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: D4.15

Signal processing algorithms for extraction of frailty related indicators (vers b)

Due date of deliverable:	M24 (31st December 2017)
Actual submission date:	31st January 2018
Version:	2.0
Lead Author:	Konstantinos Deltouzos (UoP)
Lead partners:	Konstantinos Deltouzos, Spyros Kalogiannis, Emilia Papagiannaki, Thomas Papastergiou, Athanassios Kordas, Evangelia Pippa, Evangelia Zacharaki, Dimitrios Vlachakis, Yiannis Ellul, Vasilis Megalooikonomou (UoP), Andreas Vasilakis, Ilias Kalamaras, Konstantinos Moustakas (CERTH)

Horizon 2020
European Union funding
for Research & Innovation

Change History

Ver.	Date	Status	Author (Beneficiary)	Description
0.1	24/11/2016	Draft	Dimitrios Vlachakis (UoP)	First draft from WP leader
0.2	6/12/2016	Draft	Konstantinos Deltouzos, Evangelia Pippa, Yiannis Ellul (UoP), Andreas Vasilakis, Ilias Kalamaras (CERTH)	First round of contribution collected
0.3	6/12/2015	Draft	Konstantinos Deltouzos (UoP)	Feedback by deliverable responsible: 1st draft sent out to all partners.
0.4	15/12/2016	Draft	Athanassios Kordas, Evangelia Pippa, Aimilia Papagiannaki (UoP), Andreas Vasilakis, Konstantinos Moustakas (CERTH)	Second round of contribution collected
0.5	20/12/2016	Draft	Konstantinos Deltouzos (UoP)	Feedback by deliverable responsible: 2nd draft sent out to all partners.
1.0	28/12/2016	Final	Dimitrios Vlachakis (UoP)	Deliverable finalized taking into account internal review's comments.
1.1	08/01/2018	Draft	Konstantinos Moustakas (CERTH)	First draft of the final deliverable containing contribution from involved partners.
1.2	16/01/2018	Draft	Aimilia Papagiannaki, Evangelia Zacharaki (UoP)	Second draft containing updates on the Frailty Indices.
1.3	22/01/2018	Draft	Thomas Papastergiou, Aimilia Papagiannaki, Evangelia Zacharaki (UoP)	Third draft containing updates on the analysis using tensors and multi-association rules.
1.4	25/01/2018	Draft	Spyros Kalogiannis, Aimilia Papagiannaki, Konstantinos Deltouzos (UoP)	Fourth draft containing updates on the data fusion.
1.5	26/01/2018	Draft	Konstantinos Deltouzos (UoP)	Fifth draft prepared by the deliverable responsible after reviewing all contribution.
1.6	30/01/018	Draft	Matteo Toma (SIGLA), Kosmas Petridis (Hypertech)	Sixth draft containing internal review comments.
2.0	31/01/2018	Final	Konstantinos Deltouzos (UoP)	Deliverable version (B) finalized considering internal review's comments.
2.1	28/06/2016	Revised	Evangelia Zacharaki (UoP)	Revised version to address the reviewers' comments after the Final Review in Luxembourg

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, behavioral, cognitive and physical activities of frail older people and apply them to manage and analyze new data. Results from the analysis of existing and new data will be also used to create user-profiling virtual models of older patients. Towards this direction, the deliverable D4.15 aims to examine methods of fusing information to extract frailty related indicators. These methods need to manage uncertainty in the system generated by incompleteness and noise of wearable sensor data.

In the preliminary deliverable, our primary efforts were focused on discovering a set of relevant and informative indicators for frailty. During this process, the state of the art was analyzed, and the clinical experts of our consortium gave their valuable input. Our preliminary work was performed on data mining techniques towards discovering associations between frailty, and physiological or behavioral patterns. These techniques aim to discover a novel way of mining multi-level association rules, in a distributed environment, from multiple heterogeneous data sources. Finally fueled by previous work on data fusion, three schemes were designed: (i) Early Integration scheme, (ii) Late Integration scheme with local (sensor dependent) training models, (iii) Late Integration scheme with global (sensor independent) training model.

In this final deliverable, we present our work which progressed in all of the areas mentioned above. We have revised section 2 and have included the new proposed Frailty Indicators which are based on the findings of the data analysis performed in WP4. In section 3, our progress on data analysis using tensor decomposition and multi-level association rules and its results are presented. Finally, in section 4 we present the fusion schemes developed for the multivariate data generated from the FrailSafe study and the results of the analysis methods which were performed.

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:	4.15	**Title:**	Signal processing algorithms for extraction of frailty related indicators (vers b)
Work package number:	4	**Title:**	Data Management and Analytics

Date of delivery	**Contractual**	**31/12/2017 (M24)**	**Actual**	**31/01/2018**
Status	Draft ☐		Final ☒	
Nature	Report ☒ Demonstrator ☒ Other ☐			
Dissemination Level	Public☒ Consortium ☐			
Abstract (for dissemination)	In this final deliverable, we present our work in defining frailty indicators, signal processing and data fusion. We present the state of the art on frailty indicators and proposed our own, which are based on the findings of the data analysis performed in WP4. Additionally we present our progress on data analysis using tensor decomposition and multi-level association rules and its results. Finally, we present the fusion schemes developed for the multivariate data generated from the FrailSafe study and the results of the analysis methods which were performed.			
Keywords	FrailSafe, signal processing, frailty indicators, data fusion			

Contributing authors (beneficiaries)	Evangelia Zacharaki, Spyros Kalogiannis, Aimilia Papagiannaki, Thomas Papastergiou, Konstantinos Deltouzos, Evangelia Pippa, Dimitrios Vlachakis, Athanasios Kordas (UoP), Ilias Kalamaras, Andreas Vasilakis, Konstantinos Moustakas (CERTH), Matteo Toma (SIGLA), Kosmas Petridis (Hypertech)			
Responsible author(s)	Konstantinos Deltouzos		**Email**	deltouzos@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*)

ALS	Alternating Least Squares
CANDECOMP	CANonicalDECOMPosition
ClFI	Clinical Frailty index
CoFI	Combined Frailty index
ECG	Electrocardiography
eFI	Electronic Frailty Index
FI	Frailty Index
GPS	Global Positioning System
HER	Electronic Health Records
EI	Early Integration
HOSVD	High-Order Singular Value Decomposition

LI	Late Integration
LSA	Latent Semantic Analysis
MMSE	Mini Mental State Examination
MoCA	Montreal Cognitive Assessment
PARAFAC	PARAllel FACtor analysis
PCA	Principal Components Analysis
SSH	Secure SHell
SVD	Singular Value Decomposition
TFI	Technical Frailty Index
THIN	The Health Improvement Network
XML	eXtensible Markup Language

1 Introduction

One of the key objectives of the FrailSafe project is the better understanding of frailty and the development of new quantitative and qualitative measures to define it. Towards this direction, the current state of the art of frailty definition was analyzed in order to discover the strengths and limitations of each method. In Chapter 2 , a summary of this work is being presented, and a special attention is being given to the electronic Frailty Index which can be generated automatically by health record data. During this process, the clinical experts of our consortium gave their valuable input. Fueled by the results of the data analysis performed in the WP4, we propose new Frailty Indicator.

Flowingly, in Chapter 3 we present our work on signal processing and data mining techniques. This work is focused so far in two directions. First is the modelling of the multimodal data that are being collected in FrailSafe, using tensors. By doing so, we can use their strong mathematical background and achieve several advantages such as data compression, identification of clusters and exploitation of patterns. The second direction aims to discover a novel way of mining multi-level association rules, in a distributed environment, from multiple heterogeneous data sources. The general architecture model and the future work on this are presented in section 3.2

Finally, in Chapter 4 our work in data fusion is presented. Several different approaches for fusing data from different sensor units/dimensions were explored, resulting in three schemes: (i) Early Integration scheme, (ii) Late Integration scheme with local (sensor dependent) training models, (iii) Late Integration scheme with global (sensor independent) training model. These schemes were tested and validated using the FrailSafe sensor data. In the end of the Chapter, the final fusion schemes developed for the multivariate data generated from the FrailSafe study and the results of the analysis methods which were performed are presented.

2 Frailty Indicators

2.1 Frailty phenotype and Frailty Index

The frailty syndrome has been widely discussed among the scientific community and its foundations are generally well established in the literature. However, its practical interpretation, particularly in the ordinary clinical practice, remains questionable. The combination of frailty measures in clinical practice is essential for the mediations and interventions design against age-related conditions (such as disability) in older people (1). Several methods have been developed lately in an effort to address this geriatric multidimensional syndrome.

The authors in (2), focused initially at some basic clinical manifestations of frailty, which were then projected into the frailty phenotype as it was described in the Cardiovascular Health Study (3). Along these lines, Rockwood et al. (4) utilized the Canadian Study of Health and Aging to create and approve their proposed Frailty Index. Additional methods to quantify frailty have been proposed over the last years, mainly expanding on these two models (5) (6) (7). Indeed, the frailty phenotype and the Frailty Index are considered by the scientific community as the cornerstone of frailty definition. These two methods follow a different approach, and thus should be considered complementary (8). Their main characteristics are summarized in Table 1.

Frailty phenotype	**Frailty Index**
Signs, symptoms	Diseases, activities of daily living, results of a clinical evaluation
Possible before a clinical assessment	Doable only after a comprehensive clinical assessment
Categorical variable	Continuous variable
Pre-defined set of criteria	Unspecified set of criteria
Frailty as a pre-disability syndrome	Frailty as an accumulation of deficits
Meaningful results potentially restricted to non-disabled older persons	Meaningful results in every individual, independently of functional status or age

Table 1: Main characteristics of the frailty phenotype and the Frailty Index

The frailty phenotype uses five distinct criteria that assess the appearance of signs or manifestations related to frailty (involuntary weight loss, slow walking speed, poor handgrip quality, reported exhaustion and mobility issues) (3). The quantity of criteria being met by the subject leads to a 6-level ordinal variable extending from 0 to 5. This is then sorted into a 3-level variable portraying a fit older person (none of the criteria), a pre-frail person (meets one or two criteria) and a frail person (meets at least three criteria). The frailty phenotype can be performed at the first meeting with the subject and does not require an in depth clinical assessment. In this way, it serves as a general categorization of the population into three distinct profiles. Overall, the frailty

phenotype does not give any specific guidelines about preventive or helpful mediations or interventions. The problem is that it is composed of extremely broad signs or side effects, which are only able to raise an alert about a potential health issue. This is not enough though to design a quick preventive or restorative intervention due to the fact that there is no information about the underlying cause of frailty. For instance, it is clinically impossible to treat sudden weight loss or slow gait speed without knowing the basic causal conditions. This is only possible by a thorough geriatric clinical assessment, in which the overall health status of the older person is being assessed through a multidimensional, interdisciplinary analytic process leading to particular clinical interventions.

The Frailty Index proposed by Rockwood et al. is made by a long list of clinical conditions, disorders, and diseases. There are more than 70 parameters as an initial screening tool that must be addressed. The actual goal of this list is to address and bring to the surface more critical deficiencies that have accumulated over the years. Although the Frailty Index has many times been revised and updated, the end goal remains to be able to clearly reflect on dichotomous conditions (e.g. robustness versus frailty). It is clear that the Frailty Index is impractical and inapplicable as a first contact tool for frailty diagnosis, since an extensive geriatric assessment of the older person must be conducted at the same time. Once the full assessment is complete, the Frailty Index can be used as a tool for monitoring the continuous follow up of the older person. Actually, the Frailty Index is more sensitive to changes than the overall frailty phenotype. Consequently, the Frailty Index might be of more use to help the clinician determine the effectiveness of any intervention that was designed and to depict the health status progress of the individual over time. In any case, the clinical intervention dependably goes through the Frailty Index's categorization into classes of frailty, separating normal ageing from anomaly. The categorization into risk groups of the frailty phenotype makes it more powerful as a tool that will link a typical clinical condition to frailty. In a clinical world that is constantly dominated by new advancements and developments, it can be of great value to formulate a complete geriatric assessment tool that can be generated by an electronic health record and serve as a reference for following assessments.

To sum up, it can't be overlooked that there are two noteworthy theoretical contrasts at the heart of the two frailty assessment tools:

(1) Relationship amongst frailty and age-related grouped conditions. As said, the frailty phenotype depends on the assessment of signs and clinical manifestations. This implies, as indicated by Fried et al. (3) (9) that frailty may hypothetically exist even without medically characterized conditions. Under such viewpoint, the frailty phenotype for sure portrays a novel age-related field of research for medical sciences (10). Then again, the Frailty Index revolves to a great extent around medical grouped conditions. It depicts a likeliness profile that is close to the one assessed by the clinician, which is possibly capable to characterize the phenotype frailty and to link it to its early signs as a preventive tool.

(2) Relationship of frailty with disability. In their review assessing the phenotype, Fried et al. (3) suggests that frailty causes disability that may not be linked to

(sub)clinical disorders. They clarify that 'the syndrome of frailty may be a physiologic precursor and etiologic factor in disability'. This implies a verifiable identification of frailty as a key element for the design and conduction of interventions against episodes that may result in disability. Along these lines, the frailty phenotype finds its optimal application in non-disabled more independent subjects. Then again, the Frailty Index incorporates measures of everyday incapacity (e.g. issues with getting dressed, issues with washing and reduced versatility) in its calculation (4). At the end of the day, the Frailty Index does not make a clear distinction between frailty and disability. It is more focused at impartially evaluating the measure of accumulated deficits of each individual, whichever they are.

These conceptual differences between the two instruments obviously and consequently differentiate the target populations to which they might be applied. As mentioned individual, while we may meaningfully estimate the Frailty Index in every case, the frailty phenotype may lose some of its clinical relevance when assessed in older persons already experiencing disability.

To summarize, the frailty phenotype categorically defines the presence/absence of a condition of risk for subsequent events (most specifically, disability). By differentiating a normal (i.e. robustness) versus an abnormal (i.e. frailty) status, the frailty phenotype may facilitate the implementation of the frailty concept into clinical practice. It provides the clinical-friendly dichotomous variable on which deciding the possible need of adapted care and/or interventions. Differently, the Frailty Index acts as measure of the organism capacity to accumulate deficits. It tells us how many clinical conditions are present and concur at exhausting reserves. Thus, the Frailty Index seems to act as an objective marker of deficits accumulation.

2.2 Electronic Frailty Index (eFI)

Recently there has been some research efforts on developing an electronic frailty index (eFI) that can be automatically populated from routinely collected data contained within the primary care EHR. In the work of (11) a study was performed using anonymized primary care electronic health record data contained in massive databases (ResearchOne, THIN). The eligible patients were aged 65–95 years and had permanently registered at the practice. Using a scoring system, the patients were categorized into four categories of frailty:

1. **Fit (eFI score 0 - 0.12)** – People who have no or few long-term conditions that are usually well controlled. This group would mainly be independent in day to day living activities.
2. **Mild frailty (eFI score 0.13 – 0.24)** – People who are slowing up in older age and may need help with personal activities of daily living such as finances, shopping, transportation.
3. **Moderate Frailty (eFI score 0.25 – 0.36**) – People who have difficulties with outdoor activities and may have mobility problems or require help with activities such as washing and dressing.

4. **Severe Frailty (eFI score > 0.36)** – People who are often dependent for personal cares and have a range of long-term conditions/multimorbidity. Some of this group may be medically

The deficits that were identified by the authors and were used to generate the eFI score are listed in Table 2. These deficits are not homogenous, as some refer to symptoms (i.e. dizziness, memory/cognitive problems), some are connected to disability (hearing/visual impairment) and some state the appearance of a disease (osteoporosis, Parkinson). A more graphical way to organize the deficits is depicted in Figure 1.

Activity limitation	Memory and cognitive problems
Anemia and hematinic deficiency	Mobility and transfer problems
Arthritis	Osteoporosis
Atrial fibrillation	Parkinsonism and tremor
Cerebrovascular disease	Peptic ulcer
Chronic kidney disease	Peripheral vascular disease
Diabetes	Polypharmacy
Dizziness	Requirement for care
Dyspnea	Respiratory disease
Falls	Skin ulcer
Foot problems	Sleep disturbance
Fragility fracture	Social vulnerability
Hearing impairment	Thyroid disease
Heart failure	Urinary incontinence
Heart valve disease	Urinary system disease
Housebound	Visual impairment
Hypertension	Weight loss and anorexia
Hypotension/syncope	
Ischemic heart disease	

Table 2: List of the 36 deficits used in the eFI

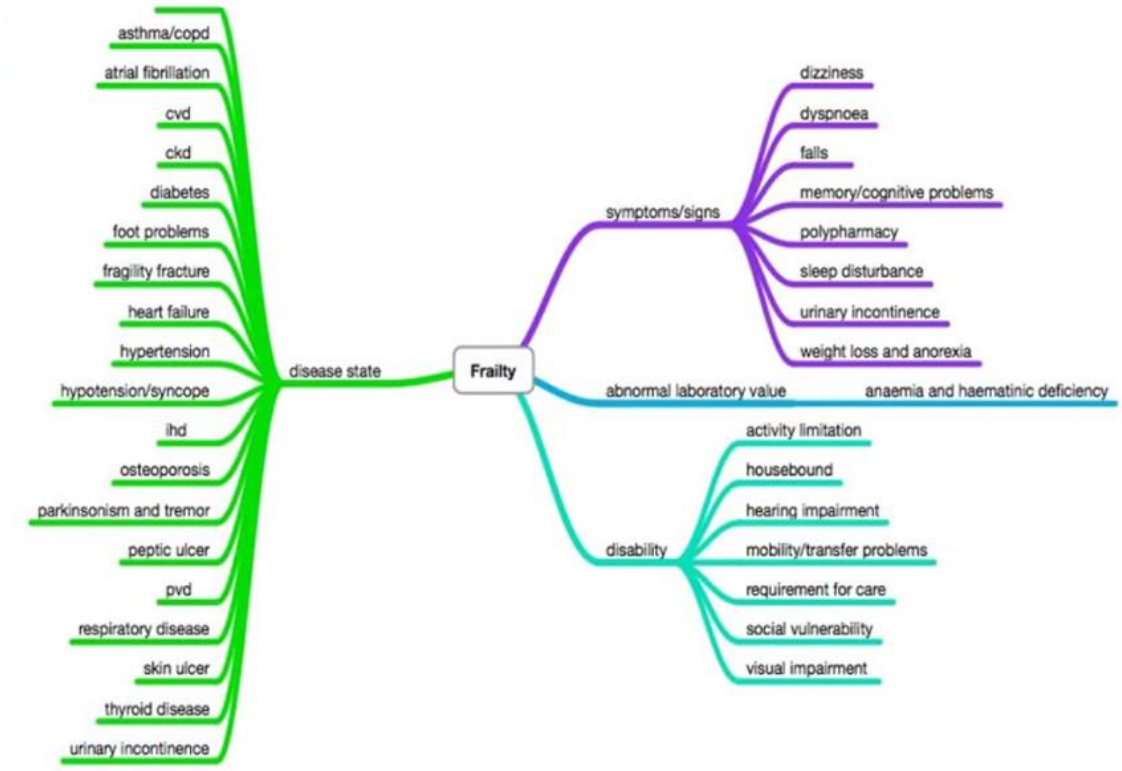

Figure 1: Map of the 36 deficits used in the eFI

The eFI has robust predictive validity and good discrimination for nursing home admission, hospitalization, and mortality. These outcomes are of particular importance for older people and health and social care systems internationally, and the predictive validity and discrimination characteristics of the eFI across all three outcomes add considerable weight to the clinical utility of the tool in terms of individual and population health planning.

