instant alerts, which are produced due to an unusual measurement at a specific time, i.e. elevated blood pressure at a specific time or suicidal manifestation in a text posted in social media.
b. Daily alerts, which are produced when daily user measurements exceed predefined thresholds, i.e. elevated average heart rate for a specific day.
c. Periodic alerts, which are produced due to unusual measurements over a period of time, i.e. elevated breathing rate for one hour.

2.3.1 Example 1: Export all alerts stored in the DSS

Command: GET /fetch_all_alerts

Parameters: No parameters required

Response: Response is a json object containing all alerts

The parameters in the json response are described below:

pid: participant id

date: date of alert

type: type of the alert (allowed: daily, instant, period)

from: beginning of alert period in case of period type alert

to: ending of alert period in case of period type alert

time: time of the alert in case of instant type alert

value: alert value

description: description of the alert

activity: activity performed by participants when alert occurred in case of high/low hr (heart rate) or br (breath rate) alerts (values: 1.0 for standing/sitting, 2.0 for lying, 3.0 for walking, 4.0 for walking upstairs, 5.0 for walking downstairs).

Use case:

Request: curl http://172.16.2.50:5052/dss/fetch_all_alerts

Response:


```
{"Results": [
   {
    "data": {
     "activity": "1.0",
     "avg": "7.61180657333",
     "description": "low_br"
    },
    "date": "20161210",
    "from": "20161210T03:39:25.000",
    "pid": "1042",
    "to": "20161210T03:40:09.000",
    "type": "period"
   },
   {
    "data": {
     "description": "high_avg_diastolic",
     "value": "96.5"
    },
    "date": "20170109",
    "pid": "3001",
    "type": "daily"
 },
 {
    "data": {
     "description": "low_systolic",
     "value": "99"
    },
    "date": "20170123",
    "pid": "3033",
    "time": "14:52:00",
```

```
    "type": "instant"
  }
]}
```


2.3.2 Example 2: Store alert data coming from social media analysis

Command: GET /add_social_alerts/<pid>/<date>

Parameters:

<pid>: participant' s id

<date>: datetime of the alert (in format YYYYMMDDTHH:mm:ss)

Response: Response is a json object containing message about the query result.

Use case:

Request: curl -H "Content-Type: application/json" -X POST -d '{"social_media":"facebook", "description":"suicidal_manifistation", "field1": "value1","field2": "value2}' http://172.16.2.50:5052/dss/add_social_alerts/1060/20170909T14:52:00

Response:


```
{
 "message": "Alert for participant with id=1060 and ts=20170909T14:52:00 has been added",
 "status": 200
}
```


2.4 Alert visualization panel

The alerts are visualized by the Decision Support System User Interface (DSS UI) which is described in detail in D5.5 and is part of the Clinical Web Platform. The Clinical Web Platform is the main interface between users (older persons, families, clinicians, researchers) and the underlying FrailSafe database. Through the DSS UI users can view the collected data, according to their access rights, through intuitive visualizations.

The alert visualization tab of the DSS UI, visualizes the alerts produced for the older persons. Currently, the visualizations are table views of the alerts, where three types of alerts are distinguished as described earlier.

The visualization of these alerts in the DSS UI is differentiated based on the type of user that has logged in. The older person can view the alerts generated for him/her as shown in Figure 3, while the clinician is provided with an alert panel for all the older persons he/she is supervising as shown in Figure 4.

The information displayed for each alert includes the older person's ID, the alert description, the date/time of the alert, as well as any other value related to the alert, e.g. the blood pressure value for a "high blood pressure" alert. In future versions of the DSS UI, the alert visualization tab will also include graphical visualizations of alerts, such as timeline plots, showing the amount of alerts produced through time, etc.

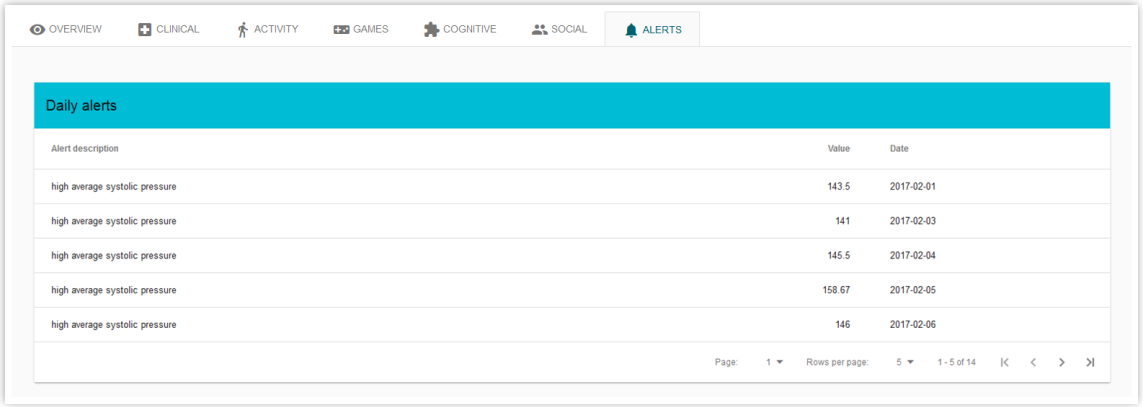


Figure 3: Visualization of alerts in the older person.

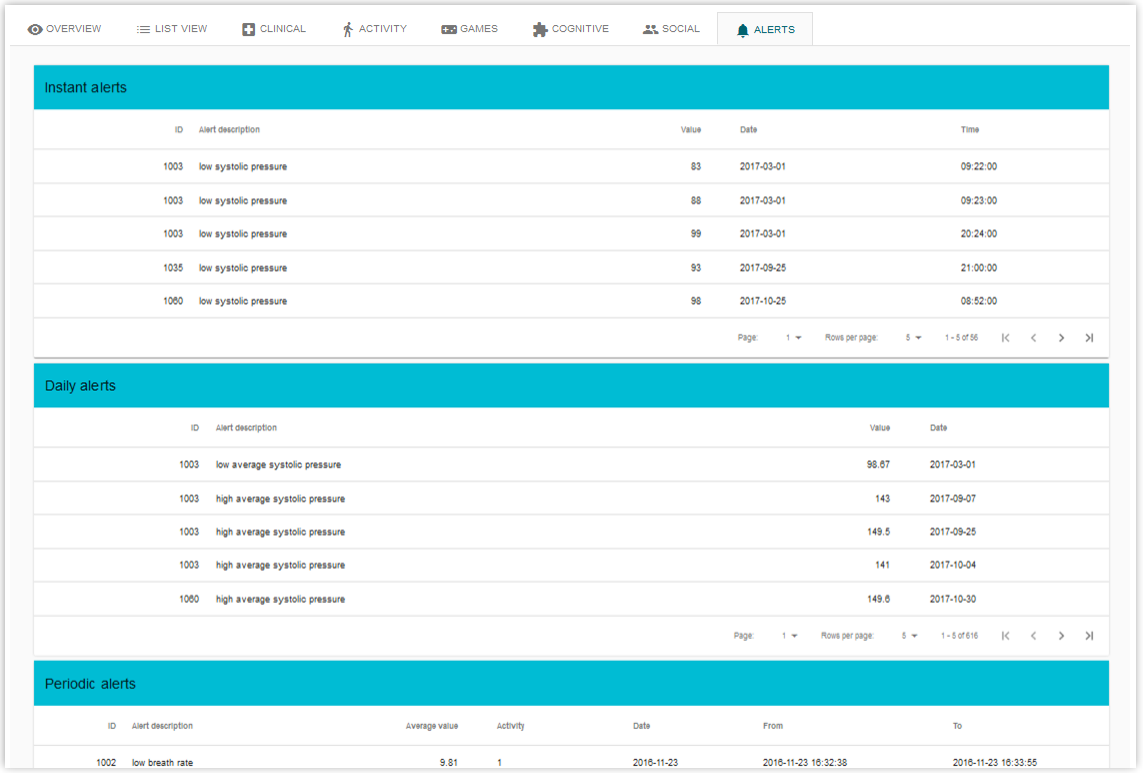


Figure 4: Visualization of alerts in the clinician interface.

The clinician is also allowed to modify the alert generation rules and thresholds, or add new ones, in order to personalize the alerts for specific older persons, as shown in Figure 5.

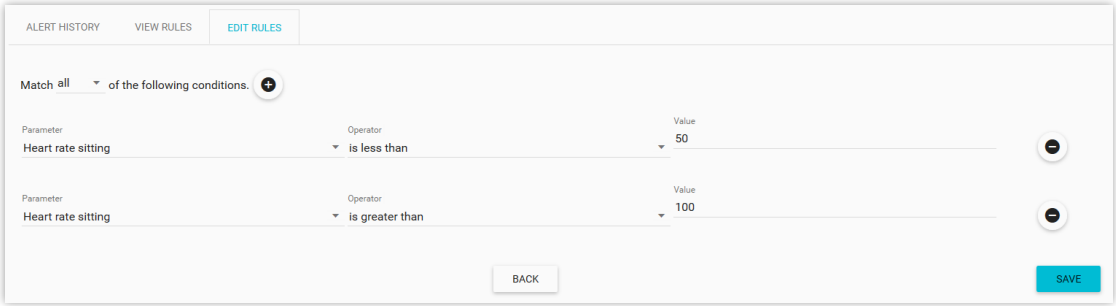


Figure 5: Alert rule editing screen in the clinician interface.

2.5 Role of inflammation as a risk factor for frailty

The section aims to examine the influence of the inflammation on our health. The role of inflammation in the pathogenesis of frailty has been hypothesized, and so far many studies have been performed in order to understand the mechanism of action underlying this association. Recent studies support this hypothesis and show a clear association between inflammation, frailty, and age-related disease. Chronic inflammation is a key pathophysiologic process that contributes to the frailty directly and indirectly through other intermediate physiologic systems, such as the musculoskeletal, endocrine, and hematologic systems. The etiology of frailty is so complex and multifactorial, also including obesity and other age-related specific diseases. This has a semantic effect on quality of life in the later years.

Inflammation is our natural mechanism of maintaining our body's homeostasis [3]. The occurrence of inflammation is inextricably combined with the exposure of our inner systems to potentially harmful microorganisms, well known as pathogens. However, verified proof reveals that in certain cases, this mechanism falsely targets well-functioning cells of the body too, a situation creating the autoimmune syndromes.

Firstly, triggered by the intervention of inflammatory mediators, vasodilation escalates the blood flow. Simultaneously, increased permeability of the blood vessels results in the leakage of proteins and fluid into the tissue. At this stage, it is evident to the human eye that the area has become red, has increased its temperature and has become swollen. The loss of function is believed to occur in order to avoid the feeling of pain. Afterwards, neutrophils and macrophages take action. Neutrophils migrate outside the blood vessels and adhere first into the tissue with the assistance of chemotactic gradients, which enable them to become attached firmly onto the endothelial cells [4]. Phagocytes have non-specific microbe affinity and efficacy and immediately initiate to extinguish the harmful agent by encapsulating it. Next to arrive are the macrophages, which move slower inside the blood vessel, but, induce a more effective impact than their partners. During the apoptosis of neutrophils,

antimicrobial substances, such as NO, OH and H_2O_2, are released into the blood stream destined to destroy the pathological stimuli. However, their effect is not specialized towards each specific factor, meaning that the surrounding healthy tissue may be damaged. In parallel, this process is enhanced by a group of preformed proteins released into the plasma in an inoperative form, the complement system, which destroys the injurious factor without encapsulating it, but, by simply creating pores on the microbe's cellular surface. At this point, it is important to point out that certain microorganisms and particularly the bacteria have adjusted to this strategy by developing an additional protective surface. In these cases, chemical factors, named opsonins[5] make contact with the macrophage and the microorganism and create a complex. C3b (member of the MAC) and CRP (C-reaction protein) belong to this category [6, 7]. The procedure described above concerns cases where the human body becomes exposed to the harmful factor for the first time.

Human body is so resourceful that it has developed a method of "saving" the different types of pathologic agents. This results to even more immediate and acute response if one "saved" agent enters the human body. In this situation, it is again the macrophages who initiate the process. After encapsulating the agent, they divide it into smaller protein molecules. Subsequently, each molecule gets attached to MHC || proteins and becomes exposed to the cellular surface. At this point, the CD4 cells, through a unique receptor specialized for this specific antigen comes in contact with the macrophage. The connection between the two, is also enhanced by the presence of other proteins found on the cell membrane, while, IL-1 and TNF are produced by the macrophage in order to activate fully the CD4 cells.

The CD4 cells, then, release IL-2 which leads to the increase of their population forming clones and the activation of B cells. B cells are responsible for three major roles [8]. Firstly, they mature into plasmacytes that produce antibodies which enter the blood stream. Each antibody is genetically destined to recognize one and only particular antigen. Secondly, a part of B cells population transforms into memory B cells. Moreover, it is possible that they behave like the macrophage, revealing an injurious antigen to the CD4 cell.

The complex of an antibody and an antigen enables the destructive efficacy of the macrophage.

Inflammation appears to be a mechanism triggered by one single incident and to be completed with the elimination of the pathogen. However, in certain cases it may become established in a particular area of the human body and cause continuously the mentioned symptoms.

Nowadays, science has progressed significantly enough to comprehend most of the extremely specific and complicated steps our body takes to protect itself. Thanks to this, we are able to examine how inflammation contributes among others to frailty. The base of the hypothesis is a linkage between high levels of cortisol and the activation of inflammation markers [9]. It is believed that specifically the incidents of chronic inflammation are responsible for the "weakening" of the individual's stamina [10].

Since it has come to our realization that despite our material wealth, increased access to goods like sterilized water, safely preserved food and advanced drugs, the human population faces an even wider range of health issues seemingly relevant to a series of inflammation incidents, we ought to wonder: why is frailty so common among the population?

The question raised above derives from the simple example of the Galapagos Islands. Even though the islands' biodiversity is under threat from several sources, we cannot help but notice that the Galapagos tortoise has managed to become the longest living of its species. A probable explanation would be that the Galapagos Islands have been a secure and isolated environment in general terms. That is, stress-increasing factors have been almost eliminated and each species has been successfully adjusted to its natural habitat [11].

The example described above brings us to a second question: is anxiety valued sufficiently based on the influence it imposes on our quality of living?

Aiming to reach a reasonable conclusion, it is necessary to evaluate in detail each and every aspect of the majority's way of living that is potentially stress provoking. A balanced psychological status may determine the outcome of everyday life. Nevertheless, depression affects an alarming number of citizens from all social strata. Based on WHO, there are 350 million people from all age groups worldwide suffering from major depression symptoms and over 800.000 of them are led to suicide each year. Depression is a mental disorder of which most frequent symptoms consist of uncontrollable sadness, isolation from social activities, inability to concentrate, random emotional fluctuations, constant tiredness and abnormal behaviour [12, 13]. Such symptoms are also intensified by solidarity -which is, in most times, a conscious personal choice-, financial and social discrepancies like bankruptcy and racism and in certain cases, the unfortunate choice of profession along with inappropriate working conditions. According to studies, depression seems to affect the female population more often than the males[14].

Secondly, it is essential to take each person's genetic background into account. It is probable that most of the population have inherited altered biochemical paths encrypted into the DNA, resulting to, for example, the abnormally excessive production of cortisol into the blood stream or even, falsely triggering the inflammation factors after no severe threat [15]. This, automatically, leads to general inflammation incidents that target no particular pathogen. In the first case mentioned, the gene defect would be translated as a benign mass in the adrenal glands while, in the second one, the conditions referring to this situation are the autoimmune syndromes, like rheumatoid arthritis [16].

On the other hand, health wise, cardiovascular abnormalities in combination with chronic high blood pressure, hyperlipidemia and diabetes exhaust the resources of the human body [17, 18]. The simultaneous effect of the conditions mentioned may be found in a vast amount of men over the age of 45 and women over the age of 55 (women can be protected thanks to estrogens) and also be linked with genetic predisposition [19]. Such patients are advised to receive medication for life to

prevent the formation of atheromatic plaque in the blood vessels and to prolong their life expectancy. Such conditions could develop into constriction of the blood vessels in the extremities, sudden ischemic and bleeding strokes, aneurisms etc. However, it is common act that not only do they neglect the regular visits to their specialists, but also, in some cases, they take no action about. Unfortunately, physicians confirm that prevention could be accomplished in condition that citizens are thoroughly informed and educated about the importance of regular check-ups.

The excessive release of cortisol in the blood stream may also contribute to a generalized hormonal alteration. Automatically, the human body will start producing hormones uncontrollably and under no specific mechanisms of action [20]. For example, CRP and PTH will increase, the first one triggering the mechanism of inflammation while, the second one, destructing the bone tissue. This may also lead to loss of appetite and malnutrition. The latter is responsible for not only causing our metabolic rhythms to cease, but it also leads to major muscle tissue loss. It comes as logical conclusion that major bone and muscle tissue loss will result in the so called frailty of the elderly [1].

Furthermore, examining the example of the Galapagos Islands, we ought to point out that throughout history the organisms inhabiting the islands have developed unique ways of surviving [21]. For instance, the iguana of the Galapagos has developed the ability to dive up to 10 metres underwater and feed on sea creatures. Its capacity might be the answer to a previous shortage of food on land, or, a sudden increase in the iguana's population. It is the instinct for survival that led the iguana for food underwater, a plan that proved successful and was later on encrypted into the species' genome. In this situation, it is clear that there is a positive effect of stress provoking situations too, because they contribute to the progression of species.

As a conclusion, cortisol and inflammation may become an ally to the amelioration of human race, but, they also create numerous health issues, sometimes minor, but others, severe which decrease the quality of life.

3 Clinical State Prediction

3.1 Probabilistic models of frailty by Multi-Instance Learning (MIL) techniques

3.1.1 Background

In many real-life applications, collected data come in a format where a single example cannot always be represented as a single feature vector. For example, images depicting several objects, text documents covering several topics or molecules with different conformations having different chemical properties. In all these cases, representing an example as a collection of feature vectors (e.g. patches or segments of an image, paragraphs in a document etc.) is a more efficient way of preserving as more information about an example as possible. On the other hand, this representation needs a more refined level of annotation which means that each of the feature vectors describing an example (i.e. instances) need to be annotated (which is not always possible). Furthermore, there may be instances inside a bag that don't provide any information about the bag's class or even more commonly, instances that are closer related to another class (than the one the whole object belongs to), thus providing misleading information.

To deal with such problems supervised learning techniques have been reformulated and extended to multiple instance learning (MIL) schemes, with the first work appearing in 1997 [23]. In this setting subjects are sets (*bags)* of feature vectors (instances) and labels are provided only for the subjects.

In order to deal with the MIL problem a classifier f_B must be trained which will be able to classify new unseen bags (i.e. collection of instances). According to [24] MIL classifiers can be categorized in three sets: (a) Instance space classifiers, which consider that the discriminative information lies at the instance level, (b) Bag space classifiers, which consider that the discriminative information lies at the bag level and since the bag space is a non-vector space, a distance function between bags *D(X, Y)* or a similarity kernel must be defined that can be used to construct the classifier, and (c) Embedded Space classifiers, where each bag is mapped to a single feature vector which contains relevant information about the whole bag.

MIL basic concepts

Formally a bag is a set $X = \{\overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_N}\}$ of feature vectors called instances, describing an example (e.g. an image or a document). The cardinality of the bags can vary. All instances $\overrightarrow{x_i} \in \mathbb{R}^d$ belong to a d-dimensional space, called instance space. The objective of the MIL problem is to train a model that can predict class label of an unseen bag. In other words, our task is to estimate a classification function $F(X) \in [0,1]$ that provides the likelihood that X is positive. In order to estimate F we are given a training set of M bags and their corresponding labels: $\mathcal{T} = \{(X_1, y_1), (X_2, y_2), \dots, (X_M, y_M)\}$. When required, (such as in the case of instance space classifiers), it is assumed that all instances of a given bag inherit the label of the corresponding bag.

In the MIL framework the label of each bag, as described earlier, is known but the individual labels of each instance of a bag are unknown. From the introduction of MIL in [23] a strong assumption was made regarding the labels of the instances of a bag and the label of the corresponding bag, referred to as *standard MI assumption*. Under this assumption each instance of a bag has a hidden positive or negative class label characterising the instance as positive or negative. The bag is considered to be positive if and only if contains at least one positive instance. Although this assumption is believed to be suitable for many MIL datasets (e.g. the MUSK drug activity prediction problem), alternative MI assumptions are proposed [25] making MIL appropriate for diverse datasets and problems. For example *the collective assumption* proposed in [26] assumes that all instances in a bag contribute equally to the bag's label. Under this assumption instances are assumed to have class labels according to some unknown probability distribution and the bag label is determined by the expected class value of the instances of the corresponding bag.

3.1.2 Problem formulation of frailty prediction

Before formulating properly the problem of frailty status prediction in the context of MIL we will give some intuitions in our context. In the daily life recordings, the tracked activities are neither discrete and predefined (as in controlled experiments), nor the duration of the recordings is the same across subjects. Using the sliding windows technique, we can split the recordings of each individual in (overlapping) time windows (of constant length for all subjects) that describe a local (in time) aspect, and thus we can capture the diverse unknown activities of each subject. On the other hand, the frailty status is a characteristic of the person, which means there is no label information for each "activity" (i.e. each time window). The term "activity" here should not be considered identical with the *Activities of Daily Living* (ADL) defined and automatically classified in the deliverable D4.2. These included characterizations of the subject's motion, such as walking, sitting, laying down etc. Here we generalise the notion of the term "activity" to include additionally to motion (captured by the accelerometer) also other physical conditions (captured by ECG and breathing sensors). However, our goal here is not to segment the biosignals into predefined discrete activities, but to use the activity patterns (no matter what they represent) as clusters of data with reduced intra-cluster variability. This latent space allows to summarize the data and facilitates inference. As activities of each individual are unknown, and not every activity is related to the frailty status of the individual (e.g. recordings while resting are not expected to provide discriminative measurements (i.e. features) for frailty classification) the problem of frailty prediction cannot be formulated as MIL problem using the standard MI assumption (where the bag is considered to be positive if and only if contains at least one positive instance), since we do not really know which activities are related to frailty status. We can assume, furthermore, that there must be patterns of activity that are not discriminating regarding the frailty status and patterns of activity that can be associated to frailty.

In order to formulate the problem of frailty prediction in the MIL framework we represent each subject as a *bag* and the time windows of the recordings as the

instances of each bag. We represent the data as tensors $\mathcal{X}^{(i)}$, i = 1,2,...,nrOfSub which are considered to be the bags, while the slices $\mathcal{X}^{(i)}_{r,:,:} \in \mathbb{R}^{J\times K}$ are considered to be the instances of Bag *i*. Labels are only provided for bags, so $Y_i \in \{0,1,2\}$ are the labels corresponding to non-frail, pre-frail and frail group. Initially instances inherit class labels from the corresponding bag and subsequently class probabilities are updated based on the assumptions that not all instances of a bag are informative in terms for frailty status prediction. The calculated probabilities provide a measure of discrimination and allow us to detect the most informative patterns. Next, we explain the method in more details and assess it based on the currently available recordings.

3.2 Development of MIL framework and application to FrailSafe data

3.2.1 Constructing the training set

In the context of FrailSafe, data are collected from individuals in different sessions, using the two different developed products of Smartex for the measurement of the physiological signals: the WWS (Wearable Wellness System)[1], which was used in the first phase of the project and which we will also call "strap", and the WWBS (Wearable WBan System) which is the new wearable solution, referred to as "vest" here. The WWBS takes its origin from the WWS (Wearable Wellness System)[2], with a further integration of some Inertial Measurement Units (IMUs) in order to have information of higher quality with regards to movement analysis. Together with data on movement, posture and physical activity, it records also data from the heart (a full ECG lead, similar to standard Einthoven DI lead) and respiration.

Signals from 7 different channels are monitored during the sessions:

1. Respiratory raw signal (by the piezoresistive sensor)
2. Acceleration
3. Breathing amplitude
4. Breathing rate
5. ECG Heart Rate
6. ECG Heart Rate variability
7. ECG RR interval

First time synchronization of the channels is performed by interpolating all recordings at 25 Hz, then signal segments of low quality are discarded and frames (e.g. of 1 minute duration) are extracted using the sliding windows technique. The multiple frames of each subject are concatenated in a 3 dimensional tensor $\mathcal{X}^{(i)}$ of dimensionality $I_i \times J \times K$, *i=1, 2,..., nrOfSub*, where I_i is the *number of time windows* available for each subject*, J* is the number of time points corresponding to each time window and *K* is the number of channels monitored. The number of time windows varies across subjects, but *J* and *K are fixed for all subjects.*

[1] http://smartex.it/index.php/en/products/wearable-wellness-system
[2] http://smartex.it/index.php/en/products/wearable-wellness-system

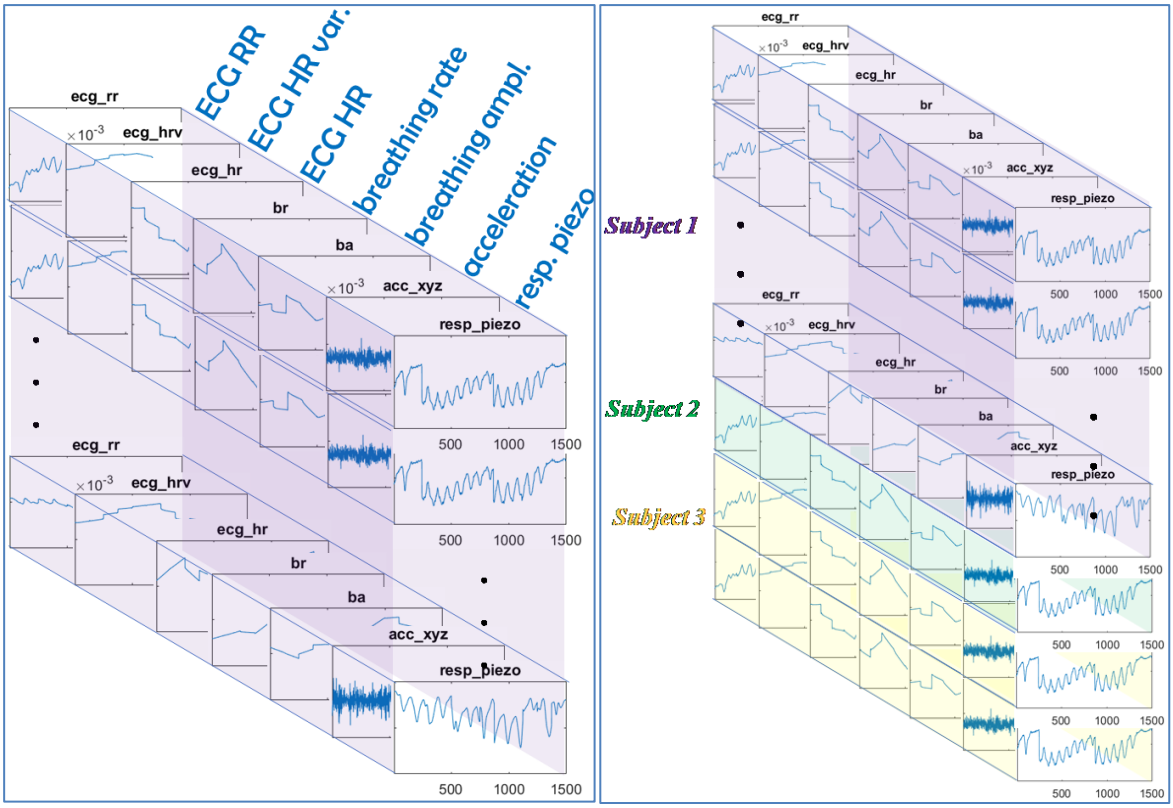


Figure 6. 3d-tensor for one subject (left) and 3d-tensor of all subjects (right).

In order to construct our training set we concatenate all tensors $\mathcal{X}_i$ along the first dimension to produce a new 3D-tensor $\boldsymbol{\mathcal{X}}$ containing all the time windows of all the subjects as is shown in Figure 6. Labels $y_i \in \{0,1,2\}$, (where 0 corresponds to *non-frail*, 1 to *pre*-frail and 2 to *frail* group) are also provided for each subject.

3.2.2 Data cleaning and representation

Besides the preliminary cleaning of the data, based on the quality of the recordings, we performed a series of more elaborate cleaning procedures since we discovered outliers in some measurements that are probably due to bad recordings. These procedures are summarized below.