An effort to link eFI to the FrailSafe data was made, with the aim of evaluating our population using an automatically-extracted score. For the purpose of adjusting eFI to FrailSafe's parameters, the variables (deficits) that eFI uses for extracting the frailty index had to be mapped to the FrailSafe corresponding variables. Since the 36 eFI variables constitute of 2000 sub-parameters and are therefore somehow abstract, the mapping could only be performed intuitively at the first place. 27 FrailSafe variables out of 36 eFI variables were identified as most similar and transformed to binary scale. The mapping between eFI variables and FrailSafe variables is shown in Table 3. As in the case of eFI score, all variables were equally weighted. The eFI score was estimated for 375 participants of FrailSafe. The histogram of the estimated eFI as well as the boxplot of eFI score versus the Fried status are depicted in Figure 2 and Figure 3 respectively.

Table 3: eFI to FrailSafe variables.

eFI variables	Corresponding FrailSafe variables
Activity limitation	Physical activity
Anemia and hematinic deficiency	Anemia
Arthritis	Arthralgies
Atrial fibrillation	Arrhythmia
Cerebrovascular disease	
Chronic kidney disease	
Diabetes	Diabetes Mellitus
Dizziness	Dizziness and/or Vertigo
Dyspnea	
Falls	Falls
Foot problems	
Fragility fracture	Fractures
Hearing impairment	Hearing problem
Heart failure	Heart Insufficiency
Heart valve disease	
Housebound	Leisure activities
Hypertension	Blood pressure
Hypotension/syncope	Blood pressure
Ischemic heart disease	Ischemic heart disease
Memory and cognitive problems	Impaired Cognitive Function, Cognitive evaluation (MoCA, MMSE, Memory complaint)
Mobility and transfer problems	ADL index
Osteoporosis	Osteoporosis
Parkinsonism and tremor	Parkinson's disease
Peptic ulcer	
Peripheral vascular disease	
Polypharmacy	Medication List
Requirement for care	Living conditions
Respiratory disease	Respiratory disease
Skin ulcer	
Sleep disturbance	Sleep problem
Social vulnerability	Exchange visits & telephone calls
Thyroid disease	Thyroid Gland Pathology
Urinary incontinence	Urinary incontinence
Urinary system disease	
Visual impairment	Eye disease
Weight loss and anorexia	Unintentional weight loss

As it can be seen from the boxplot, there is a relevant progression among the three Fried categories, according to the eFI estimated score. Although an overlapping exists between the categories, the results are rather encouraging considering the abstract way in which the variables' mapping was performed. The effort of exploiting eFI score is planned to be continued, by collaborating with the clinical team in order to define

the corresponding FrailSafe variables more accurately, with respect to the eFI deficits. Additionally, we plan to request for a more detailed description of the eFI variables, in order to systematically migrate from the eFI standard score, to a corresponding score for the FrailSafe participants.

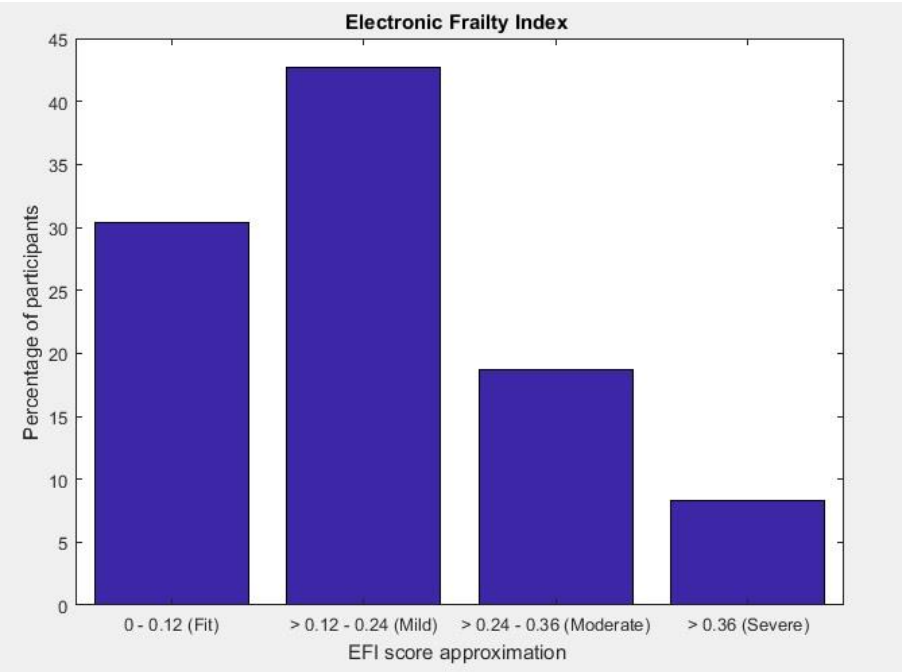

Figure 2: 2The estimated eFI score for the FrailSafe participants

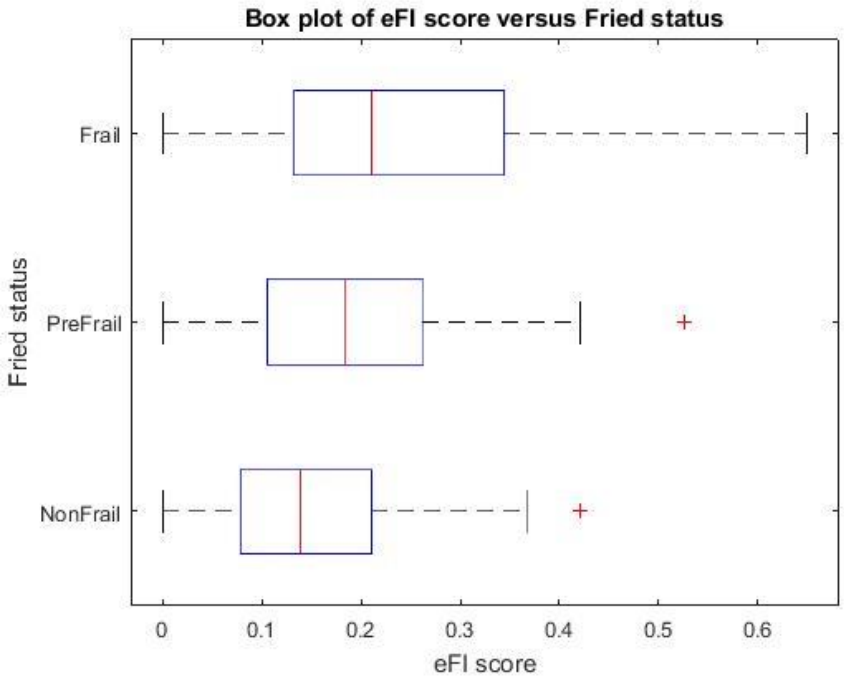

Figure 3: The range of estimated eFI values for the 3 groups classified by Fried at M24.

2.3 Definition of Frailty Indices in FrailSafe

The clinical partners of the consortium analyzed the state of the art of frailty definitions as part of deliverable D2.1, and defined new frailty indices which are going to be used in FrailSafe. These indices aim to define the loss of reserve, independently of frailty status as this is defined by Fried's criteria, in order to render clinical results measurable.

On the other hand, FrailSafe Database contains variables at different time points from:

- Clinical Evaluation
- Follow up assessment
- FrailSafe system metrics

In this scope, a new combined index (**Combined Frailty index: CoFI**), that will express frailty status relevant to the study's measurements, will be created by adding up two other frailty indices derived from the study, the **Clinical Frailty Index (ClFI)**, corresponding to the results of the clinical evaluation, and the **Technical Frailty Index (TFI)**, corresponding to the metrics derived from the FrailSafe system devices.

Each time a programmed clinical evaluation is effectuated, a ClFI score will be calculated, which will be composed by several items that correspond to various aspects of frailty, as they are described by the clinical evaluation sub-questionnaires. Similarly, a TFI will be calculated for each FrailSafe system installation, practically, for each FrailSafe home visit. Finally, a combined FI, by adding up ClFI and TFI will be calculated. A summary of these indices in shown in Table 4 and more can be found in deliverable D2.1.

Table 4: Frailty Indices definition

ClFI (Clinical Frailty Index): score corresponding to the findings of the clinical evaluation in a time-spot
TFI (Technical Frailty Index): accumulated score derived from the FrailSafe system metrics during certain time intervals of observation
CoFI (Combined Frailty Index): combined Clinical and Technical frailty score

Toward defining the Clinical Frailty Index, two different indices were computed and their predictive ability towards the development of a frailty index was examined. The first one aims to predict the discrete Fried classification score (FI_1) and the second one tries to estimate a continuous score as a linear combination of the 5 criteria related to Fried classification (FI_2).

For the FI_1 score, lasso linear regression was performed to select a subset of variables and estimate their β coefficients, in order to build a predictive model having the best possible correlation with Fried's score:

$$FI_1 = \sum_{j=0}^{m} \beta_j x_j \approx Fried,$$

where x_j are the variables from the clinical evaluation,m is the number of β_j coefficients (number of variables), β_0 is the intercept and $x_0 = 1$.

For the *FI₂* score, a similar analysis was performed but this time the variable to be predicted is a continuous frailty index (let's denote it with Y) expressed as a linear combination of the 5 criteria related to Fried score:

$$Y = \sum_{k=1}^{5} \alpha_k f_k, \text{ subject to } \sum_{k=1}^{5} \alpha_k = 1,$$

where $f_k \in$ {involuntary weight loss, slow walking speed, poor handgrip quality, reported exhaustion, low physical activity}. Again, the score that takes into account all clinical variables x_j is calculated by:

$$FI_2 = \sum_{j=0}^{m} \beta_j x_j \approx Y$$

where m is the number of β_j coefficients (number of variables), β_0 is the intercept and $x_0 = 1$.

The details of these scores and some preliminary results can be found in deliverable D4.2. The ultimate goal is to investigate whether the proposed frailty indexes are more reliable predictors of frailty transition than standard classification scores. This hypothesis will be assessed in the evaluation phase.

Finally, for defining the Technical Frailty Index, we are investigating tensor decomposition approaches (described later in this deliverable) and deep learning techniques for an in-depth analysis of the time series data and for the seamless extraction of a features' hierarchy that will be linked through a deep neural network to a frailty index. An intrinsic challenge lies in the non-uniformity of the data in respect to duration, activities performed during the recordings, as well as the low quality of some signals due to their acquisition in a real-life home environment and not in a controlled experimental setting. A detailed description of the method can be found in the deliverable D4.16, while final results will be reported in D4.17 where the largest amount of data will be available, absolutely necessary for deep learning techniques.

3 Signal processing and data mining techniques for extracting frailty related indicators

3.1 Analysis using tensors

The traditional approach to data representation utilizes a matrix structure, with observations in the rows and features in the columns. Although this model is appropriate for many datasets, it is not always a natural representation because it assumes the existence of a single target variable and lacks a means of modeling dependencies between other features. Additionally, such a structure assumes that

observed variables are scalar quantities by definition. This assumption may not be valid in certain domains where higher-order features predominate, or in domains which have strong spatiotemporal components, such as ECG signals.

Traditionally, these problems have been solved by reducing the features to scalars and fitting the dataset to a matrix structure. However, as well as potentially losing information, this strategy also employs a questionable approach from a philosophical standpoint: attempting to fit the data to an imprecise model rather than attempting to accurately model the existing structure of the data. Finally, while it may be possible to model dependencies between features by repeating the methodology multiple times, each with a different target variable, this yields suboptimal performance and may not be computationally feasible when real-time performance is required or when the dataset is very large.

To address these issues, we propose to model such datasets using *tensors*, which are generalizations of matrices corresponding to multidimensional arrays. To formulate the use of tensors, we first need to establish some basic notation. The *order* of a tensor is the number of its dimensions. So, a 3rd order tensor is a three-dimensional array, like the one shown in Figure 4.

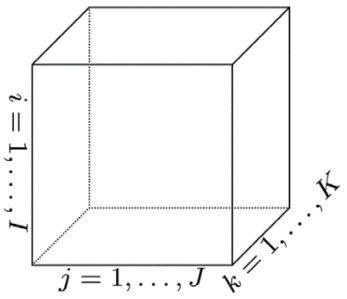

Figure 4: Third order tensor (source (12)).

Now if someone is dealing with *N* dimensions, the corresponding tensor will be an Nth order tensor. It is very common that tensors are treated either as sets of *fibers* or as sets of *slices.* A fiber is the higher order analogue of matrix row and column and it is defined by fixing all but one indices of a tensor. A slice is a two-dimensional part of a tensor and it is defined by fixing all but two indices of a tensor. Figure 5 depicts shows all possible slices and fibers of a 3rd order tensor.

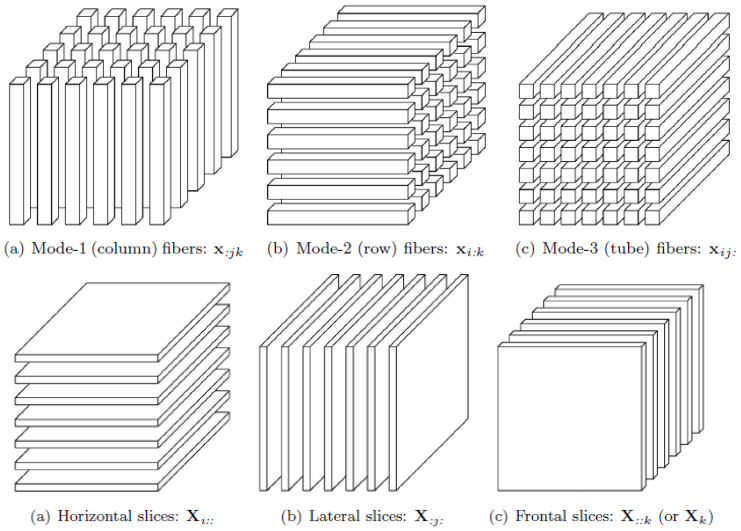

Figure 5: Slices and fibers of a third order tensor (source (12)).

3.1.1 Background on tensor decomposition and motivation for the FrailSafe project

Having defined the structure of a tensor, one must go further to examine the value of representing their data with a tensor. This is where tensor decompositions enter to manipulate even the most pretentious sets of datasets with high dimensionality. Tensor decompositions, which are an extend of matrix decompositions coming from linear algebra, have a wide range of application including data mining, information retrieval, neuroscience, signal processing and many other problems. Their success lies on their ability to capture multi-linear and multi-aspect structures of high-dimensional datasets. The two most widely used tensor decomposition models are PARAFAC and Tucker. PARAFAC, as well as Tucker decomposition are the higher dimensional analogous of the widely known methods *Principal Component Analysis (PCA)* and *Singular Value Decomposition (SVD)*.

Singular Values Decomposition: Singular values decomposition (SVD) is a unique matrix factorization by which and $m \times n$ **A** matrix is decomposed into two projection matrices and a core matrix, as follows:

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^{\mathbf{T}}$$

where $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{U}$ is an $m \times r$ column-orthonormal projection matrix, $\mathbf{V}$is an $n \times r$ column-orthonormal projection matrix, and $\mathbf{\Sigma}$ is a diagonal $r \times r$*core matrix*, where r is the rank of the projection. Singular value decomposition has a wide variety of applications: for example, truncation of the SVD coefficients provides an optimal low-rank approximation (i.e. minimizes the Frobenius norm). This indicates a close relationship between SVD and Principal Component Analysis (PCA). SVD is also used to discover the rank of a matrix, find the pseudoinverse, and solve least squares minimization problems. Additionally, the solution to SVD may be used in an

unsupervised summarization technique known as Latent Semantic Analysis (LSA) (13). In this technique, $\mathbf{A}$ is treated as a term-document matrix. Here, singular value decomposition automatically derives a user-specified number of latent *concepts* from the given terms which form a basis for the rows and columns of the matrix. The projection matrices $\mathbf{U}$ and $\mathbf{V}$ then contain term-to-concept and document-to-concept similarities, respectively. Thus, SVD can be used to provide simple yet powerful automatic data summarization. This technique may be naturally viewed as a form of co-clustering, in which the rows and columns of a matrix cluster to the same space. An alternative graphical interpretation exists, in which clusters represent shared "waypoints" through which edges pass between vertices. Use of the eigendecomposition or SVD is also common in a graphical context, where it is known as *spectral graph theory*; here a common technique is to cluster on the eigenvector corresponding to the second smallest eigenvalue of the Laplacian matrix, thereby partitioning vertices along edges which are likely to be minimal cuts. This technique is known as *Fiedler retrieval*. It is also possible to project new query vectors into the space defined by the SVD, known as *folding in*; this enables recommendation as the query projects to the same space as both the rows and columns and can be assessed using a distance metric.

Matrix decomposition (factorization): In general matrix decomposition is the factorization of a $m \times n$matrix **X** in a product of two matrices $\boldsymbol{A} \in \mathbb{R}^{m \times R}$ and $\boldsymbol{B^T} \in \mathbb{R}^{R \times n}$, where R is the rank of matrix **X**.

Let us define the rank of a matrix in a slightly different manner that can be immediately applied to higher order tensors. A rank-one matrix **X** is a matrix that can be written as the outer product of two arrays:

$$\boldsymbol{X} = a_1 \circ b_1 = a_1 {b_1}^T$$

where$a_1 \in \mathbb{R}^m, b_1 \in \mathbb{R}^n$. The rank of a matrix can now be defined as the minimum number of rank-one components that sum up to **X**.