1. For the channel measuring the breathing rate we consider only recordings with values in the range $[8, 50]$ that are accepted values of breathing rate, and discard all other recordings.
2. For the channel measuring the heart rate we consider only recordings with values in the range $[40, 200]$ as they are accepted values of heart rate.
3. For all other channels we discard values that belong up to the 5% quantile (i.e. the lower 5% of the values) as well as the values belonging to >95% quantile (i.e. the highest 5% of the values)
4. We keep only these time windows, which have in each channel more than 10% non-outlier values.
5. We construct for each instance (i.e. time window) per channel histograms using as bin centres the 30-quantiles.

The output of this procedure is a tensor of dimensions $10503 \times 30 \times 7$ containing at each slice the histograms per channel of each instance. Thus, we have 10503 time windows, 30 bins per instance's histogram and 7 channels.

Since two different devices are employed for acquiring measurements we briefly summarize the number of instances (i.e. time windows), the number of bags (i.e. subjects) per class and per device in Table 2. Some subjects have more than one session and possibly not both with the same device, thus the total number of subjects is smaller than the sum of data with strap and vest.

Table 3 Summary of the data

Data Summary per class			
	Strap Data	**Vest Data**	**Total**
Time windows	7003	3500	10503
Number of subjects	89	27	104
Non-Frail	38	9	43
Pre-Frail	42	13	49
Frail	9	5	12

3.2.3 Feature extraction by tensor decomposition

In order to extract features from the multidimensional data described earlier, we employ the well-known tensor decomposition method PARAFAC. We will briefly resume the PARAFAC method and we will explain how we can extract features for MIL problems.

PARAFAC decomposition is a powerful tool in tensor analysis. In PARAFAC decomposition the aim is to approximate the $I \times J \times K$ tensor $\mathcal{X}$ by a sum of rank-one tensors, referred to as latent factors. Formally

$$\mathcal{X} \approx \sum_{r=1}^{R} \boldsymbol{a_r} \circ \boldsymbol{b_r} \circ \boldsymbol{c_r} \quad (1)$$

where $\boldsymbol{a_r} \in \mathbb{R}^I, \boldsymbol{b_r} \in \mathbb{R}^J, \boldsymbol{c_r} \in \mathbb{R}^K$ are the columns of the factor matrices $\boldsymbol{A} \in \mathbb{R}^{I \times R}, \boldsymbol{B} \in \mathbb{R}^{I \times R}, \boldsymbol{C} \in \mathbb{R}^{I \times R}$ respectively, "$\circ$" denotes the outer product and R is the rank of the tensor. Element-wise the PARAFAC decomposition can be written as $x_{i,j,k} \approx \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$, $i = 1,2, \dots, I, j = 1,2, \dots J, k = 1,2, \dots, K$.

Sometimes PARAFAC decomposition can be written as

$$\mathcal{X} \approx \sum_{r=1}^{R} \lambda_r \boldsymbol{a_r} \circ \boldsymbol{b_r} \circ \boldsymbol{c_r} \quad (2)$$

where λ_r illustrates the significance of each latent factor. Figure 7 depicts in a graphical way the PARAFAC decomposition.

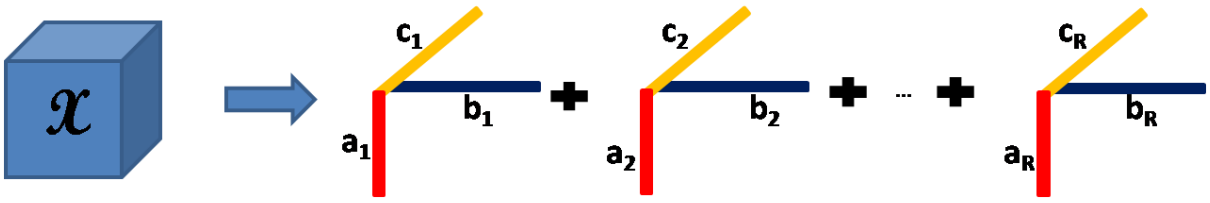


Figure 7 The PARAFAC decomposition

Our goal is to extract features for each instance (i.e. time window) $x_{i,:,:} \in \mathbb{R}^{J \times K}$, where $x_{i,:,:}$ is the i-th slice of tensor $\mathcal{X}$. Each instance can be written as $x_{i,:,:} \approx \sum_{r=1}^{R} a_{ir}(b_r \circ c_r)$, meaning that each instance is a linear combination of $\boldsymbol{b_r} \circ \boldsymbol{c_r}, r = 1,2,\dots,R$, with coefficients $a_{ir}, r = 1,2,\dots,R$. Thus, we can choose as features, for describing an instance, the coefficients $a_{ir}, r = 1,2,\dots,R$, that is the i-th row of factor matrix **A**. Furthermore, we can normalize the columns of **A** and choose as features the matrix $\widetilde{\boldsymbol{A}}$, where $\boldsymbol{A} = \widetilde{\boldsymbol{A}}\, diag(\lambda_r)$, and $\lambda_r, r = 1,2,\dots,R$ are the scaling factors of each column.

3.2.4 One class-SVM for frailty status prediction

Here we will describe our MIL approach for training a model to predict the frailty status based on the data collected from FrailSafe. Firstly, we will give the concept and some intuitions that led us to formulate this approach and subsequently we will describe our methodology.

As aforementioned, in the setting of FrailSafe, subjects wear devices and perform tasks in their daily life that not all of them associated with their frailty status. As a consequence, we need to be able to distinguish instances that are more informative for discriminating the frailty status of the subjects and relying only on them to predict the subject-wise labels.

Figure 8 depicts the class boundaries of the three classes (Non-Frail in green, Pre-Frail in blue and Frail in red) using the three significant features, extracted as described in the previous section. We can observe that only a part of the instances can be informative in terms of predicting the frailty labels and these are the instances that lay in the region of each class that does not overlap with other classes. As this figure is a 3-dimensional projection of our 30-dimensional features it is only indicative and obviously does not provide any proof.

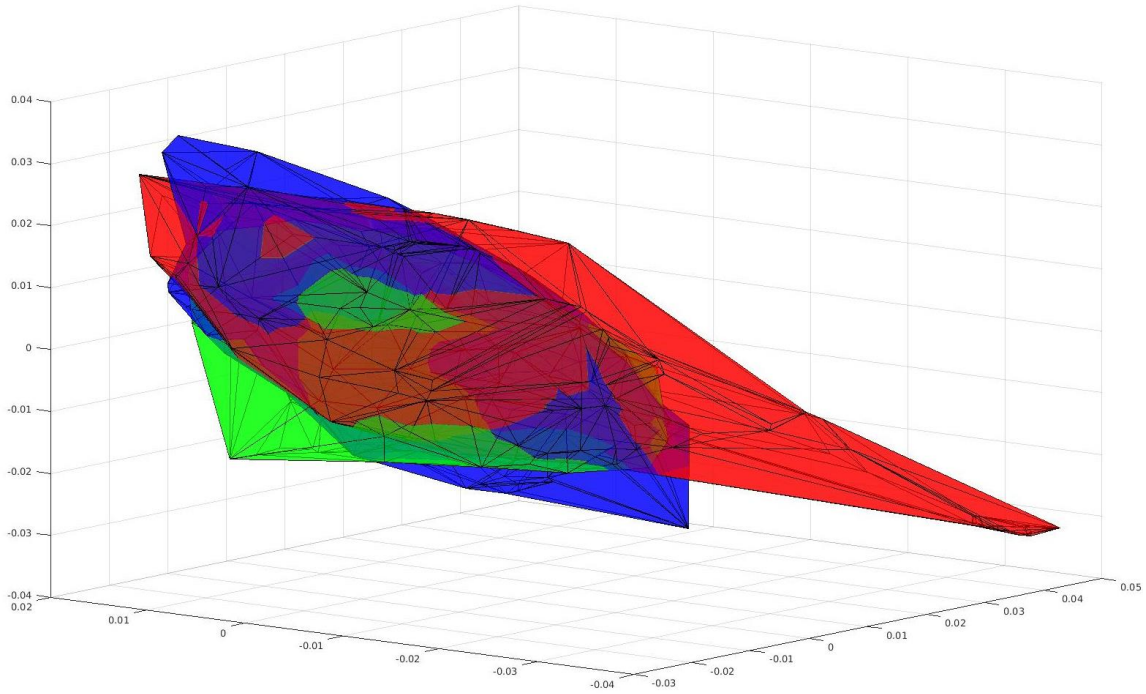


Figure 8 The class boundaries of instances for the 3 significant features: Non-Frail in green, Pre-Frail in blue and Frail in red.

In order to deal with the problem of searching the informative instances and rely on them for predicting bag labels we train 3 distinct classifiers for each class separately to learn class probability distributions and then we annotate as informative instances only these that have a clear classification based on all the three estimated probabilities. With the term clear (i.e. not ambiguous) classification we mean that there are not conflicts in the prediction, i.e. only one classifier recognizes the instance as positive for its class, with the other two classifiers labeling the same instance as negative for their class.

One-class SVM

In our approach, features are extracted using the PARAFAC decomposition method discussed in a previous section. In order to maintain consistency of our model and factor matrices **B** and **C** (both for training and test set), we concatenate along the first dimension the training and test tensors ($\mathcal{X}_{train}$, $\mathcal{X}_{test}$) to a model tensor $\mathcal{X}$ and perform PARAFAC decomposition to $\mathcal{X}$. The training features are the rows of factor matrix **A** that correspond to $\mathcal{X}_{train}$ while the test features are the rows of **A** corresponding to $\mathcal{X}_{test}$.

Subsequently we train three one-class SVM classifiers each one on instances of each of the classes, using Gaussian kernels. In that way, each classifier is trained to discriminate one of the three classes. We then use each of the above classifiers to predict instance-level labels on the test set. In this stage we have 3 distinct predictions for each instance, one from each classifier.

In the next step we want to discover the informative instances. An informative instance can be defined as an instance that has been classified from exactly one

classifier to belong to the class, which the classifier is trained to recognize, and classified by the other two classifiers as not belonging to their corresponding classes. Using these extracted informative instances, we will predict the bag-level labels.

For predicting subject-wise (i.e. bag-wise) labels we rely on the informative instances that we found in the previous step. The label decision is made by counting by a majority voting scheme, i.e. the class assigned to each subject is the one with the most votes. If no informative instance is found, a label of -1 is assigned to the subject meaning no decision could be made. The full algorithm is shown in Table 2.

Table 4 Fusion of one class SVM models in MIL setting

Algorithm: 3 one class_SVM for frailty prediction

Input: training instances' feature X_{tr}, training labels Y_{tr}, test instances' feature matrix X_{ts}, test subject ID's $SubID_{test}$,
Output: predicted subject level labels Y_{ts}

1. Extract positive training examples of each class P_0, P_1, P_2
2. Train 3 one-class SVM models SVM_0, SVM_1, SVM_2, for recognizing classes 0, 1 and 2, using X_{tr} and P_0, P_1, and P2 respectively.
3. **for** i=0 to 2
   using SVM_i and X_{ts} predict the labels $Y_{ts}^i \in \{-1,1\}$of the test set
   end **for**
4. **for each** subject i
   a. find the instances of each subject ***Inf_Inst_i*** that satisfy the equation $\sum_{i=0}^{2} Y_{ts}^i = -1$ (i.e. the informative instances)
   b. assign to each subject the class label of the most predicted class out of ***Inf_Inst_i.***
   c. **if** ***Inf_Inst_i*== *empty***
      do not assign any label to this subject
      end if

end for

3.2.5 Results

We evaluated this method separately on data collected by the FrailSafe devices. In the next section we will use the term "vest" for the "Wearable WBan System (WWBS)" and the term "strap" for the "Wearable Wellness System (WWS)". We conducted 5-fold cross validation and we report the average testing accuracy over all folds. The features were extracted as reported using the PARAFAC decomposition; the rank of the decomposition was chosen to be 30. We assess performance on normalized features (matrix **A** as presented in equation (2)) as well as in original features (matrix **A** as presented in equation (1)). In the first case the λ parameter (the parameter that denotes the significance of each feature) is not taken into account. In contrast, in the second case (i.e. not normalized features) the λ parameter is incorporated into the feature matrix **A** so the features are scaled by their significance value as calculated via the PARAFAC decomposition (see section 3.2.3).

In Table 5 we can see the obtained average accuracy across 5 folds based on cross validation for classification of the frailty status into 3 classes. It can be observed that the accuracy is 48.7% for the vest data when features are not normalized and 46.14% for the strap data, while for normalized features the accuracy is 40.7% for the vest data and 40.6% for the strap data. We observed that when using not normalized features better performance in acquired for both the strap and vest recordings. The data from strap (WWS) were acquired in the first phase of the project when the vest was not available yet and their collection will not continue. Therefore, the accuracy of the method on the data from WWBS is more important. Although the classification accuracy (48.7%) seems low, it is slightly increased in comparison to random guess (33.3%) and also it is expected that the availability of more data samples and more variables (from the rest of the FrailSafe devices and clinical scores) will help to improve prediction.

Table 5 Average 5-fold Cross Validation accuracy

	Data from strap (WWS)	**Data from vest (WWBS)**
Normalized Features (matrix A from eq. (2))	40.6%	40.7%
Not Normalized Features (matrix A from eq. (1))	46.14%	48.7%

3.3 Convolutional Neural Networks for prediction of frailty status

Our previous methods for classification or modelling, presented in previous deliverables of WP4, generally relied on the usage of domain specific features normally selected based on literature or by clinical experts. Finding the best features was the subject of a lot of research and the performance of the classifier was heavily dependent on their quality. The advantage of deep neural networks and especially convolutional neural networks (CNNs) is that they can learn such features by themselves, reducing the need for human experts. A recent review on deep learning techniques for time series analysis can be found in [27].

Building upon our previous work on deep CNNs [28], we aim to implement a deep architecture that employs convolution and pooling operations to capture the salient patterns of the multi-channel time series data at different time scales. The architecture will be similar to the one proposed by Yang et al. [29], which showed very competitive performance. In this architecture, the convolution and pooling filters in the CNN are applied along the temporal dimension for each sensor, and all the feature maps for different sensors are unified as a common input for the neural network classifier.

3.3.1 CNN architecture

The architecture we implemented is shown in Figure 9.