If matrix **X** is of rank R then it can be written as sum of exactly R rank-one components:

$$\boldsymbol{X} = \sum_{i=1}^{R} a_i \circ b_i = \boldsymbol{AB^T}$$

Even though the rank of a matrix and his SVD (given the orthogonality of the projection matrices) are unique, that is not the case for the matrix decomposition. If the SVD of **X** is $\mathbf{U\Sigma V^T}$ we can choose $\mathbf{A = U\Sigma}$ and **B=V**. Furthermore, we can easily also choose $\mathbf{A = U\Sigma W}$ and **B=VW** where **W** is a random orthogonal matrix. This fact clearly shows the non-uniqueness of the matrix decomposition, since we have different solutions due to the random W orthogonal matrix.

The rank of a matrix, as defined before, can easily be extended to the high order case. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$be an order 3 tensor. We say that $\boldsymbol{\mathcal{X}}$is of rank-one if it can be expressed as an outer product of three arrays:

$$\boldsymbol{\mathcal{X}} = a_1 \circ b_1 \circ c_1$$

We can further define the rank R of tensor $\boldsymbol{\mathcal{X}}$ as the minimum number of rank-one components needed to express $\boldsymbol{\mathcal{X}}$. So, if $\boldsymbol{\mathcal{X}}$ is an order 3 rank R tensor it can be written as:

$$\boldsymbol{\mathcal{X}} = \sum_{i=1}^{R} a_i \circ b_i \circ c_i$$

where a_i, b_i, c_ican be thought as the i-th columns of factor matrices$\boldsymbol{A} \in \mathbb{R}^{I\times R}$, $\boldsymbol{B} \in \mathbb{R}^{J\times R}$, $\boldsymbol{C} \in \mathbb{R}^{K\times R}$ respectively. Even though the rank of a tensor is defined exactly as the rank of a matrix the two ranks have different properties. One major difference is that, in contrast to the matrix rank, the problem of finding the rank of a tensor is an NP-hard problem (14).

The definition of the rank of a tensor as a sum of rank-one components leads us straightforward to the definition of the CANDECOMP/PARAFAC decomposition. PARAFAC(16) is a generalization of Principal Component Analysis (PCA) and forms the basis of our tensor analysis approach. Given a user-specified number of concepts *R*, PARAFAC decomposes an order-*k* tensor $\mathcal{A}$ into a sum of R rank-one tensors, each of them consisting of the outer product of the columns of the projection matrices, denoted $\mathbf{U}^{(1)} \ldots \mathbf{U}^{(r)}$. Formally, we define the decomposition as follows:

$$\mathcal{A} = \sum_{\mathrm{i}=1}^{\mathrm{R}} \lambda_{\mathrm{i}} \mathbf{U}_{:,i}^{(1)} \circ \mathbf{U}_{:,i}^{(2)} \circ \ldots \circ \mathbf{U}_{:,i}^{(\mathrm{k})}$$

Where the $\mathbf{U}$ matrices represent projection matrices containing mode-to-concept similarities and λ represents a *R*-element scaling vector, in which each element represents the strength of a concept. The notation $\mathbf{U}_{:,i}$ refers to the entire *i*-th column of $\mathbf{U}$. The number of columns in each projection matrix will therefore be equal to the number of user-specified concepts and the number of rows in each individual projection matrix equal to the number of observations made on the corresponding mode of the tensor.

Unlike for the matrix factorization problem where uniqueness is not the case, the PARAFAC decomposition can be unique under very mild conditions. That means that there is only one combination of factor matrices $\mathbf{U^{(i)}}$ such that the sum of the outer product of their columns can be equal to the tensor. The uniqueness of the decomposition is true except of the permutation of the columns of the factor matrices and the scaling of them. The most general and well-known result of the uniqueness comes from Kruskal (18) and it is based on the k-rank (kruskal-rank). The k-rank of a matrix can be defined as the maximum number k such that any k columns of a matrix are linearly independent. The criterion of the uniqueness of the PARAFAC decomposition according Kruskal is:

$$k_{\boldsymbol{A}} + k_{\boldsymbol{B}} + k_{\boldsymbol{C}} \geq 2R + 2$$

where R is the rank of a 3rd-order tensor and **A**, **B** and **C** are the factor matrices of the PARAFAC decomposition.

An important aspect of PARAFAC is that it can easily handle missing values. There are plenty of methods exploiting different algorithms such as ALS (21),(22) or the Proximal Algorithm (23) that can be applied to big data. The PARAFAC model is a good model for handling missing values and noise.

Moving a step beyond the original use of the tensor decompositions, there has been a great effort to exploit the tensors' structure and tools considering several data mining problems including clustering, feature extraction and classification. A feature extraction and classification problem from a tensor decomposition point of view can be defined as follows:

Consider a set of K training samples (a set of arrays formulated as slices within a tensor) corresponding to C classes and a set of test data (test slices of a tensor). The challenge is to find appropriate labels for the test data. The latter can be performed in the following steps:

- *Find a set of basis matrices and corresponding features for the training data.*
- *Perform feature extraction for test samples using the basis factors from the previous step.*
- *Perform classification by comparing the test features with the training features.*

One can easily understand that the above problem is an extension of the Tucker decomposition model, considering the factor matrices as the basis matrices, and the core tensor as the feature representation. The compressed core tensor is of much lower dimension than the original data, making it a fruitful option for dealing with the classification problem using as little resources and time as possible. The above method was proposed by Phan and Cichockiin(25)and was tested for handwritten digits, BCI motor imagery and image classification. A simplified scheme is shown below.

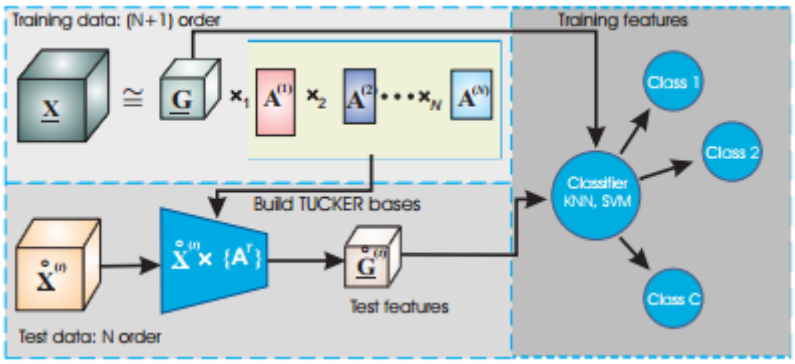

Figure 6: Classification diagram based on TUCKER decomposition (source (25)).

Another interesting approach on the co-clustering problem was introduced in the work of (26), where the idea was to use PARAFAC decomposition with sparse latent factors in order to extract tri-clusters from the original data. The uniqueness of the decomposition along with the sparsity constraint impose that a large number of possibly overlapping co-clusters will arise.

Considering the multimodal nature of the data collected through the FrailSafe system, the analysis should be headed towards representing sensory, physiological and device-related data with tensors. More specifically, each data source can be thought of as a dimension (a mode) in an *N*-dimensional tensor (where *N* is the number of the different kind of data sources). For each different sample of the data collected (a sample can be a set of data coming from a specific time course), there is a distinct tensor created, which in fact belongs to a set of *K* training tensors. For each of these samples we suppose that the knowledge about the frailty condition exists (aka the class to which a training tensor belongs to). By decomposing the whole set of *K* (concatenated) tensors we obtain the core tensor which is equivalent to a set of features for each sample.

Tensor-based aided methods such as the one mentioned above for the purpose of knowledge discovery through multiple-sources' datasets are considered a great option in the context of FrailSafe system development. Properly extracted underlying structures through the joint analysis can bring into light hidden frailty-related components, which in turn will be given as an input to the Virtual Patient Model as well as the monitoring system of FrailSafe. Picking robust methods in order to set up a valuable human-oriented system is the cornerstone of future success.

3.1.2 Applying tensor decomposition to FrailSafe data

In the context of FrailSafe, data are collected from individuals in different sessions, using the two different products developed by 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

Signal Preprocessing. The signals acquired from the sensors have different frequencies, thus, time synchronization of the channels had to be performed. All signals have been synchronized at 25Hz by interpolating the missing values. In the second step of the preprocessing procedure signal segments of low quality are

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

discarded according to the quality indication provided by the devices. In the third phase time frames (e.g. of 1-minute duration) are extracted using the sliding windows technique. The intuition behind this segmentation is that we want to capture separate activities of the older people and thus a time frame of 1 minute is considered to be satisfying for that purpose.

Tensors construction. 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. At this phase we have a distinct tensor for each individual. These tensors have constant dimensionality in the 2nd and 3rd way (dimension), since they share the same time points (*J* =1500) and the same channels (*K* = 7). A challenge in our analysis is that the tensors, corresponding to each individual have different dimension in the time frame mode. This is due to the different acquisition lengths, as well as the different amount of signal dropouts due to bad quality.

Model construction. In order to construct a consistent model suitable for frailty prediction we concatenate the tensors corresponding to each individual, along the first dimension. The resulting tensor is of dimensionality 19740x1500x7 containing 207 million observations. We will use the PARAFAC decomposition for modeling the data and in order to do so we need to have a coherent model for all the individuals.

Unwanted Outliers. Besides the preliminary cleaning of the data, based on the quality of the recordings, we performed a more elaborate cleaning procedure since we discovered outliers in some measurements that are probably due to bad recordings. For example, we discovered in some channels recordings that lie outside of the normal values. These channels include the breathing rate and the heart rate channel. We will briefly mention the data cleaning procedure, which is the same (in respect to features and thresholds) with the procedure followed in D4.17 (section 3.2.2).

1. **Breathing rate channel:** we discard recordings with values outside the range $[8, 50]$ $breaths/minute$.
2. **Heart rate channel:** we discard recordings outside the range $[40, 200]$ $beats/minute$, as they are non-accepted values of heart rate.
3. **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). This is the same step as the one described in D4.2.
4. We keep only these time windows, which have in each channel more than 10% non-outlier values.

We should clarify that the range of acceptable values of breathing and heart rate that defines inliers is much larger (in both directions) than the acceptable range used for generating alerts (defined in D4.17). Outliers correspond to undoubtedly erroneous measurements that should be discarded and that are outside the range of measurements obtained in pathological conditions.

In steps 1-3 we discard some unwanted values from the data and in step 4 we discard time windows having less than 10% of information. This procedure produces data that

have missing values up to 90%. This will not be a problem since as aforementioned we can easily handle missing data using the PARAFAC model.

The output of this "cleaning" procedure is a tensor of dimensions $10757 \times 1500 \times 7$ containing 101,275,188 observations and 11,673,321 missing values, making the missing values percentage be 10,34% (a very low missing value rate).

Handling missing values. After the cleaning process that we described earlier we come up with a tensor containing about 10% missing values. We can model our data using the PARAFAC model relying only on the observed using ALS (21) or a proximal algorithm like in (23) for the optimization. The algorithms will compute the factor matrices of the model relying only of the observed values (having excluded the unwanted outliers). Using the factor matrices, we can reconstruct the full tensor by calculating the outer products of the corresponding components and summing up all the individual components as described earlier. Our intent is rather having a model that contains less noise than the upper limit for efficient tensor reconstruction. For that purpose, we will use only the factor matrices.

Both methods that we use make an efficient memory management relying on sparse tensors where missing values are represented as zeros. The fact that our data contain also zeros will cause that these observed zeros will be treated as missing values. To overcome this difficulty, we make some preprocessing steps before we run the algorithms we first denote missing values with NaN, then add a small number (eps – the machine precision) to all our data, in order not to have zeros in the data set, and finally we replace NaN's by zeros. After the third step all missing values are modelled by zeros, zeros have a machine precision small values (that adds a little noise to the data, but as PARAFAC can handle noisy data this will not have a big effect on the factor matrices) and so we can apply the aforementioned memory efficient algorithms.

Tensor rank selection. A crucial aspect of modeling multidimensional data using the PARAFAC model is the selection of the decomposition rank of the model. As we mentioned earlier the problem of finding the rank of a N-dimensional ($N \geq 3$) tensor is not trivial as in the matrix case. In fact, this problem is an NP-hard problem and no straightforward algorithm exists. On the other hand, we do not need to find the exact rank of the tensor as we are interested in a low rank approximation of it, since we want our model to capture the latent concepts inside the data. As we will see below if we select a very low rank for the model, the recordings will be smoothed out and the model will capture only a small part of the underlying concepts and in a fuzzy manner. On the other hand, if we select a very high rank for our model, we over-represent the data, losing the opportunity to capture the underlying concepts.

In Figure 7 we observe the recordings per channel for 3 random individuals belonging to the three classes (non-frail, pre-frail, frail). The recordings were scaled across channels and missing values at this moment are modelled as NaN's, thus they don't appear in the figure. In the same figure we illustrate also the corresponding landscapes for all three time-windows of time points versus the channels.

When we use a very small rank for the PARAFAC decomposition (e.g. 7, as the number of the smaller dimension) we obtain a smoothed version of the signals, that are informative capturing some underlying concepts, but still not informative enough, as we see in Figure 8.

In order to gain some insights of the PARAFAC modelling and since the visualization for high values of rank is not so easily interpretable, we report in Figure 9 the loadings of the PARAFAC model with rank 7.

The first row of Figure 9 shows the loadings of factor matrix **A** that correspond to the time windows dimension. These loadings will be used as features for the frailty prediction procedure.

In the second row of the Figure 9 we depict the loadings that correspond to the time points dimension. These time series of length equal to a time window correspond to the latent signal segments that are present in our data. In other words, these are the principal time series patterns that describe the data.

In the third row of Figure 9 we can see the loadings corresponding to the third dimension. These graphs inform us about the different patterns that exist in the data concerning values across all the channels.

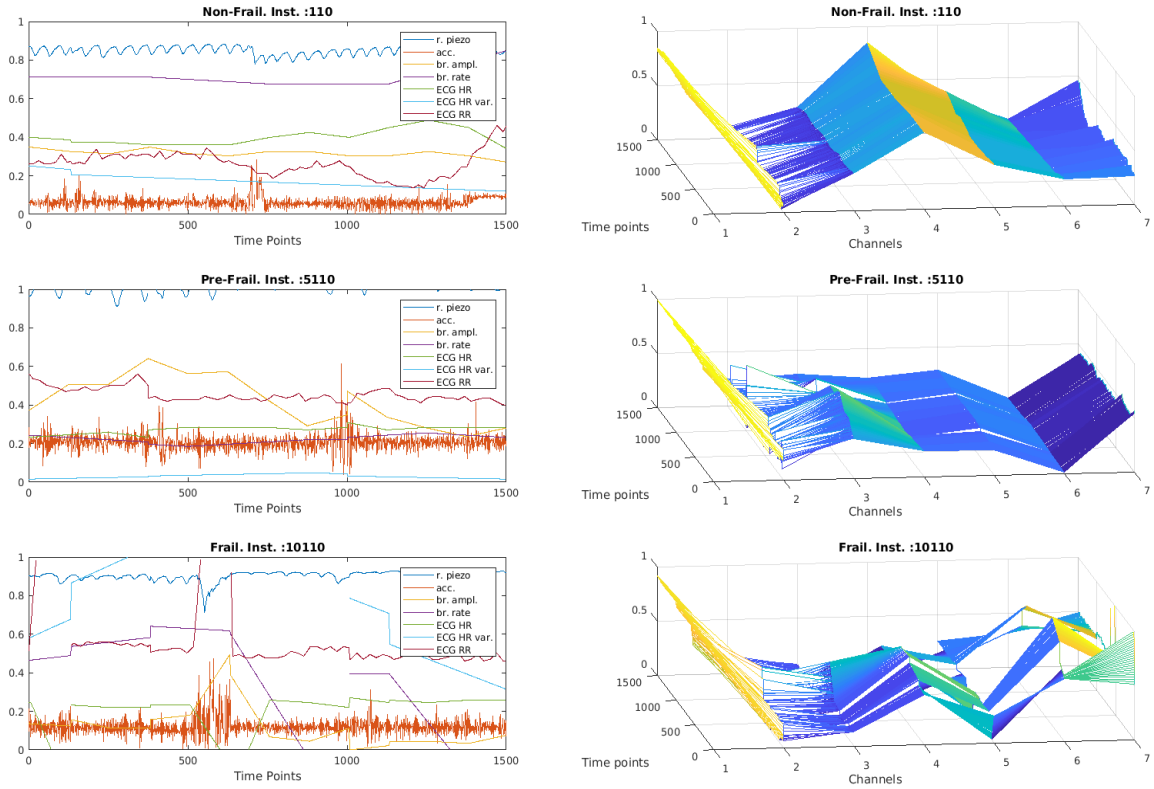

Figure 7: Three random time windows from three distinct subjects (from top to bottom: non-frail, pre-frail, frail). The first column shows the scaled recordings (containing missing values) and the second column shows the corresponding landscapes of time versus the channels.