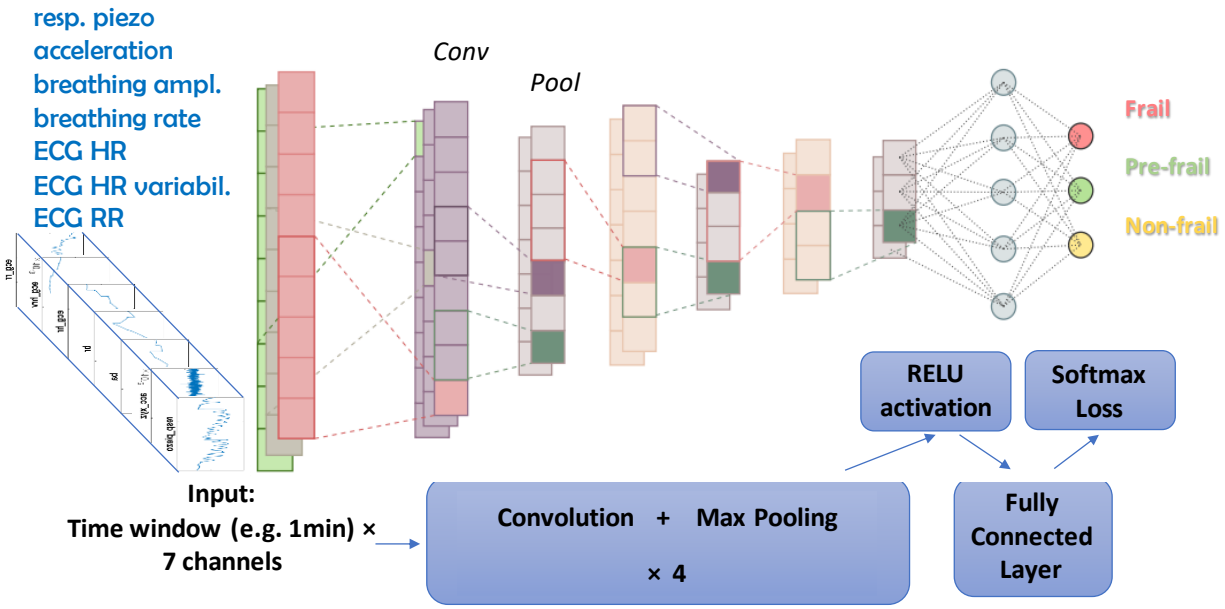


Figure 9: Recognition of frailty using recordings from strap/vest by a deep convolutional neural network (CNN).

The method has been implemented using the MatConvNet functions. The specific parameters of the several layers are included in the Figure 10.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
type	input	conv	mpool	relu	conv	mpool	relu	conv	mpool	relu	conv	mpool	relu	conv	softmxl
name	n/a	layer1	layer2	layer3	layer4	layer5	layer6	layer7	layer8	layer9	layer10	layer11	layer12	layer13	layer14
support	n/a	100x7	2	1	100x1	2x1	1	100x1	2x1	1	100x1	2x1	1	1	1
filt dim	n/a	1	n/a	n/a	20	n/a	n/a	20	n/a	n/a	50	n/a	n/a	500	n/a
filt dilat	n/a	1	n/a	n/a	1	n/a	n/a	1	n/a	n/a	1	n/a	n/a	1	n/a
num filts	n/a	20	n/a	n/a	20	n/a	n/a	50	n/a	n/a	500	n/a	n/a	3	n/a
stride	n/a	1	2	1	1	2	1	1	2	1	1	2	1	1	1
pad	n/a	2	0	0	0	0	0	0	0	0	0	0	0	0	0
rf size	n/a	100x7	101x8	101x8	299x8	301x8	301x8	697x8	701x8	701x8	1493x8	1501x8	1501x8	1501x8	1501x8
rf offset	n/a	48.5x2	49x2.5	49x2.5	148x2.5	149x2.5	149x2.5	347x2.5	349x2.5	349x2.5	745x2.5	749x2.5	749x2.5	749x2.5	749x2.5
rf stride	n/a	1	2	2	2	4	4	4	8	8	8	16	16	16	16
data size	1500x7	1405x5	702x2	702x2	603x2	301x1	301x1	202x1	101x1	101x1	2x1	1	1	1	1
data depth	1	20	20	20	20	20	20	50	50	50	500	500	500	3	1
data num	256	256	256	256	256	256	256	256	256	256	256	256	256	256	1
data mem	10MB	137MB	27MB	27MB	24MB	6MB	6MB	10MB	5MB	5MB	1000KB	500KB	500KB	3KB	4B
param mem	n/a	55KB	0B	0B	156KB	0B	0B	391KB	0B	0B	10MB	0B	0B	6KB	0B

Figure 10: CNN architecture and parameters of individual layers

3.3.2 Fusion of predictions for inter-subject analysis

A main difference in published works using CNNs for time series classification problems is that in the former the classification (e.g. for activity) is performed for each temporal window, whereas in our case a single decision score (prediction of frailty) has to be made for the whole recording (subject). As described in the previous section this is a Multiple Instance Learning problem. Thus, a fusion operation is required to assign a class label on each subject based on the decision scores of the time windows that constitute the recordings of the subject. We have examined several fusion approaches for prediction of the frailty status of each subject:

1. Maximum likelihood and majority voting across time windows:

   The class label is predicted for each time window based on maximum probability and then the most frequent class was assigned to the subject.

2. Maximum similarity of the probability profile:

   A feature vector is extracted to summarize the distribution of probability values of being frail for all time windows. Multi-variate regression is subsequently performed to learn a linear model that predicts the class labels of the training set based on this feature vector. The model is then applied on the test set for prediction.

3. The MIL algorithm MILES [30] is implemented and tested in an iterative scheme

   MILES algorithm extends ideas from the diverse density framework [31] and the wrapper model in feature selection [32]. The diverse density framework is based on the assumption that there exists a single target concept, which can be used to label individual instances correctly. The target concept that is most likely to agree with the data is then determined by maximizing this probability. MILES identifies instances that are relevant to the observed classification by embedding bags into an instance-based feature space and selecting the most important features. A subset of mapped features that is most relevant to the classification problem of interest are selected by defining a similarity measure between a bag and an instance. Although any feature selection approaches could be applied for this purpose, in MILES a joint approach is selected that constructs classifiers and chooses important features simultaneously. Classification is performed using the 1-norm SVM method because of its excellent performance in many applications. The 1-norm SVM is also referred to as sparse SVM can be formulated as a linear program, which, in general, can be solved efficiently from an optimization point of view.

   Our method iterates between CNN and MILES. The deep network extracts class probabilities for every time window which are used as features in MILES to select the most relevant instances (i.e. time windows). Once the time

windows are selected, these are introduced into the pretrained in the previous iteration CNN to get updated results (learn new parameters for the network). This iterative process resembles the Expectation-Maximization (EM) algorithm.

Results

Results have been produced by all fusion schemes. The obtained accuracy for all fusion schemes is similar or slightly less than the MIL framework with tensors described in section 3.2. Thus results are not analytically reported. This is possibly attributed to the small number of samples (subjects), large dimensionality, uncertainty in labelling, and high variability of data that are collected from daily living and not in a controlled environment or an experimental setting. All these issues make the problem very ill-posed. This is the reason why we tried to tackle this problem using standard feature extraction and machine learning techniques. We used two approaches, one based on clustering and one based on classification. Details on new analysis frameworks are presented in the next sessions along with corresponding results.

3.4 Aggregation of temporal parameters towards assessment of health transition

Most of the methods we have developed until now in WP4 focus on the extraction of relevant variables and the calculation of regression or classification models for prediction of the frailty status or for prediction of intermediate clinical metrics (defined in D2.1 as proxy outcomes). Although this was the main goal of the project, i.e. to extract new, more sensitive, frailty indices that cover all clinical domains, part of the clinical prediction engine includes the very ambitious goal of prediction of frailty transition which will facilitate early intervention. A first attempt to study the temporal transition and identify relevant variables has been performed in D4.2 where we calculated the distribution of transition values of the extracted frailty indices built from variables from the clinical evaluation, games, and wearable garments, while in D4.13 we analysed variables from the social domain. Our newest work for the final version of this deliverable (D4.17) includes the dynamic prediction based on all (identified as relevant) combined FrailSafe variables or individual prediction scores. The availability of measurements over multiple time points allows us to track the evolution of the clinical state of the elderly and therefore model frailty transition. This was performed following the strategy described next and illustrated in Figure 11.

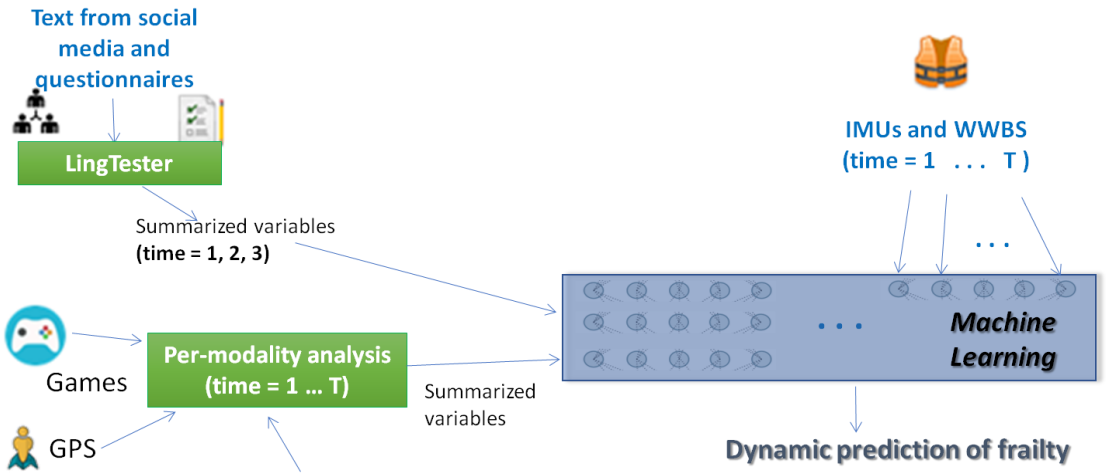


Figure 11: Prediction of frailty transition using multiparametric data by advanced machine learning techniques.

Variables from different domains were collected together into a data tensor X_{PxTxM}, where *P* is the number of different statistical properties, such as mean and standard deviation, *T* is the number of time points and *M* the number of modalities, which include different devices or different channels within the same device (e.g. from WWBS). The first step in this aggregation procedure is the synchronization of variables in time. Variables that are collected in sparse intervals, such as during clinical evaluation, were propagated to the reference time points either by interpolation (for quantitative variables) or by nearest neighbour classification (for ordinal variables). If we denote with the binary variable Y the incidence of an event (1 if it happened and 0 if it didn't happen) or the range of values for quantized variables (e.g. 1 for small values, 2 for middle values and 3 for high values), the aim is to predict Y given $X=\{X^i\}$, where *X* is the set of available samples. We have investigated several classification algorithms for this goal. As a note, the calculation of differences (change in feature values) is a special case of this approach, since the mathematical difference is just a linear mapping of the feature space.

All samples were concatenated and used within a machine learning algorithm to predict clinical scores over time (used as proxy measures of health deterioration) or possible adverse events, such as falls, not planned hospitalization and death (hard outcomes). Results of the prediction of the scores in the different clinical domains will be reported in the section, while the prediction of adverse events will be discussed in the next section.

3.4.1 Methods

The proposed method first builds a multi-dimensional profile of each participant by processing the multiple physiological signals to extract meaningful secondary measurements (e.g. heart rate from raw ECG). Statistical features are extracted from the raw or secondary measurements representing physiological and cognitive state, as well as indoor and outdoor mobility behavior. The multi-parametric features are subsequently fused into a long feature vector and introduced into a linear and non-

linear dimensionality reduction technique for extracting a small number of distinct patterns.

Clustering is then performed in this low dimensional embedding space in order to discover coherent and well separated groups. The results are evaluated by clustering validity criteria and the identified clusters are also compared with the groups determined by CGA in respect to several clinical metrics from multiple domains. A schematic diagram of the proposed methodology is illustrated in Figure 12, while more details on data and algorithms are provided next.

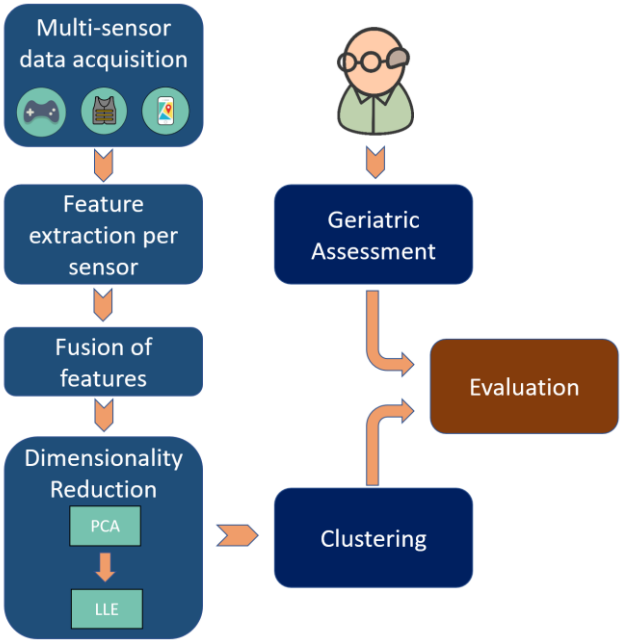


Figure 12: Proposed methodology for finding clusters of the aging population

3.4.1.1 Devices and feature extraction

The monitoring system consists of an aggregation of sensors, devices, and developed software that capture several aspects of the participants' health status (physiological, behavioral, cognitive etc.). In this work we extract features from a wearable device for physiological and kinetic monitoring (WWBS), a dynamometer for strength measurement, a game platform and a global positioning system (GPS) for outdoor monitoring. The devices and the extracted parameters are briefly described next. Details about the extracted variables can be found in deliverables D4.2 and D4.15. Next we provide a few more details on the time synchronization of variables as well as the extraction of variables by taking into account the estimated activity of daily living (ADL).

Temporal mapping of variables

FS variables: Even though all of the FrailSafe devices were given simultaneously to each participant, there were several cases in which there were no recordings from

one or more devices. This is due to the fact that the participant did not use all the devices on the same day, e.g. the participant played the game on a specific day but didn't perform any outdoor activity. In order to be able to use the total dataset available, we needed to perform an estimation on the values of the measurements of a specific sensor in time points where there was no recording. Hence, the empty cells were filled by performing linear interpolation in the range between known time points.