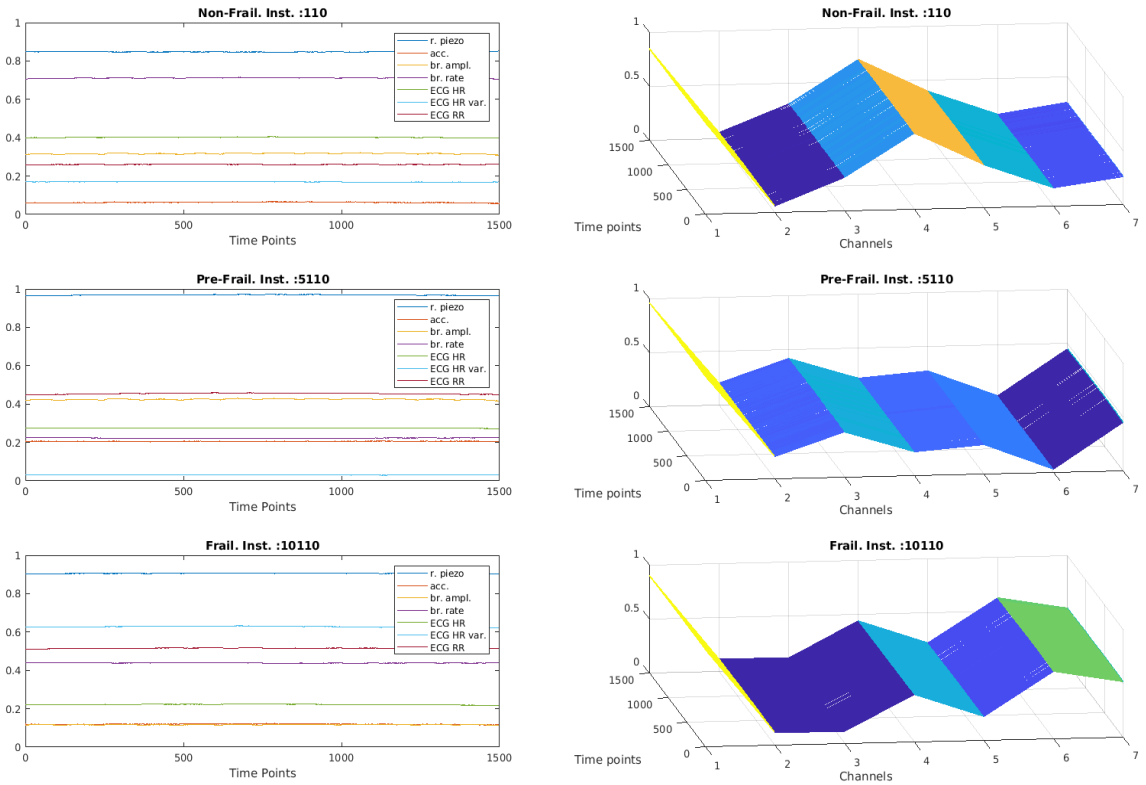

Figure 8: The reconstructed time frames from Figure 7 when using factorization rank = 7.

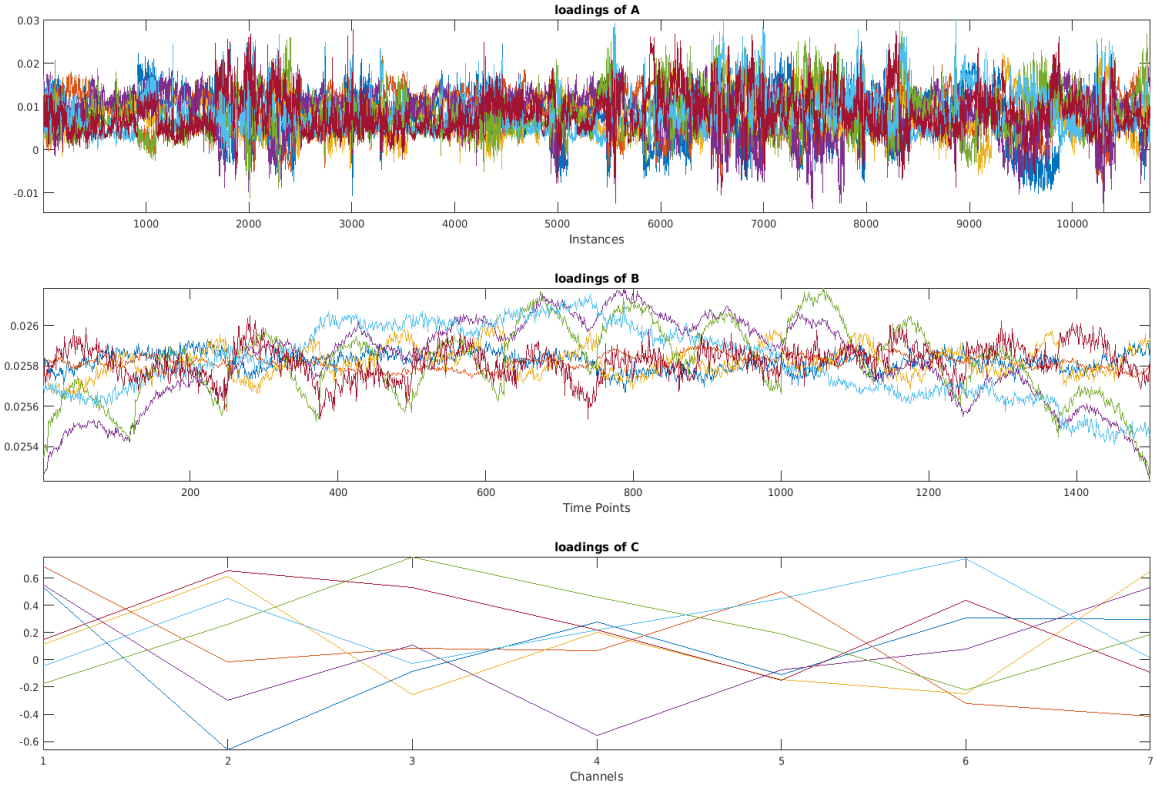

Figure 9: The loadings of the PARAFAC model with rank 7. From top to bottom: the loadings of the 1st dimension (time windows), the loadings of the 2nd dimension (time points) and the loadings of the 3rd dimension (channels).

Feature extraction. After the preprocessing steps mentioned above we are ready to extract features for the classification task. The features will be extracted using the PARAFAC model. Our goal is to extract features for each time window $x_{i,:,:} \in \mathbb{R}^{J \times K}$, where $x_{i,:,:}$ is the i-th slice of tensor $\mathcal{X}$. Each time window can be written as $x_{i,:,:} \approx \sum_{r=1}^{R} a_{ir}(b_r \circ c_r)$ using the PARAFAC decomposition, 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.

3.1.3 Predicting the frailty status from sensor data

After modeling the data using the PARAFAC decomposition we need a method that classifies the subjects based on FrailSafe variables. The challenges that arise in this procedure are the following:

1. The recordings for each individual are not of the same length and furthermore due to the preprocessing each individual ends up having different number of time frames.
2. The labels provided for the classification task correspond to the overall status of the subject and not to each individual time frame. That means that we are not aware if a particular time frame can be descriptive of the frailty status of the older person.
3. We do not have any information about the activities of the subjects and thus we cannot match the available time frames to certain activities, a fact that could be beneficial for our analysis.
4. The whole data set is highly unbalanced.

To deal with some of these challenges we propose a two phases approach. In the first phase we train a regression model using the instance- (i.e. time window-) wise features extracted by the PARAFAC model, for predicting the frailty status for each time window. In the second phase we use the outcome of the first classification and we train a subject-wise model, where the features capture the distribution of the outcomes of the instance-base classification. The final outcome of these two phases is the prediction of the frailty status of the individuals.

Phase 1: Regression. As aforementioned, by modeling the data using the PARAFAC model we can use as features, for classifying time windows (that correspond to the first mode of our tensor) the first factor matrix of the PARAFAC decomposition. We refer to this matrix as matrix **A**. Each row of matrix **A** corresponds to a time window of our data, thus we can divide the rows of this matrix to a train and a test set.

The subject-wise labels (non-frail, pre-frail and frail) are provided by the clinical evaluation of the subjects. We model these three labels as {0, 1, 2} assuming that the frailty status of a subject is increasing when moving from non-frail (label 0) to frail (label 2). We assume furthermore that the instances (i.e. time windows) of each subject inherit these labels, (i.e. every time window gets assigned the frailty status of the corresponding subject). Although this assumption is not an accurate one, since not

all the time windows of a recording are related to his/her frailty status, we imply that the outcome of the frailty status prediction can be an indication of the degree of the relation of the corresponding activity to the frailty index of the subject. E.g. a predicted value near 1 indicates that the activity (and the subject) has to be related with the pre-frail status.

The rank of the PARAFAC model chosen corresponds to the number of "latent" variables that will be used for fitting a regression model. Since the rank of a tensor is a crucial parameter for a valid and accurate model and since the number of variables we will use to fit a regression model can affect the test accuracy of the model (e.g. too many variables chosen can lead to a model with poor generalization abilities) we perform Principal Components Analysis on the feature matrix, for acquiring the most informative variables for the regression model.

Having assigned, to each time window the corresponding label and performed PCA on the acquired feature matrix we fit a quadradic regression model to our training set. The outcome of the regression, which is for each time window, will be used in the next step for training a subject-wise classification model.

Phase 2: Subject-wise classification. In the second phase of the classification process we build a subject-wise classifier by a fusion of the outcome of the first classification phase. For this purpose, we construct normalized histograms of the prediction of the first classifier for each individual subject. In this way we capture the distribution of the frailty status predictions from the first classifier. Based on these histograms we fit an LDA (Linear Discriminate Analysis) classifier using as features the subject-wise normalized histograms of frailty status predictions.

Figure 10 depicts in a graphical manner the entire process for the frailty index prediction. Algorithm 1 presents the training process of the Regession&LDA (Reg&LDA) algorithm.

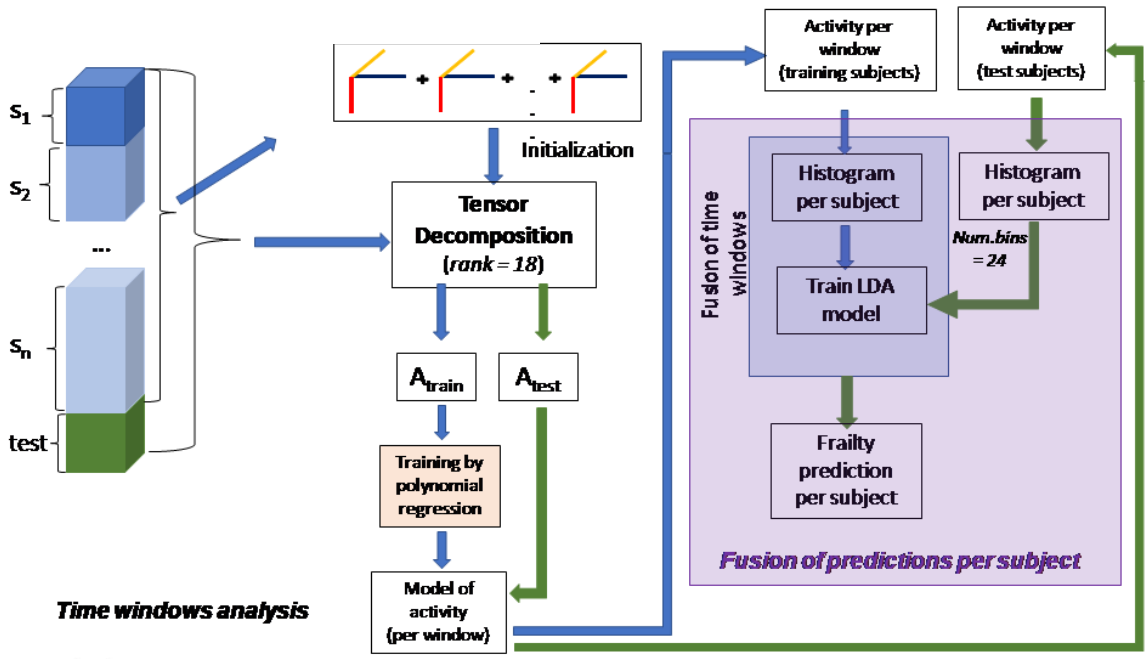

Figure 10: The architecture of the frailty status classifier.

Algorithm 1: Reg&LDA(training) for frailty prediction

Input: training instances' features$\mathbf{A_{tr}}$, instances (i.e. time window) training labels $\mathbf{Y_{tr}}$, subjects training labels$\mathbf{YSub_{tr}}$, the level of description of the feature matrix**svPer**, the **nrBins** the nuber of bins used for the histograms

Output: model for frailty status prediction

1. Concatenate $\mathbf{A_{tr}}$ and $\mathbf{A_{ts}}$ along the first dimension into a matrix $\mathbf{A_{m.}}$
2. Perform PCA on the concatenated matrix $\mathbf{A_{m.}}$ and get the scores T_{mTrunc}, using the m-leading singular values that describe the **svPer** % of matrix $\mathbf{A_{m.}}$
3. Split the truncated scores matrix $\mathbf{T_{mTrunc}}$, into the corresponding $\mathbf{T_{tr}}$ and $\mathbf{T_{ts}}$ scores matrix.
4. Train a quadratic regression model using $\mathbf{T_{tr}}$ and $\mathbf{Y_{tr}}$ and get frailty index predictions $\mathbf{FIpred_{tr}}$ for each instance (i.e. time window)
5. Find the **nrBins**-quantiles of the $\mathbf{FIpred_{tr}}$**(CutPoints)**
6. For each instance calculate the normalized histogram using as cutting points **CutPoints** and construct the $\mathbf{ASub_{tr}}$feature matrix
7. Fit a pseudo-quadratic LDA model to $\mathbf{ASub_{tr}}$ and $\mathbf{YSub_{tr}}$

Results. We present here the results of the Reg&LDA algorithm. The data used for the evaluation of the algorithm are the physiological measurements (available until M24) from the FrailSafe participants. In Table 5 we give an overview of the data. In the parenthesis we report the percentage relative to the sum of the data. In the data there are subjects that have both strap and vest sessions. We observe that our dataset is unbalanced as concerns the classes, in terms of subjects per group as well as in terms of available time windows.

Table 5: FrailSafe physiological measurements data (available until M24).

	Strap	**Vest**	**Sum**
Subjects	***90 (85,71%)***	***27 (25,71%)***	**105 (100%)**
Non-frail	*39(37,14%)*	*9 (8.57%)*	44 (41.9%)
Pre-frail	*42 (40%)*	*13 (12.38%)*	49 (46.67%)
Frail	*9 (8.57%)*	*5 (4.76%)*	12 (11,43%)
Time windows (TW)	***7006 (66.69%)***	***3500 (33.31%)***	**10506 (100%)**
TW (Non frail)	*2777 (26,43%)*	*632 (6.02%)*	3409 (32.45%)
TW (Pre-frail)	*3191 (30.37%)*	*2025 (19.27%)*	5216 (49.65%)
TW (Frail)	*1038 (9.89%)*	*843 (8.02%)*	1881 (17.09%)

We evaluated our approach using two different ranks for the PARAFAC model, namely r=7 and r = 60. In order to calculate the PARAFAC model, with missing values we used two approaches: the als_wopt function (21) that exploits the ALS (Alternating least Squares) algorithm and StrProxSGD algorithm (23), a distributed algorithm exploiting the Proximal Algorithm and SGD (stochastic gradient descent). For the case of r=60 the cp_wopt algorithm could not provide results due to the very long running time (>4

days), thus we report only the results of StrProxSGD algorithm. For the case of r = 6 (i.e. 6 "latent variables for the regression model) we did not use PCA, since the number of variables is relatively small. In contrary for the case of r = 60 we used PCA on the feature matrix using as many variables required for explaining 70% of the data variation. Parameter **nrBins** (i.e. the number of bins used for calculating the distributions of each subject) is learned using cross validation.

Table 6: Evaluation results for cp_wopt and StrProxSGD algorithms.

	r=6, bins=3cp_wopt,		**r=6, bins = 16, StrProxSGD**		**r=60, bins = 5, StrProxSGD**	
	Training	***Test***	***Training***	***Test***	***Training***	***Test***
Median Accuracy	51.33%	50%	65.08%	61.81%	**87.83%**	**72.73%**
Standard deviation	0.0273	0.1668	0.0426	0.1223	0.0289	0.1726

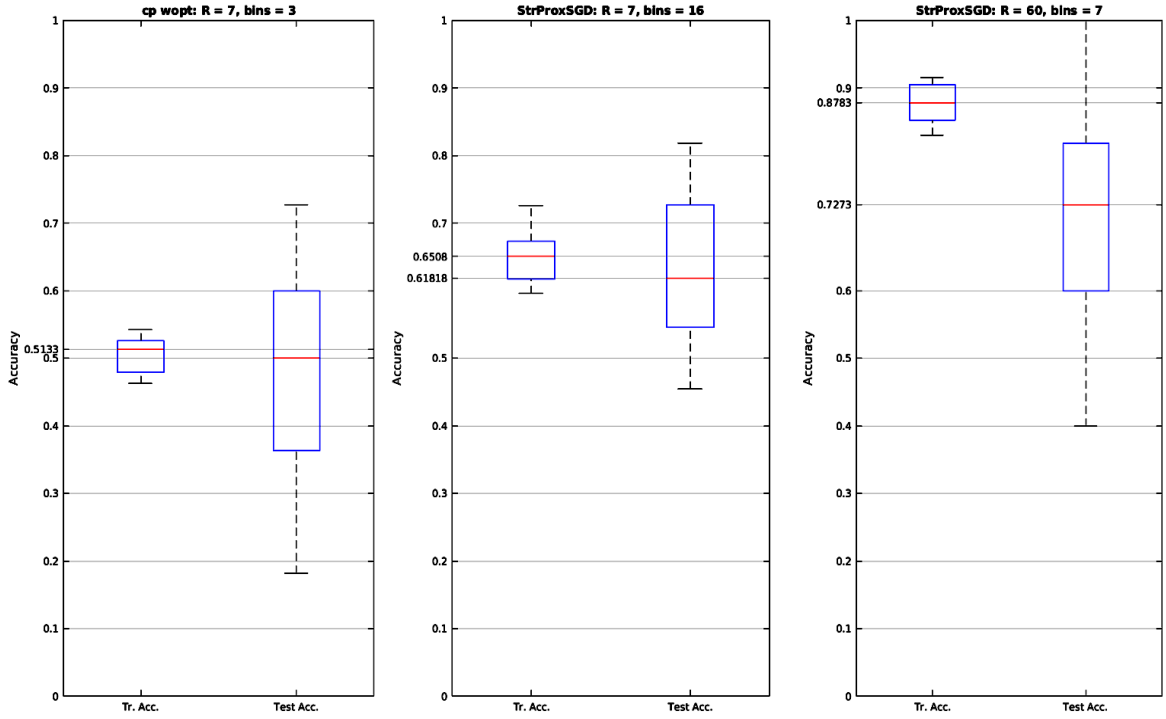

Figure 11: Training and test median accuracy for 10-fold cross validation. From right to left (cp_wopt algorithm with rank=7, StrProxSGD algorithm with rank=7 and StrProxSGD algorithm with rank=60)

Visualization using Spanning Trees

We can visualize the similarity of the subjects based on the extracted features in phase 2 of the pipeline using the spanning trees algorithm. The subject-wise features, that are used to train the LDA-classifier, correspond to the predicted frailty index by the first phase of the method. Using the spanning trees algorithm and the minkowski

distance we can visualize the proximity of the subjects. This algorithm represents subjects by dots and connects similar subjects by edges. Accordingly, the distance between subjects can be perceived as the number of edges that connect the dots. In this visualization we have colored the subjects based on their true class labels: non-frail (blue), pre-frail (orange) and frail (red) and observed their connectivity patterns.

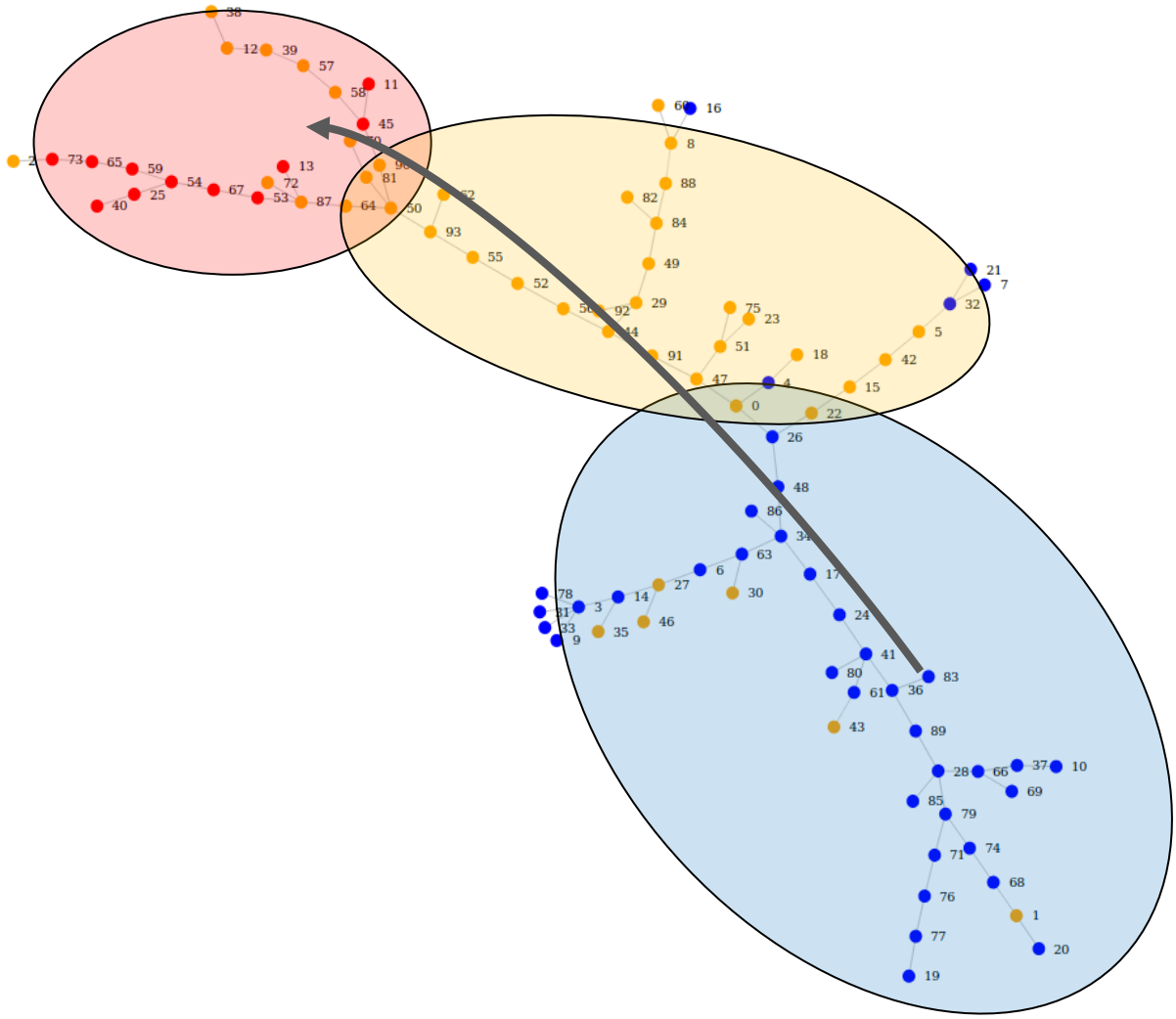

Figure 12: Visualization of training subjects using spanning trees. Blue dots correspond to non-frail, orange dots to pre-frail and red dots to frail individuals.

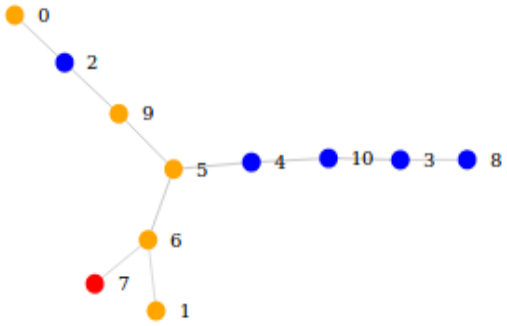

Figure 13: Visualization of test subjects using spanning trees. Blue dots correspond to non-frail, orange dots to pre-frail and red dots to frail individuals.

As can be seen in Figure 12 and Figure 13, there are distinct clusters of non-frail (blue dots), pre-frail (orange dots) and frail individuals (red dots). Furthermore, we can observe in both figures that there exists an arrangement that has a direction from non-frail through pre-frail to frail individuals.

Our plans are now to follow-up the individuals that have been estimated to lie in the transitional zone between frailty groups or are surrounded by neighbors of a different group. In other words, we want to verify whether our method has the potential to find early indicators of frailty transition. For example, the subject 2 in Figure 13, that has been marked as non-frail according to Fried's criteria, shows patterns (based on the FS devices) that are more similar to the pre-frail group.

3.2 Mining of multi-level association rules

Towards discovering associations between frailty, and physiological or behavioral patterns, a preliminary work has been made in association rules. The aim is to discover a novel way of mining multi-level association rules, in a distributed environment, from multiple heterogeneous data sources.

Nowadays, as a result of cheap storage and data availability, the volume of data is huge and is expanding rapidly and very often there is the need to discover useful and interesting knowledge from quite different data sources. While methodologies and solutions exist to mine rules effectively in a single node environment, these methodologies fail when data volume expands beyond a threshold. On the other hand, distributed systems and platforms have appeared to present an alternative processing model, capable of handling effectively massive loads of data. Despite their immense capabilities, these systems lack established methodologies in order to fully exploit their resources.

Our goal is to take the positive features from both models and combine them into a unified model, capable of handling massive data volume and performing established knowledge discovery methodologies (association rule mining) on them. Our focus is twofold:

- Combination of association rules and concept hierarchies to be able to mine multi-level rules from unified and augmented data sources
- Exploitation of the processing power and capabilities of distributed systems to effectively handle the increased data volume

To accomplish our objective, we augment input data, based on concept hierarchies (which are adjusted on the problem at hand), to produce a unified and augmented data file. This file is sent to a distributed processing system (Hadoop framework stack) to generate large frequent itemsets effectively, based on the procedure proposed by the Apriori algorithm. Multi-level association rules are then produced from these itemsets and pruned based on optimization parameters, in order to keep only those that are interesting.

3.2.1 Architectural Model

The model that is being developed is based on a multi-tier architecture. The different tiers are presented in Figure 14, where Configuration tier is marked in blue, Processing tier is marked in green, Output tier is marked in red and finally Control tier is marked in yellow. The packages out of which the tiers are consisted, are explained below.

1. **Configuration and input**: This package is responsible for every configuration that is needed at the system setup phase. System variables, hierarchy files, input and output folders, SSH connections as well as checking access rights for files and folders, where needed. Additionally, tests on the structure and format of input files (data files, hierarchy files, configuration files) are carried out, in order to ensure a smooth system initialization. By design, direct user input for system variables and hierarchies will be available, along with automated importing from xml files.
2. **Data augmentation:** Data augmentation is used to implement the idea of multi-level concept hierarchies. This procedure is performed on all input files, based on the hierarchy files provided at the configuration stage, in order to unify all data entries into a single augmented file. Information for the hierarchy structure is stored along with the actual data (that are considered to be at leaf level), following a bottom up approach.
3. **Distributed processing:** The distributed system that will be used is the Hadoop framework and its software stack. Hadoop provides many tools for distributed processing (Mahout, MapReduce, Spark), but they require coordination during the intermediate phases and the many and time-consuming iterations of the Apriori large frequent itemset generation procedure. Moreover, the actual communication, as well as input and output, with the Hadoop environment has specific issues that must be considered, especially in the case of remote SSH-based communication.
4. **Rule generation and pruning:** Large frequent itemset generation is the costliest part of the Apriori algorithm, after which the rule production follows. Rules are based on the generated itemsets and are associated with several metrics and thresholds, in order to estimate their value. The degree of interestingness determines whether a rule will be pruned or not. The various thresholds are determined during the configuration phase and metric values are exported along with the rules.
5. **Output:** All the interesting rules that have remained, along with their metrics, are stored into a report (in xml format) and the report is exported as system output.
6. **Control:** The coordination between the various stages and phases of our model is done by this package. Due to the differences in input and output data structure and of course features, there must be interfaces between the modules, in order to ensure their efficient operation.

The flow diagram of the overall system is summarized in Figure 15.

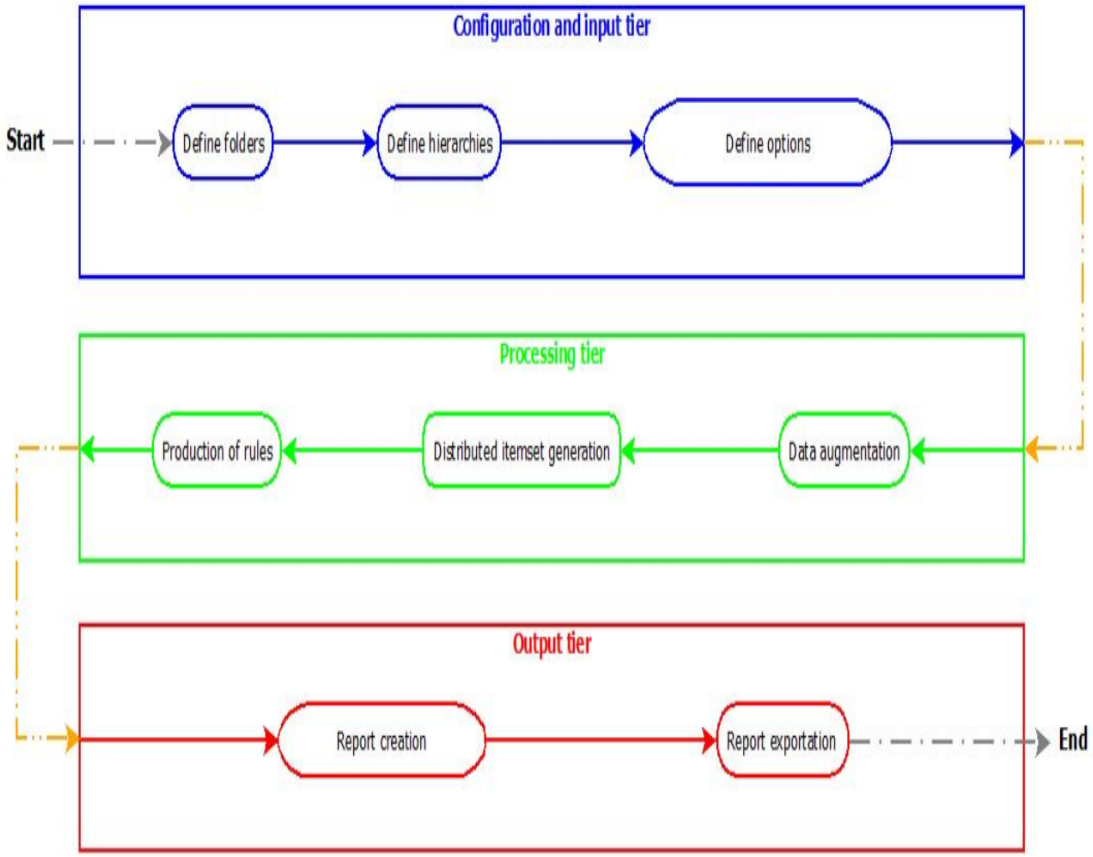

Figure 14: Multi-tier model.

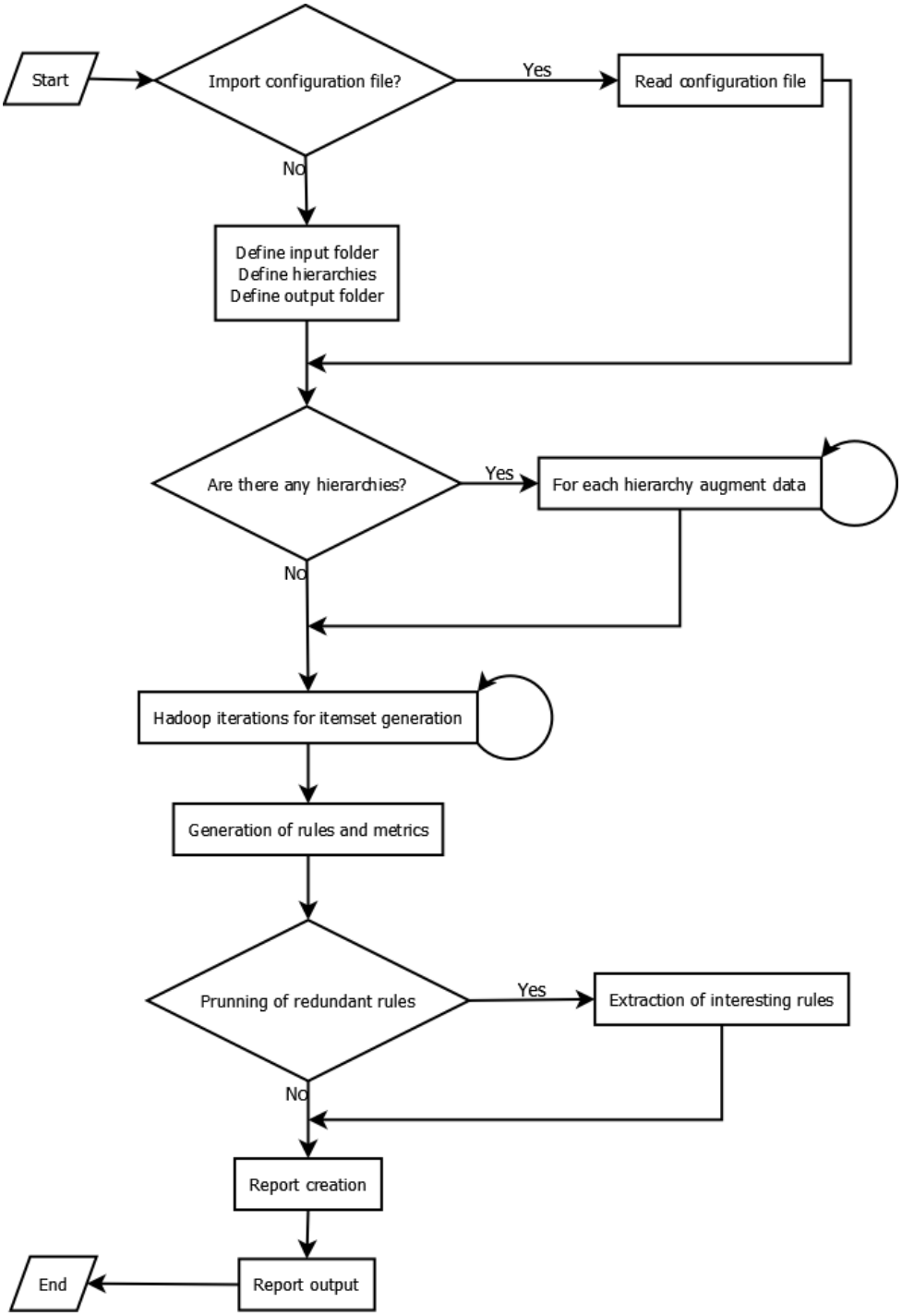

Figure 15: System control flow diagram.

3.2.2 Current and future work

The state of the art in multi-level association rule methods has been explored and several recent works ([22], [23], [24]) has been studied for the creation and the organization of a multi-level concept hierarchy, along with the data augmentation procedure. Association mining algorithms such as the Apriori in distributed architectures has been implemented in [25], [26], [27]. Motivated by the related work,

the basic architecture and packet layout have been designed and the source codes for a number of functions (retrieving hierarchy data and creating the corresponding xml files, reading hierarchy data from xml files, creation and reading of the xml configuration file) have been developed in Java. Furthermore, existing algorithms in Mahout, from the Hadoop framework stack, were examined in order to generate the frequent itemsets. As a secondary choice, an Apriori implementation based on the main map-reduce Hadoop framework stack could be used. We started to investigate the proposed framework using FrailSafe's multimodal data.

Specifically, we focused on developing, implementing and testing our model. The steps we followed until now and will be completed in the next 2 months are summarized next:

- Comprehensive study and analysis of available data and their structure, to determine the precise augmentation process
- Study and testing of the distributed system features and their interaction with the rest of the system
- Completion of the distinct software packages, regarding their basic functionality
- Implementation of the basic software functionality (for the entire system), in order to run tests and configuring the system for maximum performance (few hierarchies vs many hierarchies, distributed vs non-distributed implementation, handling multiple data sources)
- Creation of complete test cases (after the implementation of the basic system functionality), that cover everything from system input (data sources) to system output (exported report)
- Expansion of the basic software features, based on the designed multi-tier model, to their full extent
- System testing, using actual data as input and evaluation of its usability, performance and efficiency
- External user reviews (lab members that are unfamiliar with the system) for feedback on the various features
- Final system evaluation, using the complete data set and comparing the various configurations and their results.

Application to FrailSafe data

We applied the Apriori algorithm to extract association rules using as variables the statistical metrics from the VPM game parameters. The Apriori algorithm is an influential algorithm for mining frequent itemsets for boolean association rules. The Apriori algorithm is based on the following key concepts:

- *Frequent itemsets*: the sets of item which has minimum support (denoted by L_i for i^{th}-itemset)
- *Apriori Property*: Any subset of frequent itemset must be frequent
- *Join Operation*: To find a set of candidate k-itemsets is generated by joining L_{k-1} with itself.

The main parameters for the Apriori algorithm are the minimum support threshold and the minimum confidence threshold. The *support* is the fraction of transactions that contain both A and B:

$$Support(A,B) = P(A,B)$$

whereas the *confidence* is the fraction of transactions where items in B appear in transactions that contain A:

$$Confidence(A,B) = P(B|A)$$

We have used a value of 0.2 for the minimum support threshold and 0.8 for the minimum confidence threshold.