Clinical scores: Additionally, during the FrailSafe study the device and sensor recordings are collected regularly during the home visit sessions, whereas the participants' medical data are collected in larger intervals, i.e. during the clinical evaluations. In an effort to fuse the clinical data with the device recordings, it was necessary to estimate the clinical data values in intermediate time points for which there were device recordings available. Since we distinguish two types of clinical metrics - the ones quantified in a numeric scale, such as the TUG (Timed Up and Go) test, and the categorical ones, such as the frailty status according to Fried - a different approach was followed for each of them. The value of the clinical metric at the time of the session was estimated by linear interpolation if the variables were numeric, or by nearest neighbor interpolation if the variables were categorical.

Extraction of FS variables

Wearable WBAN System: The WWBS collects data from heart (a full ECG lead), respiration, posture and physical activity (IMUs with nine degrees of freedom). After time synchronization of all channels, the acquired or calculated signals were broken down into segments of main physical activities including moving and not moving conditions [36], such as sitting/standing, laying down, walking forward, walking upstairs or downstairs, and transition of activity. Such a partitioning of the signal (clustering of features) makes interpretation more consistent with human perception and allows extracting features within more uniform clusters. Previous work has also hinted at the utility of clustering to improve prediction accuracy. This has been attributed to its potential to exploit structure in the data and perform compression [37]. A set of statistical features was then calculated for each activity, resulting to an augmented feature vector containing features that correspond to all the activities. Nine statistical properties were calculated: average, standard deviation, 5% and 95% percentiles, mode, skewness, kurtosis, energy, entropy. The mode corresponds to the peak of the histogram, indicating the most frequently encountered value. Kurtosis characterizes the relative peakedness or flatness of the histogram, skewness is a measure of the distribution asymmetry and indicates the direction towards which the distribution is shifted, while energy and entropy are statistical measures of randomness and uncertainty. This feature extraction process resulted to 315 (9 statistical properties × 7 channels × 5 activities) variables for WWBS. The statistical properties were calculated within each *session*, defined as the time span of one day. That means that recordings acquired in consecutive days are treated as independent samples.

Game platform and dynamometer: The game recordings include a set of 13 variables tracked over time. The first eight features were summarized variables exported by the game application characterizing game performance. The last five dynamic measurements, concerning the height and the speed of the plane, the distance it covered, the remaining lives of the user and the force applied over the game, were aggregated by extracting their statistical properties similarly to the analysis of the WWBS recordings. Thus, the total number of features from the games was 53 (8 variables + 9 statistical properties × 5 dynamic variables).

GPS data: Finally, a number of features was extracted from the GPS data collected through the outdoor monitoring application (GPS logger).

The features from all devices were fused into a 385-dimensional feature vector (315 from WWBS, 53 from games and 17 from GPS). However, some of the extracted variables had many missing values, e.g. when the person did not perform a type of activity. Features with more than 20% of missing values were discarded from the subsequent analysis. Accordingly, we ended up with 174 features.

3.4.1.2 Dimensionality reduction

When dimensionality increases, the volume of the observations' space grows so much that the concept of similarity, distance or nearest neighbor may not even be qualitatively meaningful, thus impeding clustering or classification. Therefore, to facilitate clustering, we first perform dimensionality reduction based on a linear and a non-linear transformation of the data. Specifically, PCA [38] is first applied to extract a set of uncorrelated factors and then LLE [39] is used to find a lower dimensional embedding of the PCA scores.

In PCA the eigenvectors of the data covariance matrix corresponding to the largest eigenvalues are used to compute linear projections of greatest variance. Thus, PCA helps to eliminate redundant (zero-variance) dimensions. This is important in monitoring systems that operate also during resting phases. We used 25 components to represent the data which explained 87.72% of variance. A larger number of components was avoided because it would allow to account also for noise and random variations which would possibly lead to overfitting.

In contrast to PCA which performs only translation and rotation of the data, LLE recovers global non-linear structure from locally linear fits [39]. Let us consider a set of N data samples of dimensionality D in the ambient space R^D. If the data lie on or near a smooth non-linear manifold M of lower dimensionality $d \ll D$ then we can calculate a neighborhood-preserving mapping from the high-dimensional coordinates of each neighborhood to global internal coordinates on the manifold. The intrinsic dimensionality (d) is unknown but we used a small number, such as $d = 2$, to facilitate the subsequent cluster analysis.

3.4.1.3 Unsupervised clustering

Clustering performs partitioning of the data space into disjoint parts aiming to find hidden patterns in the data and gain insight. Since it is an unsupervised learning technique it allows an unbiased interpretation of the results, however an inherent

difficulty is how to determine the best number of clusters (K). This is usually accomplished by employing a second criterion that measures the robustness of the clustering, but we tested only the values $K = 2$ and 3, to be in accordance with the number of categories of the clinical metrics.

We investigated four popular clustering algorithms [40], namely the agglomerative (Agg), Birch (Bir), spectral clustering (Spec), k-means (KM), and also the combination of their results by majority voting (Comb). The algorithms were executed mostly with the default parameter values [40] which are described in more details next.

The agglomerative clustering algorithm recursively merges the pair of clusters that minimally increases a given linkage distance. The "euclidean" distance was used as metric to compute the linkage; the linkage criterion was the "ward", according to which the algorithm minimizes the variance of the clusters being merged.

The Birch (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm is an efficient and scalable data clustering method appropriate for large datasets and based on CF-trees, which serve as an in-memory summary of the data distribution. The radius of the subcluster obtained by merging a new sample and the closest subcluster was selected to be 0.0001 and the maximum number of subclusters in each node (branching factor) was set to 50.

Spectral clustering makes use of the eigenvalues of the similarity matrix of the data and is very useful when the structure of the individual clusters is highly non-convex. We select the "arpack" eigenvalue decomposition strategy, while the stopping criterion for eigendecomposition of the Laplacian matrix was set to 0.0, the number of neighbors to use when constructing the affinity matrix using the nearest neighbors method was 10 and the strategy to assign labels in the embedding space was "k-means".

The k-means algorithm assigns a data point to the cluster whose distance from the cluster centroid is minimum and iteratively updates the cluster centroids. The maximum number of iterations of the k-means algorithm for a single run was 300, the relative tolerance with regards to inertia to declare convergence was 0.0001 and the number of times the algorithm ran with different centroid seeds was 10.

3.4.1.4 Mapping of a new (unseen) data sample

The analysis of the currently available dataset allows to examine non-linear relations in the extracted features and identify data clusters. For new data it might be desired to classify them into the previously extracted clusters without rebuilding the models. For this purpose, the same feature extraction process should be applied followed by projection on the previously calculated principal components to reduce dimensionality and extract the scores in the PCA space (considered as ambient space R^D in the next step). Subsequently, in order to obtain the low-dimensional manifold position of the new sample from the ambient space (scores after PCA), the intrinsic coordinates based on the samples' neighborhood representation in high-dimensional space have to be inferred. We can assume an explicit mapping from the ambient space R^D to the manifold space M following the strategy described in [41], i.e.

perform the forward mapping by estimating the relationship between R^D and M as a joint distribution, such that there exists a smooth functional which belongs to a local neighborhood. This will result to a small feature vector (with $d = 2$ in our case) that can subsequently be assigned to the closest cluster.

3.4.2 Evaluation and Results

Out of all participants, 69 of them had used all the devices (one or more times) by the time the data were extracted for analysis and didn't have missing values in the CGA, thus the rest were excluded from the analysis. The number of sessions was not the same for each participant and is illustrated in the form of a histogram in Figure 13. The total number of sessions for the 69 participants was 924.

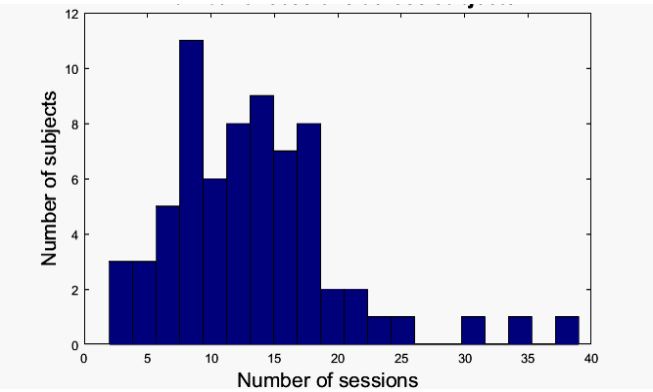


Figure 13: Histogram of the number of sessions for each participant

3.4.2.1 Clustering validity

Since clustering is an unsupervised classification problem, the lack of a gold standard makes it difficult to interpret the results and assess their accuracy. There are many different measures to check if good clustering has been achieved. The internal cluster validity indices quantify the quality of clustering using criteria such as cohesion and separation (similarity of an object to its own cluster and to other clusters, respectively). One of the most common criteria is the Silhouette index. Corresponding results are shown in Table 6 for each clustering result (with two or three clusters) and for each clustering technique. It can be observed that the quality of clustering is similar in all cases indicating robustness of the approach (coherent features).

Table 6. Silhouette index using square Euclidean distance

Num. Clusters	**ALGORITHMS**				
	Agg	***Bir***	***Spec***	***KM***	***Comb***

2	0.96	0.96	0.96	0.96	0.96
3	0.94	0.90	0.94	0.94	0.94

3.4.2.2 Clinical profile

Additionally, to clustering validation with internal criteria, we want to empirically investigate what is the predictive accuracy of the obtained clusters. For this purpose, we used clinical metrics acquired during CGA (shown in 0). Those metrics are selected under the prism of their operational function to quantify frailty, and categorized into domains taking under consideration the interrelationships that run through the implication of each variable in the various aspects of frailty. Before proceeding with classification assessment, we present the clinical characteristics of the participants, evaluated in each session, within each of the identified clusters. We distinguish two types of clinical metrics: the ones quantified in a numeric scale, such as the *TUG (Timed Up and Go) test*, and the categorical ones, such as the frailty status according to Fried [35] . Since CGA does not always coincide temporally with the use of the monitoring devices, the value of the clinical metric at the time of the session was estimated by linear interpolation if the variables were numeric, or by nearest neighbor interpolation if the variables were categorical.

The differences in the clinical profile of the participants in each cluster are presented in Table 7. Results have been produced using the combined clustering algorithm and for simplicity are shown only for two clusters ($K = 2$). The values are produced by averaging if the variables were numeric, or by frequency counting if the variables were categorical.

Table 7. Clinical profile in each cluster produced by the combined clustering algorithm with K=2.

DOMAIN	CLINICAL METRIC	Cluster 1[a]	Cluster 2 [a]
Medical	Orthostatic hypotension	No: 96	No: 100
		Yes: 4	Yes: 0
	Hearing	Poor: 27	Poor: 100
		Moderate/Good: 73	Moderate/Good: 0
	Vision	Poor: 3	Poor: 0
		Moderate: 19	Moderate: 0
		Good: 78	Good: 100
General Condition	Unintentional weight loss	No: 100	No: 0
		Yes: .0	Yes: 100
Physical Condition	Single foot standing	< 5 sec: 33	< 5 sec: 0
		> 5 sec: 67	> 5 sec: 100
	Time get up and go test	8.48 sec	12.57 sec
	Low physical	No: 90	No: 100

	activity	Yes: 10	Yes: 0
	Low grip strength	No: 61	No: 100
		Yes: 39	Yes: 0
	Gait speed (4m)	0.82 m/sec	0.86 m/sec
	MoCA score[42]	26.85	21.67
Cognitive	Subjective memory complain	No: 92	No: 0
		Yes: 8	Yes: 100
Psycholo-gical	Self-rated anxiety	3.34	2.39
	GDS-15 score [43]	1.24	5.51
	Leisure club participation	No: 16	No: 100
		Yes: 84	Yes: 0
	Leisure activities	7.55	7.0
Social	Telephone calls per week	15.1	4.0
	Visits / social interactions per week	4.45	1.0
	Self-rated quality of life	8.4	5.06
Wellness	Self-rated health status	Bad/Medium: 23 Good: 65	Bad/Medium: 100
			Good: 0
		Excellent: 12	Excellent: 0
	Self-rated pain	2.39	8.63
Lifestyle	Smoking	Never: 61 Past: 30	Never: 0 Past: 100
		Current: 9	Current: 0
	Physical activity	<2h per week: 23	<2h per week: 0
		>2h and <5h per week: 26	>2h and <5h per week: 0
		>5h per week: 51	>5h per week: 100
Nutrition	MNA screening score [44]	13.72	10.51
Frailty Status	Fried status	Non frail: 0.5	Non frail: 0.0
		Pre-Frail/Frail: 0.5	Pre-Frail/Frail: 1.0

a. For numeric variables numbers correspond to average values of the clinical metric, while for categorical variables numbers correspond to proportion (%) of sessions within each group

3.4.2.3 Predictive accuracy

We can assume that the clustering algorithm is good for prediction if it agrees well with a set of hidden labels using a small number of clusters [37]. We selected the previous clinical metrics as ground truth to assess our results. In order to incorporate the continuous variables as well, ranges of values have been selected to quantize the dataset according to clinicians' guidelines into 2 or 3 levels, such that all clinical variables of CGA were categorical. Since clustering returns unlabeled groups of observations, comparison with predefined groups (as the ones obtained by quantizing the clinical scores) is not well defined. We searched for the mapping that produced the highest overlap between the obtained clusters and the groups defined by each clinical metric and calculated the clustering accuracy (number of correctly classified samples over total number of samples) according to this mapping for the combined clustering algorithm. Accuracy was assessed using $K = 2$ for variables that were quantized into 2 levels and $K = 3$ for variables that were quantized into 3 levels. Additionally to the overall accuracy, we calculated the balanced accuracy, which is the average of sensitivity and specificity. The balanced accuracy is more informative when the classes are imbalanced since the errors in the small classes are not outweighted by correct assignments in large classes.

Table 8 illustrates the cluster overlap for each of the clinical metrics. For most of the clinical metrics the results reveal a high overlap of the identified clusters with the groups determined by the clinical metrics indicating the potential of the proposed approach to predict the outcome of clinical tests. Highest accuracy is observed for the general condition as expressed by unintentional weight loss (100%), nutrition (96%), cognitive domain (85% on the average), psychological domain (83.5% on the average), and wellness (83% on the average), whereas the lowest accuracy was obtained for the frailty status, and the social domain. While the cluster assessment facilitated our understanding of the predictive capabilities in respect to the different domains, the interpretation of the obtained results in Table 7 remains challenging due to the high cluster size imbalance ($n = 893$ sessions in cluster 1 and $n = 31$ sessions in cluster 2). Cluster 2 seems to characterize an isolated profile of subjects with mixed physical condition, poor hearing, unintentional weight loss, memory complains, mild depression and limited social life (visits, telephone). The discrimination ability of the method seems significantly lower when it is assessed by the balanced accuracy, which indicates that the smaller cluster is underrepresented. Clinical scores that are predicted with high accuracy (>70%) by both evaluation criteria (overall and balanced accuracy) include the unintentional weight loss, GDS-15 score, self-rated quality of life, and the MNA screening score.

Table 8. Clustering accuracy per clinical metric

DOMAIN	CLINICAL METRIC	Accuracy (%)	Balanced Accuracy (%)
Medical	Orthostatic hypotension	92	48

	Hearing Impairment	74	56
	Vision Impairment	68	30
General Condition	Unintentional weight loss	100	98
Physical Condition	Single foot standing	65	48
	Time get up and go test	86	60
	Low physical activity	87	48
	Low grip strength	59	47
	Gait speed (4m)	42	37
Cognitive	MoCA score	78	57
	Subjective memory complain	92	65
Psycholo-gical	Self-rated anxiety	71	48
	GDS-15 score	96	74
Social	Leisure club participation	85	59
	Leisure activities	68	36
	Telephone calls per week	52	53
	Visits / social interactions per week	57	39
Wellness	Self-rated quality of life	98	84
	Self-rated health status	62	36
	Self-rated pain	89	62
Lifestyle	Smoking	66	51
	Physical activity	42	27
Nutrition	MNA screening score	96	73
Frailty Status	Fried status	51	53

3.5 Prediction of clinical variables' outcomes from non-clinical measurements

To inspect the sufficiency of FrailSafe's non-clinical monitoring components with respect to the clinical variables that are claimed to be related to frailty, we designed a model that treats the aforementioned measurements as predictive variables and feeds them to a classifier to predict the clinical variables. A preliminary version of

this work has been reported in D4.15, but a thorough examination is described in the following subsections.

3.5.1 Description of model

The measurements and corresponding features that were used for this approach came from four different FrailSafe components: the WWBS, the GPS, the games and the text collections. The preprocessing and feature extraction pipeline has already been described in the previous section. The statistical features extracted from the corresponding components are treated as a unified feature space, meaning that the model is blind about which source has each feature been extracted from. That said, the procedure of building a predicting model includes some certain steps to ensure that the results will not be questionable in any way. More specifically, since each target variable may contain different number of samples and/or subjects, a series of rules were applied in order to exclude variables that are impossible to be explained and predicted through the features.

3.5.1.1 Exclusion of very small classes

To avoid the occurrence of a small dataset, variables that contain more than 20% empty rows are immediately excluded. As a next step, we observe the number of subjects that correspond to samples of each class. The rule is that if there are at least 10 unique subjects that belong to each class, then we can proceed to the next step. Otherwise, we check the class-cardinality of the variable, and if we are dealing with a binary variable, then this target is excluded from the analysis. But in the case of a multi-class variable, the action that we take is to merge the class that contains less than 10 unique subjects with one of its neighboring classes, choosing the class with the fewer samples. Followingly, a standard scaling of the features is performed in the range [0, 1], and then we apply PCA decomposition to extract the most meaningful components of the features. The components that were kept represented the data which explained 98% of variance. The aforementioned pipeline is depicted in Figure 14.

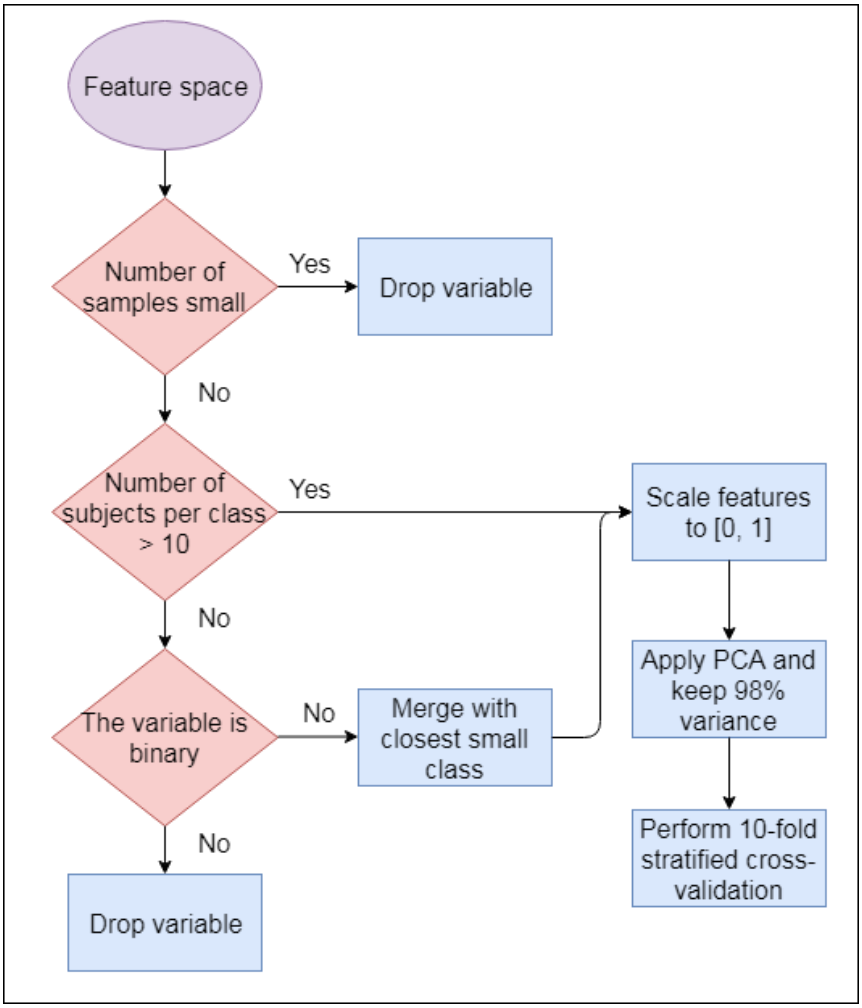


Figure 14: Samples preproscessing pipeline

3.5.1.2 Design of cross-validation scheme

After preprocessing we constructed a classification model using 10-fold stratified cross validation scheme. More specifically, cross-validation indices are created for each fold, so that the imbalance that occurs in the original dataset is still existent in each split at the same ratio. The imbalance here refers to the number of samples existing in each class, and not the subjects. To avoid inserting bias to our model, at each fold we transfer occurrences of the same subjects' samples from the training set to the test set, so that no samples of a specific subject are split in the two sets.

Since the target variables appear to be highly imbalanced, we applied an over-sampling technique (SMOTE - Synthetic Minority Oversampling Technique) in order to populate the minority class with new samples. The latter is only applied to the training set of each fold, to enrich the model's knowledge on the training samples of the minority class. The model that was chosen to be applied for the prediction of the output was k-nearest neighbors classifier with 5 neighbors, which was also the selected classifier for the previously reported analysis in D4.16. The model is fit to the training set of each fold and then is applied to the corresponding test set to

obtain the classification accuracy. At each fold three different accuracies are calculated: accuracy with respect to the samples, accuracy with respect to the subjects, and balanced accuracy that takes into consideration the imbalance of the dataset. The resulting mean value of the accuracies across the 10 folds are the evaluation metrics that are reported as a measure of our model's performance. A detailed report of the aforementioned procedure's results is presented in the next section.

3.5.2 Prediction results and model discussion

The analysis that was described above was conducted on a dataset of 1347 samples of 87 unique participants. After a first exploratory analysis of the dataset, 6 clinical scores were immediately excluded due to a critically small number of samples, while the samples that contained empty values were also dropped, resulting in a new non-missing-values dataset with 1115 samples of 80 unique participants. Regarding the phenomenon of few subjects per class that occurred in each clinical variable, the list of variables that were excluded as well as the variables whose classes were internally merged are shown in Table 9:

Table 9: Exclusion & merging of very small classes

Excluded variables	***Variables with merged classes***
Low physical activity	Vision Impairment
Unintentional weight loss	Self-rated health status
MNA screening score	Self-rated health status comparison
GDS-15 score	Smoking
Self-rated quality of life	Physical activity
Self-rated pain	Leisure activities

The mean test accuracies that resulted from the 10-fold cross-validation are depicted in the following table and the corresponding figure:

Table 10: Classification results of aggregated-devices approach

Clinical variable	***Subject accuracy***	***Instances accuracy***	***Balanced accuracy***
Fried status	40%	42%	45%
Leisure club participation	61%	60%	52%
Subjective memory complaint	71%	63%	38%
Low grip strength	33%	43%	46%

Orthostatic hypotension	68%	65%	44%
Hearing impairment	52%	53%	51%
Vision Impairment	42%	50%	48%
Self-rated health status	33%	38%	34%
Self-rated health status comparison	54%	48%	24%
Smoking	48%	53%	54%
Physical activity	51%	43%	47%
Timed get up and go test	49%	49%	41%
Gait speed (4m)	32%	33%	33%
MoCA score	62%	58%	50%
Self-rated anxiety	55%	52%	49%
Visits / social interactions per week	52%	48%	40%
Telephone calls per week	44%	40%	36%
Leisure activities	72%	69%	60%
Single foot standing	68%	65%	**68%**
Adverse event	57%	58%	60%

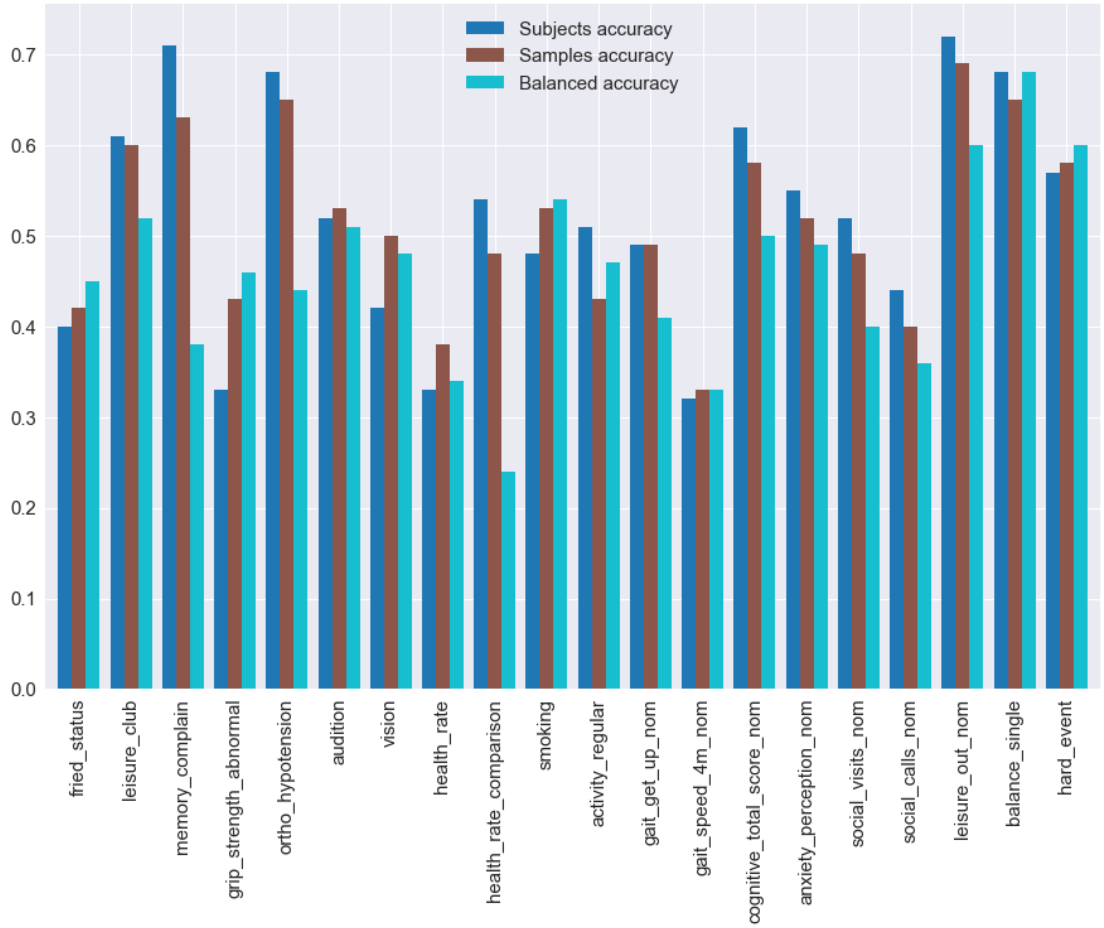


Figure 15: Plot of aggregated-devices approach accuracies

What is immediately made clear from the results is that although the classification model appears to perform very well on specific variables regarding the per-subjects and per-samples accuracy, the corresponding balanced accuracy seems to be quite lower than expected, indicating the importance of the imbalance problem. Although a series of actions were taken in order to eliminate this issue, it still remains an obstacle in creating a robust prediction model, insensitive to new entries. Since balanced accuracy is the metric that more accurately reflects the model performance, in comparison to the other two accuracies, in the next approaches we evaluate our models using this measure.

A previous analysis that had been conducted on an earlier version of this dataset was reported in D4.15, section 4.3.1. Although a strict comparison between the former and current work is not feasible due to changes in the dataset and the approach followed, one can see the previous corresponding results in Figures 25, 26 and 27 of the aforementioned deliverable. It should be mentioned thought that the previous analysis focused on building separate models for each FrailSafe component, whereas the current approach aims at exploiting a unified feature space to examine the possibility of using it as a predictive scheme for clinical variables. Regarding the model differences between the two approaches, a k-nearest neighbors classifier was selected in our former approach, while at the current study we used a linear SVC

model. This decision to change the classifier was made after experimenting with k-nearest neighbors and realizing that it did not perform as well as in the first case.