The investigated parameters, if numeric, were converted to ordinal by quantizing them to three levels: small, medium and large. The corresponding variable names were annotated with the strings 'q1', 'q2' or 'q3', for the 3 levels respectively. We used the one-hot encoding to represent ordinal variables as binary vectors. In this case each parameter value is represented as a binary vector that has all zero values except the index of the quantile it belongs to, which is marked with a 1.
Additionally to the parameters from the FrailSafe devices we added the frailty classification by Fried (from the clinical association) in order to find any associations between the FS variables and the fried classification score. Following the variable annotations in section 4.2.1 , we marked the low, intermediate and increased frailty as *fried_status-q1* (=non-frail), *fried_status-q2* (=pre-frail), and *fried_status-q3* (=frail).

1. Example with variables from the VPM game parameters

We first searched for associations in the parameters from the games. The variable names are shown in Table 8 in section 4.2.1 For example, the item *Distance_mean-q3* represents *large mean distances*.
The Apriori algorithm found 28 frequent itemsets with cardinality = 1, 71 itemsets with cardinality = 2, 46 itemsets with cardinality = 3, and 2 itemsets with cardinality = 4. As example the itemsets with the largest cardinality were the following:

1) Distance_mean-q3 , Heigth_std-q1, Force_mean-q2, Lives_mode-q1
2) avg_max_redwings_force-q1, Force_mode-q1, Lives_mode-q1, fried_status-q2

The identified association rules were 85. An example is displayed:

- ✓ Force_mode-q1, fried_status-q2 → Lives_mode-q1 (31.0714%, 100%)

This example means that when the most frequent (mode) force value was low in the case of pre-frail people, the most frequent lives (in the game) were also low. The percentages next to each rule are the values for Support and Confidence. We visualize the 85 rules in a more compact form using 2 graphs: one directed that shows the in-out connections, and one undirected that illustrates the within group connections. The previous rule for example is visualized with 2 edges in the directed graph (Force_mode-q1→ Lives_mode-q1 and fried_status-q2 → Lives_mode-q1) and one edge in the undirected graph (connecting Force_mode-q1 with fried_status-q2). These 2 graphs are shown in Figure 16 and Figure 17, respectively.

Figure 16: Directed subgraph illustrating the in-out associations of variables from the games. Edges are added to all items in each group.

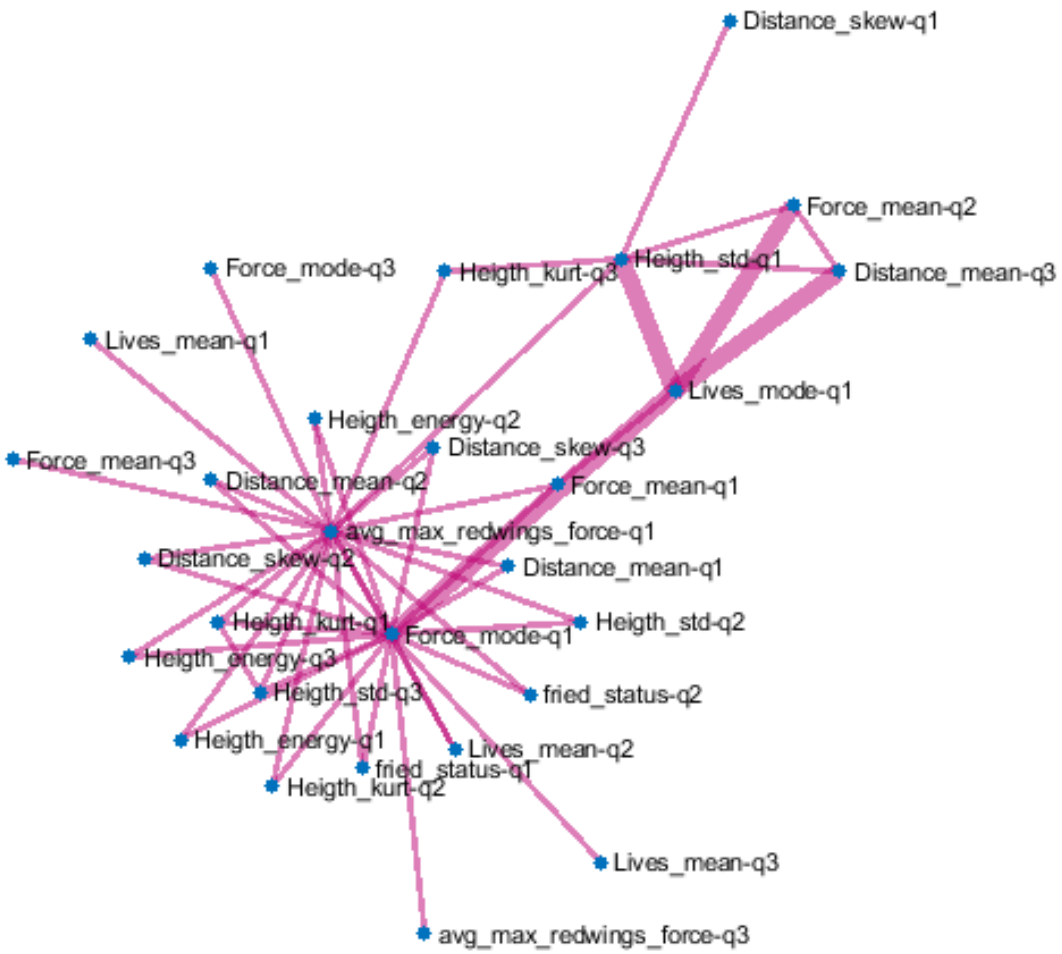

Figure 17: Subgraph with the game variables illustrating the within group connections, i.e. items that appear together. All itemsets are illustrated in the same subgraph. Thinner lines connect input items and thicker lines output items.

From the directed subgraph it is easy to see that many variables are associated to a small number of Lives and that both pre-frail and frail link to it. On the opposite the frail status seems not to be associated with any variable which might indicate that frail people did not play games. Key parameters in the undirected graph seem to be the low values of avg_max_redwings_force and force_mode.

2. Example with variables from the WWBS

Similarly as above, we applied the Apriori algorithm to extract association rules between the statistical metrics of the WWBS and the fried status. The variable names are shown in Table 7 in section 4.2.1 . The investigated parameters, if numeric, were quantized to three levels: small, medium and large, and the corresponding variable names were annotated with the strings ‘q1’, ‘q2’ or ‘q3’, for the 3 levels respectively. The calculated rules are illustrated in the form of a directed and an undirected graph (Figure 18 and Figure 19 respectively), as described previously.

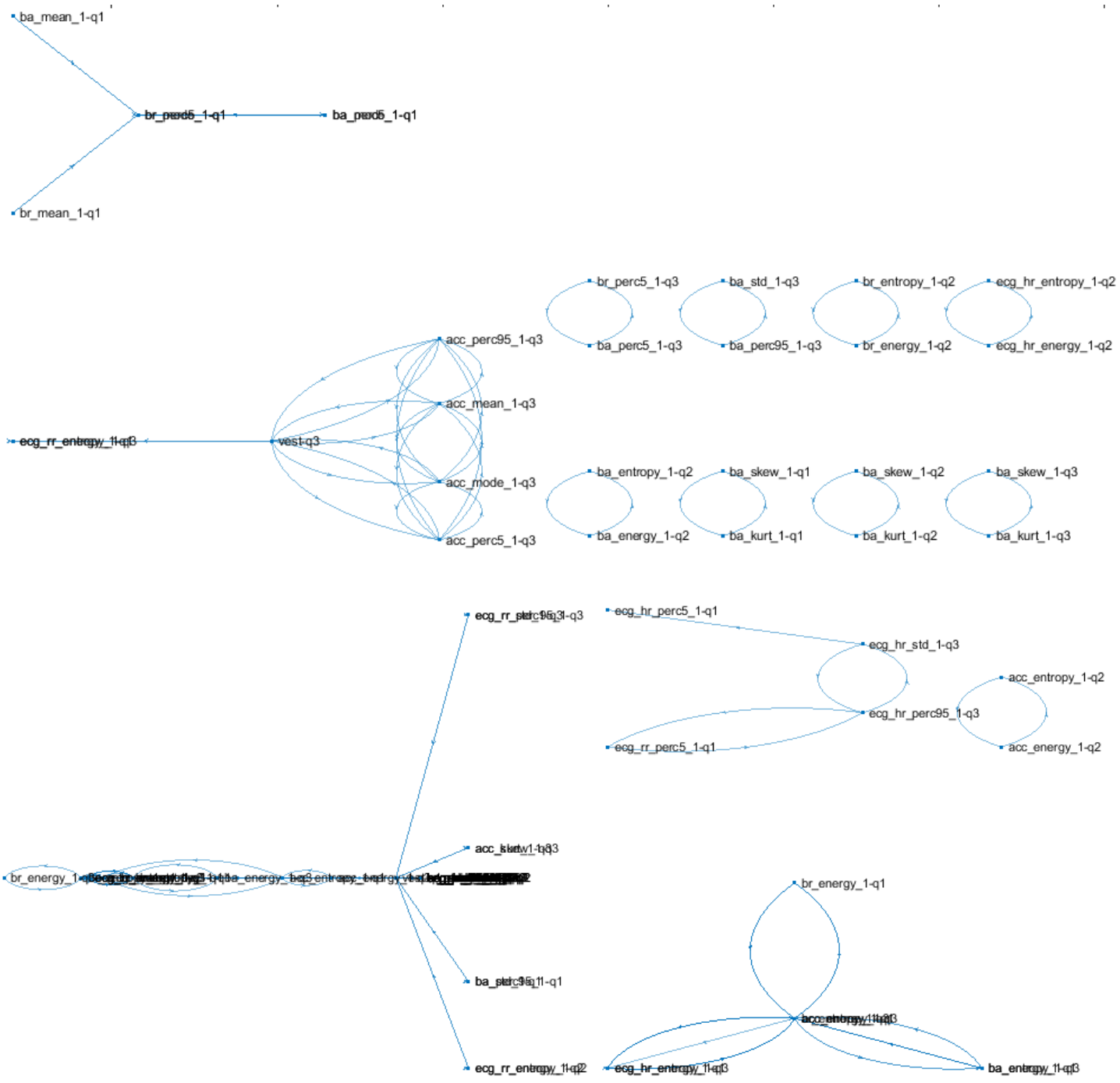

Figure 18: Directed subgraph illustrating the in-out associations of variables from the WWBS. Edges are added to all items in each group.

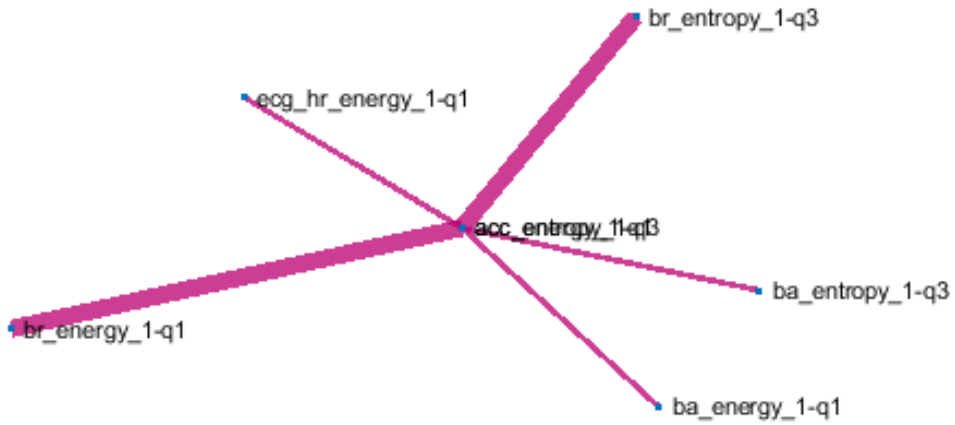

Figure 19: Subgraph with the WWBS variables illustrating the within group connections, i.e. items that appear together. All itemsets are illustrated in the same subgraph. Thinner lines connect input items and thicker lines output items.

The algorithm found 135 frequent itemsets with cardinality = 1, 449 itemsets with cardinality = 2, and 6 itemsets with cardinality = 3. As example the 6 itemsets with the largest cardinality included low energy values in acceleration, high entropy values in acceleration and one of the following: ba_energy_1-q1, ba_entropy_1-q3, br_energy_1-q1, br_entropy_1-q3, ecg_hr_energy_1-q1, ecg_hr_entropy_1-q3. This rule indicates that breathing is associated with body acceleration and heart rate.

3. Example with variables from the GPS

Furthermore, we identified association rules from GPS variables. Similarly to above the variables were quantized and the one-hot encoding was used to represent the ordinal variables as binary vectors. The variable names are shown in Table 9 in section 4.2.1 . The algorithm found 47 frequent items with cardinality = 1, 136 itemsets with cardinality = 2, and 30 itemsets with cardinality = 3. The association rules are visualized as a directed and an undirected subgraph, as shown in the next figures.

Figure 20: Directed subgraph illustrating the in-out associations of variables from the GPS. Edges are added to all items in each group.

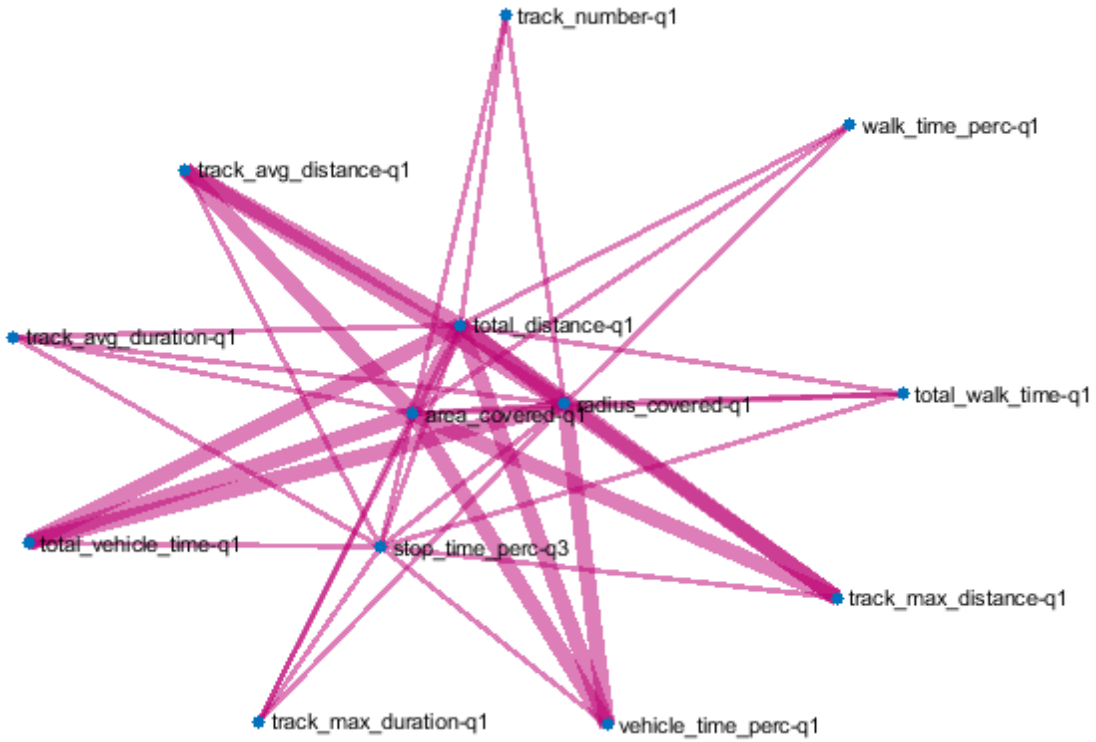

Figure 21: Subgraph with the GPS variables illustrating the within group connections, i.e. items that appear together. All itemsets are illustrated in the same subgraph. Thinner lines connect input items and thicker lines output items.

4 Data Fusion

There are two main approaches for fusing data from different sensor units/dimensions: feature-level fusion and decision-level fusion. In feature-level fusion, which is commonly used to exploit the dependencies across dimensions, the data are fused directly after feature extraction. Feature vectors from each dimension/sensor unit are fused and events are classified by one global classifier. On the other hand, in decision-level fusion, events are classified for each dimension/sensor unit by its local classifier and the results from these local classifiers are later fused in the decision layer.

Analysis of multi-sensor data is very complex and difficult to summarize with a small number of variables extracted from the multi-dimensional signals. As a result, analysis is usually accompanied by extraction of high dimensional feature vectors from data. The dimensionality is further increased in feature-level fusion approaches aiming to exploit the information across dimensions/sensor units, where already high dimensional feature vectors from several sensor units are combined to a single large feature vector. The problem of high dimensionality coupled with limited number of samples, usually available in practice, makes the analysis of multidimensional signals a challenging task.

Thus, we propose a new decision-level scheme to deal with the problem of high dimensionality in conjunction with small number of samples. The proposed scheme combines information from all sensor units in order to train a single classification model and thus is sensor-independent. The decision-level fusion scheme keeps the dimensionality quite low, while the incorporation of a global training model allows the use of more training samples (by combining all sensor units).

4.1 Data Fusion schemes

4.1.1 Feature Level fusion

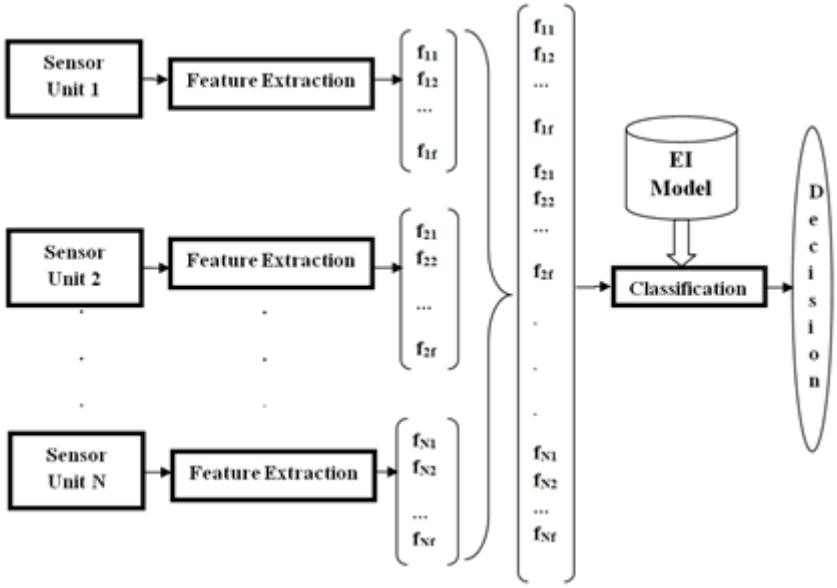

Figure 22: Feature-level fusion scheme.