3.5.3 Breaking down the FrailSafe's components' predictive capability

To further explore the potential of a FrailSafe component being able to predict a clinical score, we design a set of separate classification models – one for each component – that could serve as input to a decision fusion model. The purpose of this experiment is to examine the contribution of each component separately in the prediction process. This approach corresponds to the late integration step that was reported in D 4.15, section 4.3.2. Nevertheless, the current approach is evaluated on different set of subjects and samples, as well as more robust evaluation metric. The preprocessing steps that were reported in the previous subsection are the same that we follow for this approach as well, except for the PCA decomposition which is not applied here, as well as the oversampling technique. Since a first attempt to apply the aforementioned classification model – a k-nearest neighbors classifier – to the current approach resulted in very low classification accuracies, we decided to utilize a naïve Bayes model instead. Thus, a Gaussian naïve Bayes classifier is built for each of the four FrailSafe components, while for the fusion step a random forest was selected as classification model. The fusion is performed across the predictions of the separate classifiers, which serve as features for classification. To evaluate the performance of our models in an accurate way that reflects the extent of their functionality, we again measure the balanced accuracy of the classification process. The resulting accuracies are depicted in Table 11 and Figure 16 respectively.

Table 11: Balanced accuracy of devices-breakdown approach

Clinical variable	*GPS*	*WWSX*	*Games*	*Text*	*Fusion*
Fried status	50%	51%	58%	53%	55%
Leisure club participation	52%	63%	63%	51%	**73%**
Subjective memory complaint	56%	44%	40%	46%	42%
Low grip strength	45%	47%	54%	58%	48%
Orthostatic hypotension	50%	50%	54%	45%	51%
Hearing impairment	55%	45%	47%	51%	46%
Vision Impairment	59%	54%	59%	56%	60%
Self-rated health status	34%	32%	31%	33%	27%
Self-rated health status comparison	23%	28%	34%	18%	29%
Smoking	51%	50%	54%	44%	53%
Physical activity	37%	37%	44%	44%	53%
Timed get up and go test	48%	50%	52%	52%	52%
Gait speed (4m)	35%	31%	37%	46%	41%

MoCA score	47%	41%	56%	**70%**	53%
Self-rated anxiety	51%	41%	47%	52%	44%
Visits / social interactions per week	35%	26%	44%	29%	34%
Telephone calls per week	29%	29%	30%	29%	31%
Leisure activities	52%	53%	55%	42%	55%
Single foot standing	56%	52%	**65%**	**73%**	**72%**
Adverse event	55%	61%	58%	62%	61%

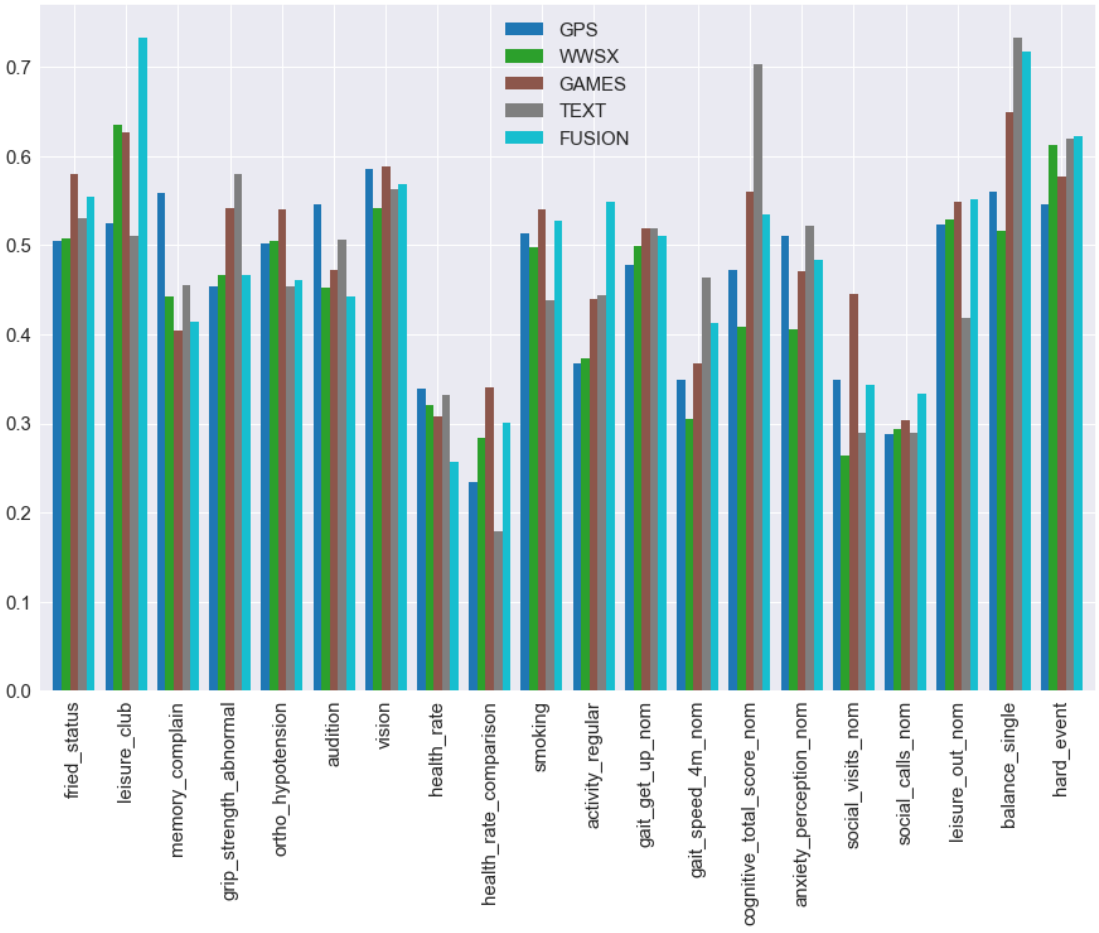


Figure 16: Plot of devices-breakdown approach balanced accuracies

The breakdown of the FrailSafe components reveals the predictive strength of specific devices with regards to particular clinical scores, e.g. the *MoCA score* is predicted with balanced accuracy 70% by the features of the text classifier. On the other side, the fusion is not likely to inherit the highest classification accuracy of a specific classifier that works correctly for a specific variable. Nevertheless, there are two paradigms that break this rule, with the fusion model appearing to achieve the highest balanced accuracy for the *leisure club participation score*, as well as in the

case of *single foot standing*. The results of this analysis point out to the fact that not enough timepoints are provided in order for the devices to be able to serve as a prediction machine for the clinical scores. We examine the validity of this assumption in the next section.

3.5.4 Mapping multiple FrailSafe samples into a delta-space

The original dataset could be manipulated as a set of deltas between different timepoints that recordings have been acquired by each device. Such a dataset inherently contains as many samples as the number of subjects, so that a sample represents the overall change that has been observed during the collection of several recordings. To be more precise, the size of this reduced data collection is equal to 67 distinct rows (subjects/samples) and 143 features from all devices (the number of features is the same as before). Since the separate-model-per-device approach seemed to perform better, we follow the same scheme here too. Thus, a separate Gaussian naïve Bayes classifier is built for each FrailSafe component and a fusion of decisions is performed at the end using a random forest classifier. Regarding the rules that aim to either exclude or merge classes of variables with very few subjects, the corresponding manipulations are depicted in Table 12 and Table 13. We evaluate the performance of the prediction scheme using balanced accuracy as metric. Results are presented in Table 14 and Figure 17 accordingly.

Table 12: Exclusion of very small classes

Excluded variables
Leisure club participation
Unintentional weight loss
MNA screening score
Self-rated quality of life

Table 13: Merging of very small classes

Variables with merged classes	
Subjective memory complaint	MoCA score
Low physical activity	GDS-15 score
Low grip strength	Self-rated anxiety
Orthostatic hypotension	Self-rated pain
Hearing impairment	Visits / social interactions per week
Vision Impairment	Telephone calls per week
Self-rated health status	Time spent on phone per week

Self-rated health status comparison	Time spent on videoconference per week
Smoking	Written messages sent per week
Physical activity	Visits / social interactions per week
Timed get up and go test	Leisure activities
Gait speed (4m)	

Table 14: Balanced accuracy of delta-space approach

Clinical variable	*GPS*	*WWSX*	*Games*	*Text*	*Fusion*
Fried status	44%	41%	45%	50%	46%
Subjective memory complaint	53%	57%	42%	43%	51%
Low physical activity	45%	41%	61%	63%	49%
Low grip strength	**67%**	50%	**70%**	57%	64%
Orthostatic hypotension	58%	36%	56%	45%	45%
Hearing impairment	60%	61%	47%	63%	48%
Vision Impairment	55%	**75%**	40%	35%	49%
Self-rated health status	42%	54%	50%	64%	55%
Self-rated health status comparison	37%	57%	43%	44%	48%
Smoking	50%	50%	50%	17%	50%
Physical activity	47%	63%	42%	40%	47%
Timed get up and go test	52%	64%	61%	47%	48%
Gait speed (4m)	51%	44%	46%	48%	38%
MoCA score	50%	38%	50%	45%	50%
GDS-15 score	50%	50%	50%	62%	50%
Self-rated anxiety	**65%**	47%	47%	52%	**65%**
Self-rated pain	63%	38%	41%	44%	50%
Visits / social interactions per week	56%	**75%**	47%	47%	51%
Telephone calls per week	53%	**69%**	54%	51%	58%
Time spent on phone per week	63%	57%	59%	51%	54%
Time spent on videoconference per week	49%	39%	44%	42%	50%
Written messages sent per week	48%	51%	64%	39%	48%
Leisure activities	50%	64%	**66%**	60%	49%

Single foot standing	44%	57%	55%	63%	49%
Adverse event	60%	37%	**66%**	50%	54%

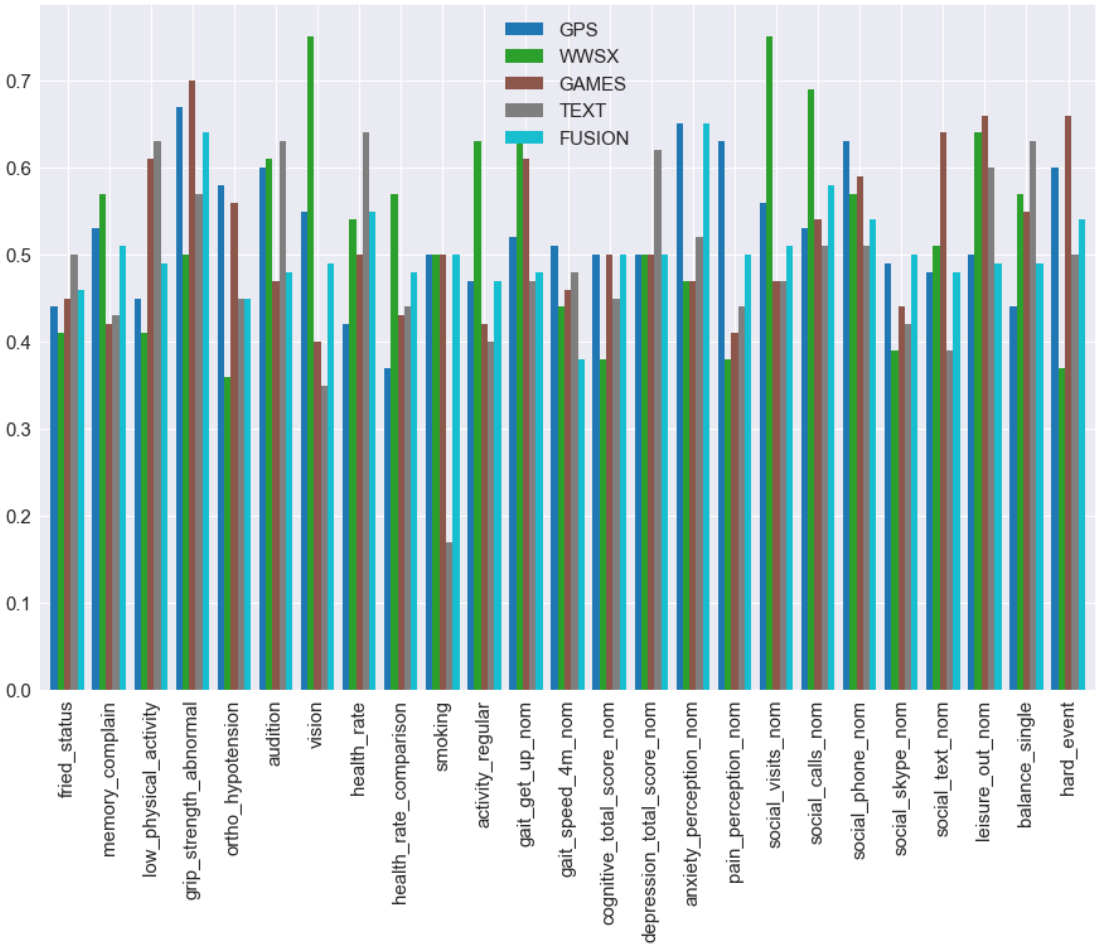


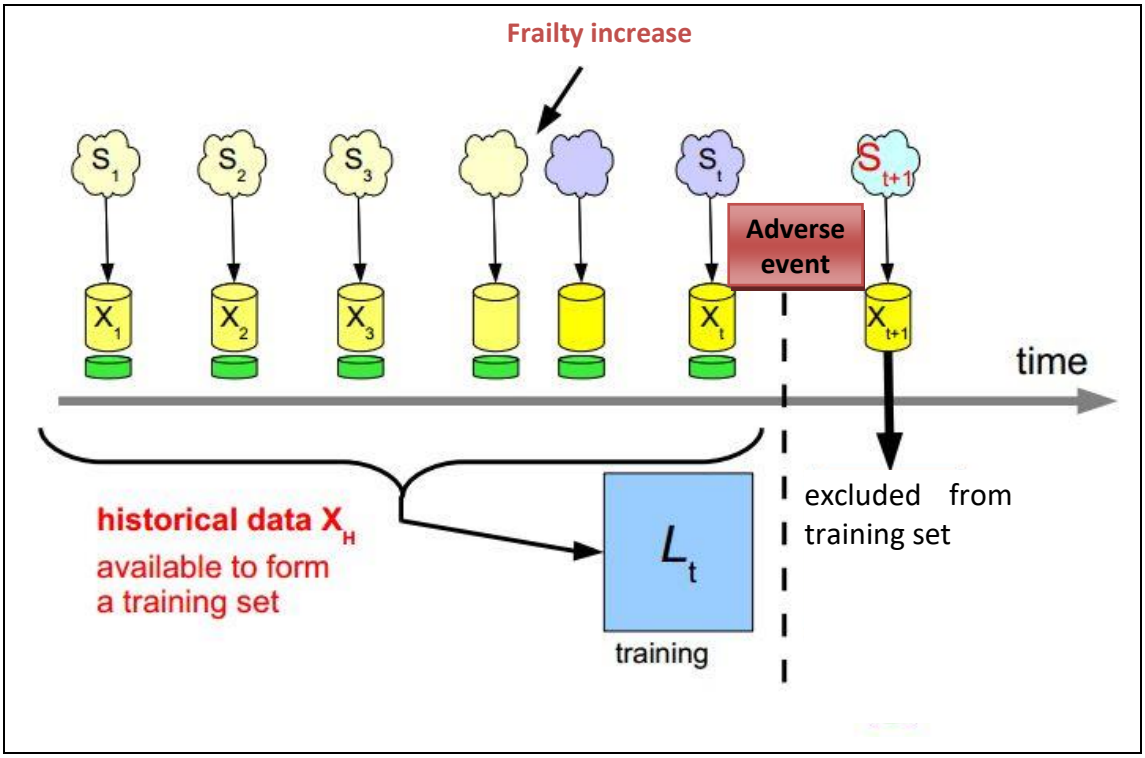
Figure 17: Plot of delta-space approach balanced accuracies

The current model appears to behave in a quite similar way to the aforementioned devices-breakdown model, with specific devices being able to predict a subset of the clinical scores. For example, the variable *Low grip strength* is correctly classified by the games' features with accuracy 70%, which makes sense since there is a couple of games that are played using a dynamometer, a device that measures the grip strength. Nevertheless, the fusion of the classifiers' decisions is still pure in terms of the resulting accuracy.