In the feature-level scheme, each one of the available sensor units from each frame is processed in parallel by the feature extraction algorithm. The estimated feature vectors from each sensor unit are concatenated into a single feature vector. This 'super' feature vector is used as a representative signature for the corresponding frame. Therefore, the training set is a data matrix *M×N×f*, where *M* is the number of frames in the training set, *N* is the number of sensor units and *f* the number of features extracted from each sensor unit. The feature level scheme is illustrated in Figure 22. Although such a scheme exploits the information from all dimensions of the data, it leads to a feature vector of high dimensionality imposing the need either for feature selection before classification, or the availability of a large number of training samples.

4.1.2 Decision-level fusion with local training models

In the decision-level fusion with local (sensor dependent) training models, a separate classification model is built for each sensor unit. Each one of the available sensor units is processed in parallel by the feature extraction algorithm and the estimated feature vectors are used to form *N* training sets, one for each sensor unit. The data matrix of each training set is *M×f* here. For each frame, *N* decisions are made by each one of the *N* local classifiers. A final decision is made by combining the *N* output class labels using a fusion rule, such as majority voting. The decision-level with local training models fusion scheme is illustrated in Figure 23. In decision-level fusion schemes the dimensionality of the feature vector is smaller than in feature-level fusion schemes. However, this scheme uses training samples only of the corresponding sensor unit.

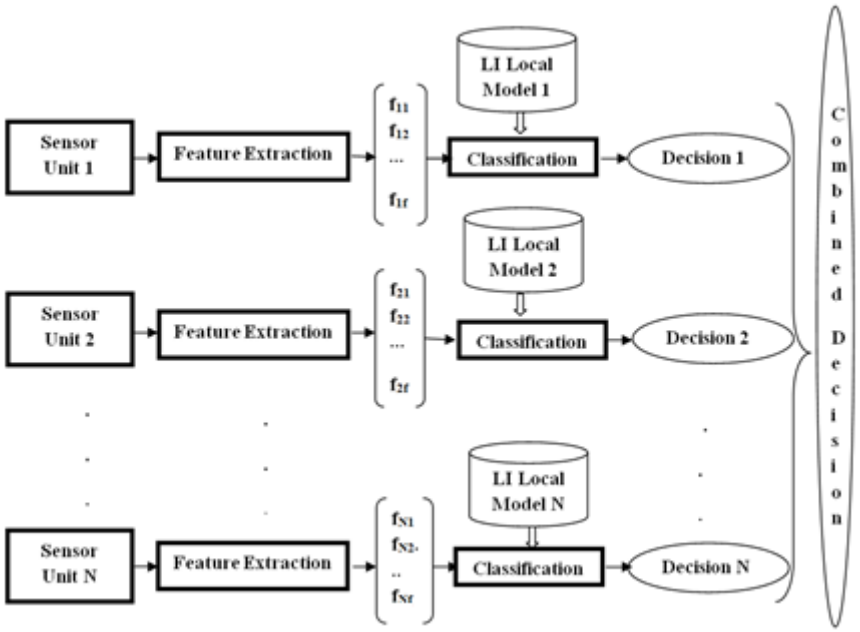

Figure 23: Decision-level with local (sensor dependent) training models fusion scheme.

4.1.3 Decision-level with global training model

In the decision-level with global (sensor independent) training model fusion scheme, a common classification model is used for the feature vectors extracted from the different sensor units. The data matrix of the training set is now *N×M×f* and is constructed by merging all training sets from the decision-level with local models fusion scheme. In this scheme the number of training samples is larger since each data frame appears in the training set *N* times, one time for each one of the available sensor units. During the test phase, for each frame, *N* decisions are made by feeding the signature from each sensor unit to the global classification model. A final decision is made at a score level by combining the *N* output class labels using the same fusion rule (majority voting) as before. The decision-level with global training model scheme is illustrated in Figure 24. Although this scheme is less specific, it handles better both the high dimensionality and the problem of small number of training instances.

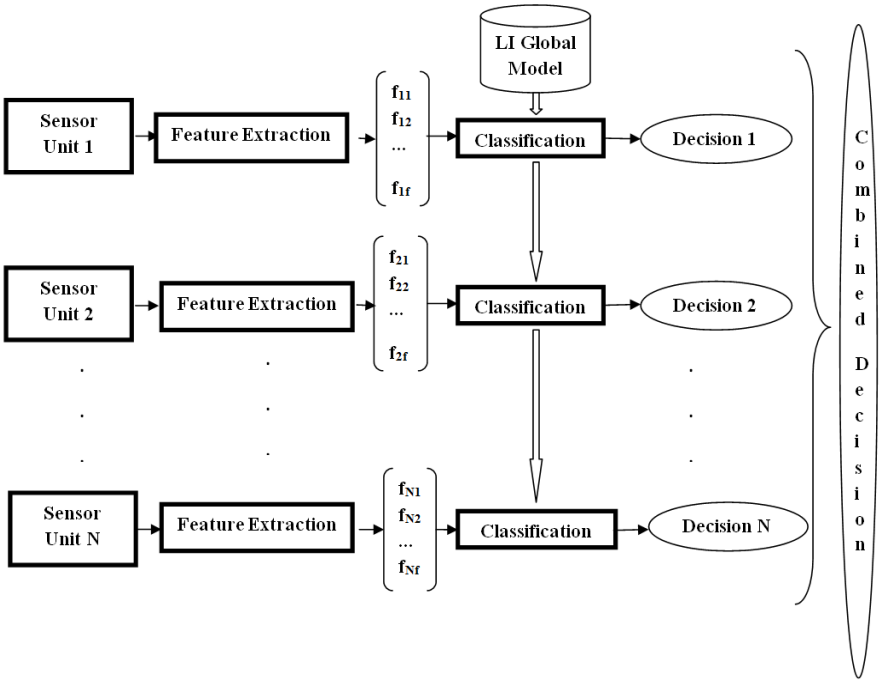

Figure 24: Decision-level with global (sensor independent) training model fusion scheme.

Our preliminary work in data fusion schemes were tested on different datasets and showed promising results (36). In the next section we validate our schemes using data collected through FrailSafe study.

4.2 Extracted parameters from the FrailSafe device recordings

4.2.1 Parameter description

In the FrailSafe project a variety of sensors, devices, and developed software is used to capture several aspects of the participants' health status (physiological, behavioral, cognitive etc.). Using the recordings of these devices we can extract features which

are fed to state-of-the-art data analysis algorithms. In this section we will present the features which are extracted from the game logs, and the outdoor monitoring application, and the sensorized strap/vest that the participants are wearing.

The strap/vest measured entities were used to extract features by calculating histogram-based features for each of the measured parameters as shown in Table 7. The extracted features were introduced for the first time in D4.2 but are also mentioned here for completeness. 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. After performing activity classification for each session, the duration of the session was split in a set of activities (e.g. a participant could be sitting but also walking during different time points of the same session). To that end, the aforementioned features were calculated for each activity of each session, resulting in an augmented feature vector containing features that correspond to all the activities performed during the session.

Table 7: Measurements from vest/strap and extracted histogram-based variables.

Parameter	Statistical metrics
Heart Rate	*average, standard deviation (std), 5% percentile, 95% percentile, most frequent value (mode), kurtosis, skewness, energy, entropy*
Respiration Rate	
Heart Rate Variability	
Breathing Rate	
Breathing Amplitude	
Acceleration	

Using the game recordings, some additional features were extracted which came as a result of calculation of histogram features from the raw values of the recorded game entities. These features were combined additionally with the summarized games data which are stored in the Virtual Patient Model (VPM) of each participant. The list of the extracted features from games are shown in Table 8.

Table 8: VPM game parameters and extracted histogram-based variables.

VPM Game parameters
Max force
Average max force
Average and max endurance
Average and max score
Average and max game duration

Game Parameter	Statistical metrics
Height	*average, standard deviation (std), 5% percentile, 95% percentile, most frequent value (mode), skewness, kurtosis, energy, entropy*
Distance	
Speed	
Lives	
Force	

Finally, a number of features was extracted from the GPS data collected through the outdoor monitoring application (GPS logger). These features which are listed in Table 9 capture the participants' outdoor mobility behaviour. This is the first time we extract multiple parameters from the GPS data and analyze them individually, as well as jointly with the rest.

Table 9: GPS logger extracted parameters.

Features
total_distance
total_duration
total_steps
radius_covered
area_covered
average_walk_speed
total_walk_time
total_stop_time
total_vehicle_time
walk_time_perc
vehicle_time_perc
stop_time_perc
track_number
track_avg_distance
track_avg_duration
track_max_distance

track_max_duration

4.2.2 Handling missing values and time synchronization with clinical evaluation data

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 map measurements for the same participant collected on different dates. This mapping was performed by filling missing values with the ones of the nearest date.

Additionally, during the FrailSafe study the device and sensor recordings are collected regularly on the home visit sessions, whereas the participants' medical data are collected sparser on the clinical evaluations. In an effort to fuse the clinical data with the device recordings, it was necessary to perform an estimation on the clinical data values in intermediate time points for which there were device recordings. Hence, the empty cells in the proxy outcomes vectors were filled by performing linear interpolation. This allowed us to have synchronized measurements for both input and output variables in our algorithms.

4.3 Exploitation of devices-generated features for classification of variables from medical domains

In the context of applying the aforementioned methodology to the FrailSafe data, we chose to use features extracted from different devices, in order to be used for classification of variables that correspond to the medical domains. Thus, we performed data fusion in two levels: a feature-level fusion and a decision-level fusion. The two approaches are described in the next section.

4.3.1 Early integration

As an early integration step, we fused features extracted from different sensors, to build prediction models of the clinical variables aiming to be able to characterize the profile of the older person in the different domains (physical, psychological, social etc) without the need to perform an extended clinical evaluation. This will allow the seamless and continuous monitoring of the participants providing the means to detect early changes in the health condition. More specifically, we built three different classification models, one for each of the following data-generator devices: WWSX, GPS, games. For each device's recordings, features were extracted and used as input to train a *K nearest neighbors* classifier. The dataset provided consisted of multiple sessions of a set of FrailSafe participants. Thus, in order to build the classification model, the dataset was split into training and test set according to the participants' IDs, meaning that the training and test sets contained unique participants. The target of the classifier was each of the proxy outcomes. It should be noted here that the

clinical variables that contained numerical values, were converted to nominal, in order to be used as classification targets. Data were preprocessed as to exclude columns with more than 1% missing values. Additionally, KNN imputation was performed for columns that contained less than 1% missing values. The models were validated by performing 10-fold stratified cross validation, to address the problem of imbalanced classes. The mean cross-validation classification accuracy per session on the test set is depicted in Figure 25, Figure 26, and Figure 27 for the WWSX, GPS and games features respectively.

A great amount of eCRF variables was successfully predicted by the classifiers constructed, with high classification accuracy in each of the three models. The targets that were predicted with low classification accuracy, were mostly the ones with three or more labels (e.g. fried_status with three labels).

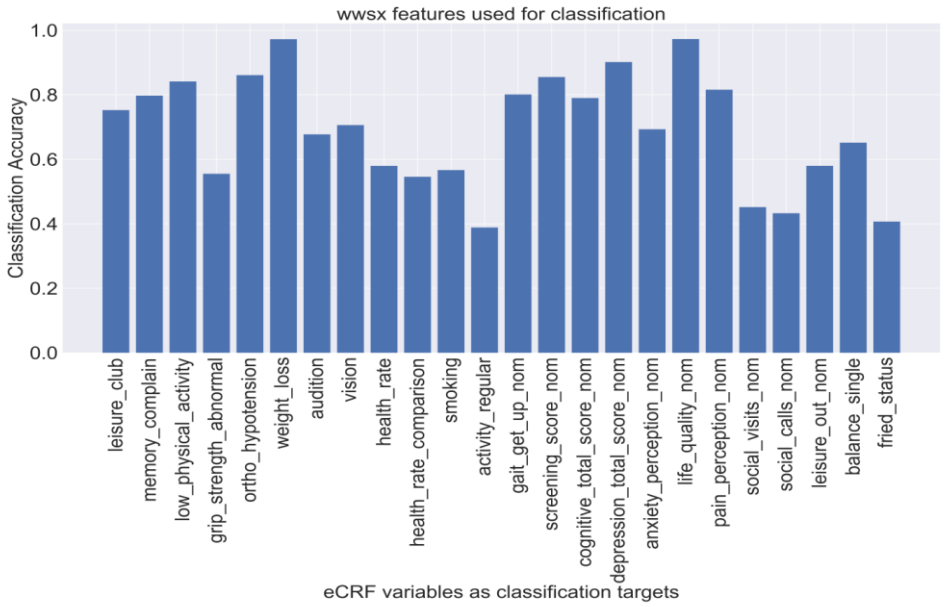

Figure 25: Accuracy of classification on clinical parameters using wwsx features.

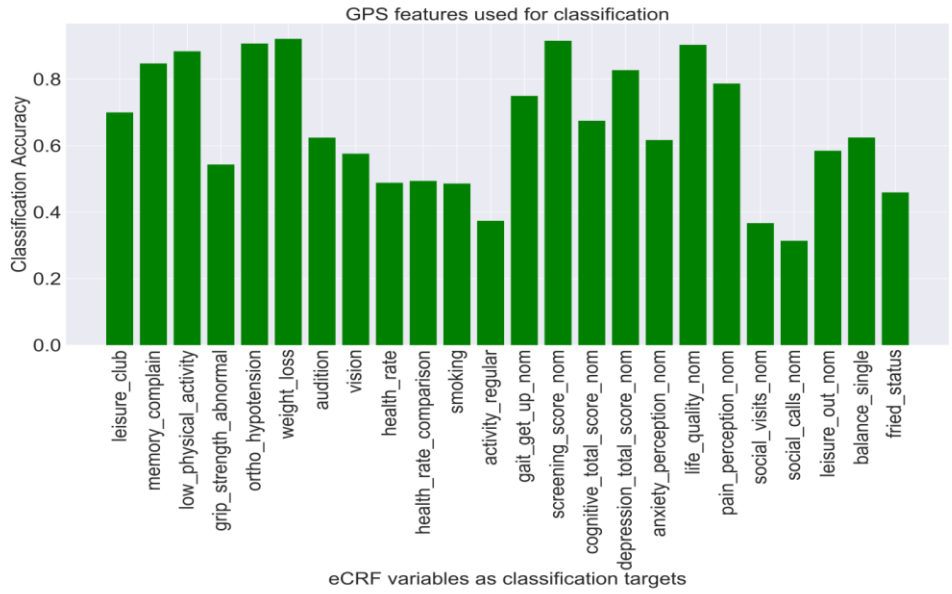

Figure 26: Accuracy of classification on clinical parameters using GPS features.

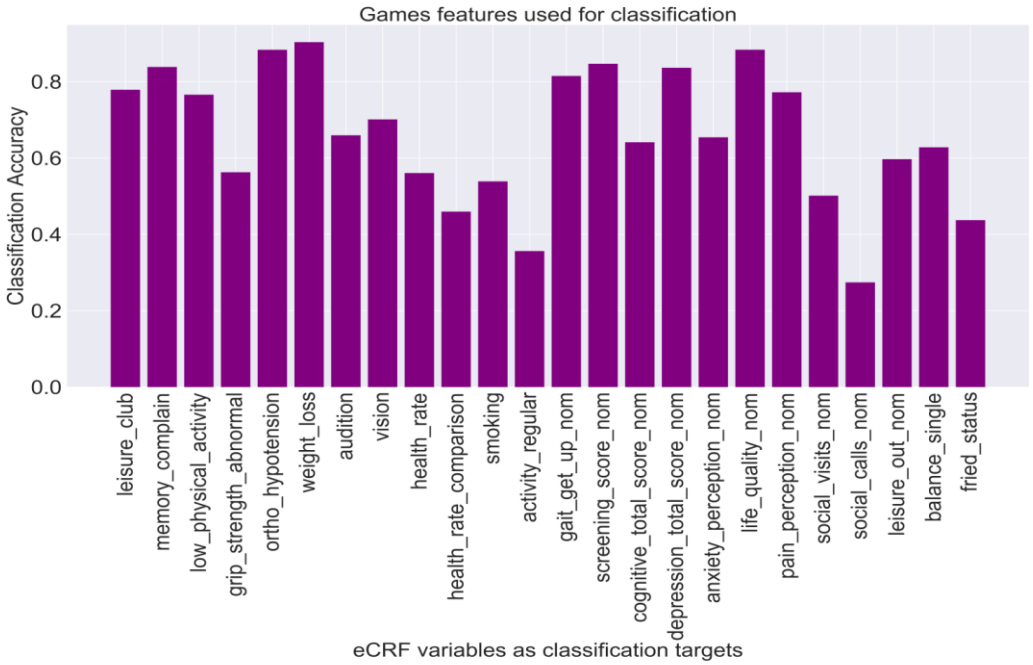

Figure 27: Accuracy of classification on clinical parameters using games' features.

4.3.2 Late integration

At a late integration step, we constructed a decision-fusion classification model, in order to combine the aforementioned separate classifiers and extract meaningful information for the same set of targets (eCRF variables). More specifically, we created a dataset that contained all sessions of each participant, for all three devices which

shared the same set of participants. The sessions of a device that did not contain any values were filled with the corresponding values of the nearest timepoint for each feature. This resulted in a merged dataset with multiple sessions for each participant, and a feature vector extracted from recordings of all three devices. To perform this feature-level fusion, the dataset was split in 70% for training and 30% for testing, according to the participants' IDs. In the training phase, the three classifiers, that were previously constructed, were used to predict the target (one target each time). The three responses were taken as features to train a new classifier, a decision tree, which would predict the target label based on the responses of the three classifiers. The same procedure was followed for the test set, and the classification accuracy was calculated for the separate classifiers, as well as the decision-fusion classifier. The classification results per session are reported in Table 10 and the corresponding chart in Figure 28.

Table 10: Classification accuracy of separate classifiers and decision-fusion classifier.