3.6 Aggregation of temporal parameters towards prediction of adverse events

In addition to the previous analysis frameworks where we examined the relationship between input and target variables for each session, here we focus on adverse

events that happen at a specific time point t. We define the time of the adverse event as the end time point and use all previous instances (historical data) to form the training set $X_H = \{X_1, X_2, \dots, X_t\}$. The subsequent instances $\{X_{t+1}, \dots\}$ are discarded because the measurements might have been affected by the incurrence of the adverse event. For example the total daily outdoor walking distance covered by an individual after a fall or hospitalization is expected to be significantly smaller and not representative of his/her (at that time point) frailty status. If no adverse event has happened, all instances are retained. The aim of this analysis is given X_H to predict whether an adverse event will happen (Y=1) or not (Y=0) within a predefined time period, such as a year. The idea is to examine whether we can extract early indicators of deterioration in the participants' health condition that might lead to dangerous events.



First we formulated this task as a multiple instance learning (MIL) problem in which the temporal alignment of the multiple instances (sessions) was ignored. The rational behind this is that (i) the calculated variables are not always consistent due to small deficiencies in hardware or software components and (ii) some of the variables (e.g. based on questionnaires during the clinical evaluation) are subjective and might change in an incoherent way. For example, cases with an apparent decrease in frailty level based on Fried score are not few, but the rest of the findings do not support this data. By solving this supervised classification problem through MIL we allow a predefined fraction of the data to be indicative of the health's deterioration without explicitly fitting a temporal model of transition.

3.6.1 Methods and Results

Several MIL algorithms were investigated using the Matlab Toolbox for Multiple Instance Learning (mil tools 1.2.0) [33]. The toolbox is an extension of the PRTools toolbox, which was used for the book [34] and in which Matlab objects for prmapping and prdataset are defined. Among the investigated algorithms were the MILBOOSTC, SIMPLE_MIL, miSVM, SPEC_MIL, MILES and sparse logistic regression (SPARSELOGLC).

The MILBOOSTC algorithm is described in [45]. The SIMPLE_MIL algorithm uses a standard untrained mapping on a MIL-dataset A. The classifier is trained on all the data in A, without considering the fact that they are organized in bags. In the evaluation phase, the instances from a bag are classified, and the outputs of all instances is combined using a preselected (as input) rule. The method miSVM applies Support Vector machines for multiple-instance learning [46]. The SPEC_MIL algorithm is a specializing multi-instance learner, that follows a generalization of the miSVM [46]. It first trains a classifier on all the data and then relabels all data according to the output of this classifier. Then it checks that at least one instance in the positive bags is labeled positive, and if not, it changes the least negative instance to a have a positive label. This is iterated a predefined number times, or until the labels do not change anymore. MILES [47] performs multi-instance learning in embedded subspaces [47]. Specifically the model is learnt by representing the bags based on the maximum kernel similarity (quantified using proximity mapping) to all other instances. On this dissimilarity representation a sparse linear classifier is trained. For SPARSELOGLC a sparse logistic classifier (implemented by the SLEP package) optimizes the logistic loss on the training set, regularized by the L_1 norm of the weights [48].

A grid search on the most crucial parameters was performed for each classifier. The best results were obtained for SPEC_MIL (with parameter frac=0.325) and are shown in **Table 15**. We also tested the method on the same variables except the ones extracted by the analysis of text. The reason for this experiment is that most of the participants do not use social media, so the text variables were mostly extracted from a test performed during the CGA during which the participants were instructed to describe a given illustrated image as well as an important event in their life. This procedure (to collect and convert to electronic form the text) is not very trivial, thus the additional contribution of these features requires investigation. The results are shown in **Table 15**, 2nd row. It can be observed that the text variables contribute the overall accuracy (balanced or unbalanced) but the performance of the classifier in respect to AUC is better without the text variables. This might be explained by the fact that less variables generally increase robustness also the number of subjects is slightly larger ($n = 80$ versus $n = 76$ with text) since some subjects that did not have all variables were excluded in the joint analysis.

In order to better assess the FrailSafe system's potential we compared with standard evaluation practices based on the CGA and specifically the Fried status. Specifically, we applied the same classification algorithm to predict the occurrence of adverse events using only the Fried status as predictor variable or using all clinical variables

(including Fried) acquired during the common geriatric assessment. The results are illustrated in **Table 15** (3rd and 4th row, respectively). It can be observed that the FS variables have a higher potential in predicting adverse outcomes than the Fried status alone or combined with the rest of the clinical metrics.

Table 15. Prediction of hard outcomes by SPEC_MIL

Features	***AUC***	***Accuracy***	***Balanced Accuracy***
All variables (clinical + FS)	0.70	0.77	0.73
All variables (clinical + FS) except text	0.74	0.75	0.70
Fried	0.60	0.72	0.49
Fried + clinical	0.56	0.69	0.53

4 Conclusions

In this deliverable, the Decision Support System has been described along with the role of inflammation in the pathogenesis of frailty. The DSS includes the clinical state prediction engine that aimed at investigating and specifying appropriate physiological and behavioral characteristics that can be used for defining biomarkers of frailty that can be of a significant predictive value. Cluster analysis helped finding groups of data that were in good accordance with the outcome of clinical tests, indicating that there is high potential in the proposed monitoring system and data analysis framework. Also, classification algorithms have been exploited in order to build prediction models for the different clinical metrics as well as to predict adverse events. The results encourage further investigation of the prognostic capacity in terms of predicting frailty transition and subsequent risk factors. Finally, a risk assessment module was developed which is responsible for the generation of alerts.

References

1. Clegg, A., et al., Frailty in elderly people. Lancet, 2013. 381(9868): p. 752-62.
2. Xue, Q.L., The frailty syndrome: definition and natural history. Clin Geriatr Med, 2011. 27(1): p. 1-15.
3. Kotas, M.E. and R. Medzhitov, Homeostasis, inflammation, and disease susceptibility. Cell, 2015. 160(5): p. 816-27.
4. Kim, D. and C.L. Haynes, Neutrophil chemotaxis within a competing gradient of chemoattractants. Anal Chem, 2012. 84(14): p. 6070-8.
5. Stuart, L.M., P.M. Henson, and R.W. Vandivier, Collectins: opsonins for apoptotic cells and regulators of inflammation. Curr Dir Autoimmun, 2006. 9: p. 143-61.
6. Du Clos, T.W., Function of C-reactive protein. Ann Med, 2000. 32(4): p. 274-8.
7. Merle, N.S., et al., Complement System Part II: Role in Immunity. Front Immunol, 2015. 6: p. 257.
8. Carter, R.H., B cells in health and disease. Mayo Clin Proc, 2006. 81(3): p. 377-84.
9. Wolkow, A., et al., Relationships between inflammatory cytokine and cortisol responses in firefighters exposed to simulated wildfire suppression work and sleep restriction. Physiol Rep, 2015. 3(11).
10. Slavich, G.M. and M.R. Irwin, From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. Psychol Bull, 2014. 140(3): p. 774-815.
11. Chovatiya, R. and R. Medzhitov, Stress, inflammation, and defense of homeostasis. Mol Cell, 2014. 54(2): p. 281-8.
12. Kessler, R.C. and E.J. Bromet, The epidemiology of depression across cultures. Annu Rev Public Health, 2013. 34: p. 119-38.
13. Hidaka, B.H., Depression as a disease of modernity: explanations for increasing prevalence. J Affect Disord, 2012. 140(3): p. 205-14.
14. Albert, P.R., Why is depression more prevalent in women? J Psychiatry Neurosci, 2015. 40(4): p. 219-21.
15. Wilcox, G., Insulin and insulin resistance. Clin Biochem Rev, 2005. 26(2): p. 19-39.
16. van Vollenhoven, R.F., New therapeutic approaches in rheumatoid arthritis. Presse Med, 2016. 45(6 Pt 2): p. e179-92.
17. Marra, A.M., et al., Cardiovascular abnormalities and impaired exercise performance in adolescents with congenital adrenal hyperplasia. J Clin Endocrinol Metab, 2015. 100(2): p. 644-52.
18. Reule, S. and P.E. Drawz, Heart rate and blood pressure: any possible implications for management of hypertension? Curr Hypertens Rep, 2012. 14(6): p. 478-84.

19. Blum, A. and N. Blum, Coronary artery disease: Are men and women created equal? Gend Med, 2009. 6(3): p. 410-8.
20. Chen, X., G. Mao, and S.X. Leng, Frailty syndrome: an overview. Clin Interv Aging, 2014. 9: p. 433-41.
21. Gentile, G., et al., An overlooked pink species of land iguana in the Galapagos. Proc Natl Acad Sci U S A, 2009. 106(2): p. 507-11.
22. Rodríguez-Molinero A. et al., Normal respiratory rate and peripheral blood oxygen saturation in the elderly population. J Am Geriatr Soc. 2013 Dec;61(12):2238-40.
23. Dietterich, T.G., R.H. Lathrop, and T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 1997. 89(1): p. 31-71.
24. Amores, J., Multiple instance classification: Review, taxonomy and comparative study. Artificial Intelligence, 2013. 201(Supplement C): p. 81-105.
25. Foulds, J. and E. Frank, A review of multi-instance learning assumptions. The Knowledge Engineering Review, 2010. 25(1): p. 1-25.
26. Xu, X. and E. Frank, Logistic Regression and Boosting for Labeled Bags of Instances, in Advances in Knowledge Discovery and Data Mining: 8th Pacific-Asia Conference, PAKDD 2004, Sydney, Australia, May 26-28, 2004. Proceedings, H. Dai, R. Srikant, and C. Zhang, Editors. 2004, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 272-281.
27. Gamboa, John Cristian Borges. "Deep Learning for Time-Series Analysis." arXiv preprint arXiv:1701.01887, 2017.
28. Zacharaki, Evangelia I. "Prediction of protein function using a deep convolutional neural network ensemble." PeerJ, 2017: e2778v1.
29. Yang, Jianbo, et al. "Deep convolutional neural networks on multichannel time series for human activity recognition." Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.
30. Chen Y, Bi J, Wang JZ. MILES: Multiple-instance learning via embedded instance selection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006 Dec;28(12):1931-47.
31. Chen Y, Wang JZ. Image categorization by learning and reasoning with regions. Journal of Machine Learning Research. 2004;5(Aug):913-39.
32. Kohavi R, John GH. Wrappers for feature subset selection. Artificial intelligence. 1997 Dec 1;97(1-2):273-324.
33. Tax, D.M.J., Cheplygina, V., MIL, A Matlab Toolbox for Multiple Instance Learning, version 1.2.1, Jun 2016, http://prlab.tudelft.nl/david-tax/mil.html.

34. Ferdi van der Heijden, Robert P.W. Duin, Dick de Ridder and David M.J. Tax, Classification, parameter estimation and state estimation – an engineering approach using Matlab", 2004.

35. Linda P. Fried et al., "Frailty in older adults: evidence for a phenotype," The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, vol. 56, pp. M146--M157, 2001.

36. Evangelia Pippa, Iosif Mporas, and Vasileios Megalooikonomou, "Feature Selection Evaluation for Light Human Motion Identification in Frailty Monitoring System.," in ICT4AgeingWell, 2016, pp. 88-95.

37. Arindam Banerjee and John Langford, "An objective evaluation criterion for clustering," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 515-520.

38. Ian T. Jolliffe, "Principal component analysis and factor analysis," in Principal component analysis.: Springer, 1986, pp. 115-128.

39. Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," science, vol. 290, pp. 2323-2326, 2000.

40. Fabian Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of machine learning research, vol. 12, pp. 2825-2830, 2011.

41. Samuel Kadoury, Guray Erus, Evangelia I. Zacharaki, Nikos Paragios, and Christos Davatzikos, "Manifold-constrained embeddings for the detection of white matter lesions in brain MRI," in Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, 2012, pp. 562-565.

42. Ziad S. Nasreddine et al., "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," Journal of the American Geriatrics Society, vol. 53, pp. 695-699, 2005.

43. Javaid I. Sheikh and Jerome A. Yesavage, "Geriatric Depression Scale (GDS): recent evidence and development of a shorter version.," Clinical Gerontologist: The Journal of Aging and Mental Health, 1986.

44. Yves Guigoz, Bruno Vellas, and Philip J. Garry, "Assessing the nutritional status of the elderly: The Mini Nutritional Assessment as part of the geriatric evaluation," Nutrition reviews, vol. 54, p. S59, 1996.

45. Babenko, B., Dollar, P., Tu, Z., Belongie, S., Simultaneous learning and alignment: Multi-instance and multi-pose boosting. Technical Report CS2008, UCSD (2008)

46. Andrews S., Tsochantaridis I., Hofmann T., Support Vector Machines for Multiple-Instance Learning, NIPS 2003.

47. Chen Y, Bi J, Wang JZ. MILES: Multiple-instance learning via embedded instance selection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006 Dec;28(12):1931-47.

48. Liu J, Chen J, Ye J. Large-scale sparse logistic regression. InProceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 2009 Jun 28 (pp. 547-556). ACM.