	WWBS model	GPS model	Games model	Decision-fusion model
leisure_club	95%	89%	93%	95%
memory_complain	92%	92%	92%	92%
low_physical_activity	89%	89%	93%	93%
grip_strength_abnormal	69%	62%	86%	86%
ortho_hypotension	95%	95%	95%	95%
weight_loss	99%	99%	99%	99%
audition	84%	77%	66%	85%
vision	72%	87%	83%	75%
health_rate	71%	73%	74%	72%
health_rate_comparison	56%	62%	67%	67%
smoking	69%	70%	72%	78%
activity_regular	45%	42%	52%	52%
gait_get_up_nom	92%	92%	92%	92%
screening_score_nom	95%	95%	95%	95%
cognitive_total_score_nom	80%	79%	85%	85%
depression_total_score_nom	99%	99%	99%	99%
anxiety_perception_nom	80%	83%	87%	80%
life_quality_nom	100%	100%	100%	100%
pain_perception_nom	81%	81%	81%	81%
social_visits_nom	62%	91%	56%	89%
social_calls_nom	45%	97%	51%	96%
leisure_out_nom	67%	64%	63%	59%
balance_single	83%	69%	70%	74%
fried_status	69%	70%	89%	87%

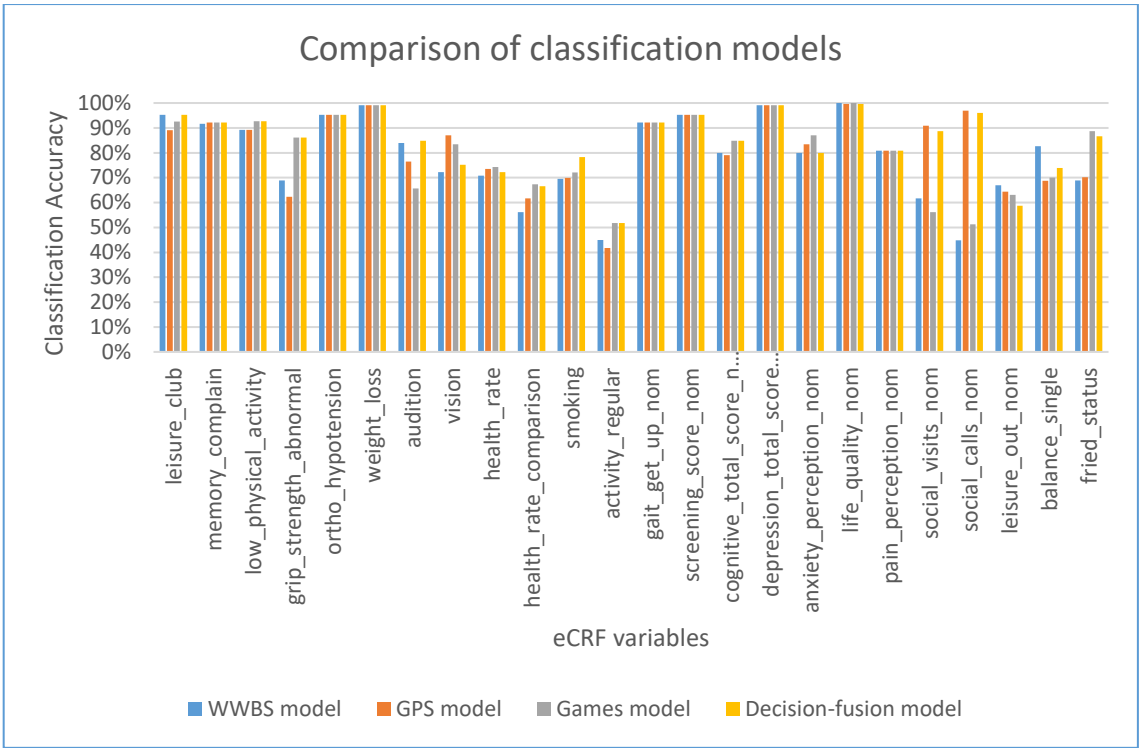

Figure 28: Comparison of the different classification models' accuracy.

It is made clear that in almost all cases, the decision-fusion model performs better or equally well. This happens due to the fact that fusing the decisions of separate classification models can produce rules based on the combination of the answers given from each model. To that end, we conclude that performing data fusion in two levels has helped towards improving the procedure of predicting eCRF variables, using features extracted from the FrailSafe devices.

4.4 Correlation of FrailSafe device recordings with proxy outcomes

The recordings from the vest/strap (ECG and IMUs) as well as from the FrailSafe games and from the GPS logger app were used for statistical analysis. More specifically, the ability of these measurements to predict the change of the clinical metrics, defined as proxy outcomes in D2.1, was examined. As mentioned in D2.1, proxy outcomes are based on the data from repeated clinical evaluations and described by the differences (delta) in clinical parameters that capture the status of separate human functions:

- MMSE (MMSE total score) and MoCa (cognitive total score) - **cognitive function**
- Gait speed (gait speed 4m) - **physical function**
- GDS (depression total score) - **psychological status**
- Weight loss - **general health**
- Health rate – **health status self-assessment**

From the aforementioned parameters, only those that are numerical and are considered to have a continuous evolution in time (MMSE total score, cognitive total score, depression total score, health rate, gait speed 4m) were included in the statistical analysis that was performed. A binary variable, such as weight loss, cannot be well predicted by regression models.

The main goal of the analysis is to investigate whether variables extracted from the FrailSafe devices can be used as predictors of the proxy outcomes. This is done by examining their correlation with each of the proxy outcomes separately and then combining the most correlated ones in a unified predictive model. Hence, four different steps were followed:

1. Examination of the correlation of vest/strap recordings with each of the proxy outcomes. To this end, we used the strap/vest measured entities to extract features for the statistical analysis by calculating histogram-based features for each of the measured parameters as described in section 4.2. The features were calculated for each activity of each session, resulting in an augmented feature vector containing features that correspond to all the activities performed during the session.
2. Examination of the correlation of games recordings with each of the proxy outcomes. For this reason, we used summarized games data from the Virtual Patient Model (VPM) as well as other features which came as a result of calculation of histogram features from the raw values of the recorded game entities.
3. Examination of the correlation of GPS logger recordings with each of the proxy outcomes. To this end, we used the GPS logger measured entities to extract features for the statistical analysis by calculating features for each of the measured parameters as described in section 4.2.

In all steps (1), (2) and (3) the statistical metrics were calculated by aggregating data collected for each participant per session to compute the evolution of the participants' health status regarding the measured parameters in day-by-day basis.

4. Examination of the correlation of combined vest/strap, GPS logger and games recordings with each of the proxy outcomes.

In steps (1), (2) and (3) lasso linear regression was performed to select a subset of variables and estimate their β coefficients, aiming at building predictive models having the best possible correlations with each of the proxy outcomes separately. The analysis was performed five times (one for each of the examined proxy outcomes) and each time a set of different values for λ, a parameter which controls the number of retained coefficients and thus the risk for overfitting, was tested and the one with the smallest fitting error was selected.

In step (1) 100 recordings constituted the input dataset of the statistical model. The reason why this dataset has a small number of recordings is that the activity

classification performed for each session is a time-consuming procedure which has not been completed yet. After the execution of the activity classification algorithm for all the sessions in the near future, we will have a sufficient dataset of at least 250 recordings and thus, we will repeat the analysis tests to get more accurate results. The current dataset resulted from the exclusion of data from: (a) participants who left the study (e.g. due to death or consent withdrawal) and (b) participants for whom there were missing values for the clinical entities defined as proxy outcomes until 15/01/2018.

The analysis in step (1) was performed by:

a) fitting all the data at once. The results in this case were the following:

Proxy	Correlation
gait_speed_4m	*0.853*
cognitive_total_score	*0.757*
mmse_total_score	*0.729*
depression_total_score	*0.696*
health_rate	*0.619*

b) calculating the predictive model using bootstrapping with 5 repetitions. At each repetition, 70% of the records were randomly selected for regression and the resulting model was applied on the remaining 30%. The results in this case were the following:

Proxy	Mean_train_corr	Mean_test_corr	Median_train_corr	Median_test_corr	Std_train_corr	Std_test_corr
gait_speed_4m	*0.754*	*0.541*	*0.760*	*0.545*	*0.036*	*0.079*
cognitive_total_score	0.671	0.390	0.674	0.376	0.027	0.065
mmse_total_score	0.416	0.174	0.423	0.171	0.035	0.092
depression_total_score	0.577	0.265	0.557	0.263	0.037	0.035
Health_rate	0.572	0.171	0.572	0.146	0.026	0.105

In step (2) recordings from the “Red Wings” game were used due to the plethora of the collected data. The total number of recordings for “Red Wings” game until 14/01/2018 was 840. For the calculation of this number, data from participants which have been excluded from FrailSafe study were not taken into account.

The analysis in step (2) was performed by:

a) fitting all the data at once. The results in this case were the following:

Proxy	Correlation
gait_speed_4m	*0.597*
cognitive_total_score	*0.312*
mmse_total_score	*0.226*
depression_total_score	*0.197*
Health_rate	*0.203*

b) calculating the predictive model using bootstrapping with 5 repetitions. At each repetition, 70% of the records were randomly selected for regression and the resulting model was applied on the remaining 30%. The results in this case were the following:

Proxy	Mean_train_corr	Mean_test_corr	Median_train_corr	Median_test_corr	Std_train_corr	Std_test_corr
gait_speed_4m	0.573	0.484	0.585	0.457	0.0316	0.072
cognitive_total_score	0.251	0.018	0.256	-0.038	0.035	0.100
mmse_total_score	0.313	0.047	0.312	0.077	0.010	0.076
depression_total_score	0.305	0.113	0.329	0.085	0.049	0.055
Health_rate	0.545	0.336	0.546	0.295	0.073	0.231

In step (3) 1296 recordings from the GPS logger constituted the input dataset of the statistical model.

The analysis in step (3) was performed by:

a) fitting all the data at once. The results in this case were the following:

Proxy	Correlation
gait_speed_4m	*0.376*
cognitive_total_score	*0.152*
mmse_total_score	*0.159*
depression_total_score	*0.121*
health_rate	*0.146*

b) calculating the predictive model using bootstrapping with 5 repetitions. At each repetition, 70% of the records were randomly selected for regression and the resulting model was applied on the remaining 30%. The results in this case were the following:

Proxy	Mean_train_corr	Mean_test_corr	Median_train_corr	Median_test_corr	Std_train_corr	Std_test_corr
gait_speed_4m	0.368	0.348	0.362	0.390	0.023	0.066
cognitive_total_score	0.153	0.039	0.140	0.059	0.033	0.078
mmse_total_score	0.216	0.019	0.219	0.026	0.025	0.087
depression_total_score	0.201	0.062	0.197	0.059	0.008	0.030
health_rate	0.268	0.052	0.271	0.067	0.018	0.043

After the determination of the β-coefficients of the FrailSafe device variables in steps (1), (2) and (3), the results were used in a correlation analysis of their combination with each of the proxy outcomes in **step (4)**. For this reason, we mapped sparse measurements to continuous scores and performed fusion of the variables as described in section 4.2. Additionally, we selected the most significant variables (by decreased order of magnitude for the coefficients) which had a cumulative percentage of 90% in steps (1), (2) and (3) respectively. By this means, it is ensured that only the most related vest/strap, GPS logger and game features are examined by the final statistical model as predictive entities.

As a result of all the above, a correlation analysis was performed in step 4 with a dataset consisted of combined vest/strap, GPS logger and games variables as input to the predictive model. This dataset was consisted of 372 recordings.

We performed analysis for each of the proxy outcomes by calculating the predictive model using bootstrapping with 5 repetitions. At each repetition, 70% of the records were randomly selected for lasso regression and the resulting model was applied on the remaining 30%, i.e. the 70% were the training subset and the 30% the validation subset. Finally, the correlation of the predicted values with the real ones was calculated for both training and validation test. At the end of all the repetitions, the average and the median values of the correlations were calculated.

Results Evaluation

As a criterion of the quality of the regression results, the Spearman's rank correlation coefficient (R) between the predicted proxy outcome values and the proxy outcome vector with the real values was used. We classified the regression outcomes in three categories based on the Spearman's correlation index: (1) Low correlation (<49%), (2) moderate correlation (50-65%), (3) high correlation (66-100%). To consider a result as

noticeable, the R in the validation test should be included in the moderate or high-correlation categories.

The results in the validation subsets for all of the proxy parameters are included in the low-correlation category for all three datasets. As an example, the results for the "gait_speed_4m" parameter are shown below:

Proxy	Mean_train_corr	Mean_test_corr	Median_train_corr	Median_test_corr	Std_train_corr	Std_test_corr
Gait_speed_4m	0.887	0.240	0.895	0.179	0.058	0.229

Despite the fact that the results of the correlation analysis performed for the combination of features from the FrailSafe devices with proxy outcomes were not very conclusive, we will repeat the experiments after the next clinical evaluation (and the completion of the activity classification procedure) in order to incorporate more data and to have more concise results. The updated results of this analysis will be presented in the deliverable D.17.

References

1. *The integration of frailty into clinical practice: preliminary results from the Gerontopole.* **Subra, J., et al.** s.l. : Springer, 2012, The journal of nutrition, health \& aging, Vol. 16, pp. 714-720.

2. *Frailty and failure to thrive.* **Fried, L. P., Walston, Jeremy and others.** s.l. : McGraw-Hill, New York, 2003, Principles of geriatric medicine and gerontology, Vol. 5, pp. 1487-502.

3. *Frailty in older adults evidence for a phenotype.* **Fried, Linda P., et al.** s.l. : Oxford University Press, 2001, The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, Vol. 56, pp. M146--M157.

4. *A global clinical measure of fitness and frailty in elderly people.* **Rockwood, Kenneth, et al.** s.l. : Can Med Assoc, 2005, Canadian Medical Association Journal, Vol. 173, pp. 489-495.

5. *Comparison of 2 frailty indexes for prediction of falls, disability, fractures, and death in older women.* **Ensrud, Kristine E., et al.** s.l. : American Medical Association, 2008, Archives of internal medicine, Vol. 168, pp. 382-389.

6. *The Tilburg frailty indicator: psychometric properties.* **Gobbens, Robbert J. J., et al.** s.l. : Elsevier, 2010, Journal of the American Medical Directors Association, Vol. 11, pp. 344-355.

7. *Frailty is associated with incident Alzheimer's disease and cognitive decline in the elderly.* **Buchman, Aron S., et al.** s.l. : LWW, 2007, Psychosomatic medicine, Vol. 69, pp. 483-489.

8. *The frailty phenotype and the frailty index: different instruments for different purposes.* **Cesari, Matteo, et al.** s.l. : Br Geriatrics Soc, 2014, Age and ageing, Vol. 43, pp. 10-12.

9. *Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care.* **Fried, Linda P., et al.** s.l. : Oxford University Press, 2004, The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, Vol. 59, pp. M255--M263.

10. *Frailty, allostatic load, and the future of predictive gerontology.* **Kuchel, George A.** s.l. : Wiley Online Library, 2009, Journal of the American Geriatrics Society, Vol. 57, pp. 1704-1706.

11. *Development and validation of an electronic frailty index using routine primary care electronic health record data.* **Clegg, Andrew, et al.** s.l. : Br Geriatrics Soc, 2016, Age and ageing, Vol. 45, pp. 353-360.

12. *Tensor decompositions and applications.* **Kolda, Tamara G. and Bader, Brett W.** s.l. : SIAM, 2009, SIAM review, Vol. 51, pp. 455-500.

13. *Indexing by latent semantic analysis.* **Deerwester, Scott, et al.** s.l. : American Documentation Institute, 1990, Journal of the American society for information science, Vol. 41, p. 391.

14. **HASTAD, J.** Tensor rank is NP-complete. *Journal of Algorithms.* 1990, Vol. 11, pp. 644-654.

15. **Kruskal, J. B.** Rank decomposition and uniqueness for 3-way and N-way arrays,. [book auth.] R. Coppi and S. Bolasvo. *Multiway Data Analysis.* Amsterdam : North, Holland, 1989.

16. *Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multi-modal factor analysis.* **Harshman, Richard A.** s.l. : University of California at Los Angeles Los Angeles, 1970, UCLA Working Papers in Phonetics.

17. *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition.* **Carroll, J. Douglas and Chang, Jih-Jie.** s.l. : Springer, 1970, Psychometrika, Vol. 35, pp. 283-319.

18. **Kruskal, J.B.** Three-way arrays: rank and uniqueness of trilinear decomposition, with application to arithmetix complexity and statistics,. Linear Algebra and its Applications, 1977, Vol. 18.

19. *Modeling and multiway analysis of chatroom tensors.* **Acar, Evrim, et al.** 2005. International Conference on Intelligence and Security Informatics. pp. 256-268.

20. **Bader, Brett W., Berry, Michael W. and Browne, Murray.** Discussion tracking in Enron email using PARAFAC. *Survey of Text Mining II.* s.l. : Springer, 2008, pp. 147-163.

21. **E. Acar, D. Dunlavy, T. Kolda, M. Morup.** Scalble Tensor Factorization for Imcoplete Data. *Chemometrics and Intelligent Laboratory Systems.* Special Issue on Multiway and Multiset Data Analysis, 2011, Vol. 106, 1, pp. 41-56.

22. **Bro, G. Tomasi and R.** PARAFAC and missing values. *Chemometrics and Intelligent Laboratory Ststems.* 2005, Vol. 75, 2, pp. 163-180.

23. *A Distributed Proximal Gradient Descent Method for Tensor Completion.* **Papastergiou T., Megalooikonomou V.** Boston, MA, USA : s.n., 2017. 2107 IEEE International Conference on Big Data (Big Data). pp. 2056-2065.

24. *Component models for three-way data: An alternating least squares algorithm with optimal scaling features.* **Sands, Richard and Young, Forrest W.** s.l. : Springer, 1980, Psychometrika, Vol. 45, pp. 39-67.

25. *Tensor decompositions for feature extraction and classification of high dimensional datasets.* **Phan, Anh Huy and Cichocki, Andrzej.** s.l. : The Institute of Electronics, Information and Communication Engineers, 2010, Nonlinear theory and its applications, IEICE, Vol. 1, pp. 37-68.

26. *From k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors.* **Papalexakis, Evangelos E., Sidiropoulos, Nicholas D. and Bro, Rasmus.** s.l. : IEEE, 2013, IEEE transactions on signal processing, Vol. 61, pp. 493-506.

27. *Understanding data fusion within the framework of coupled matrix and tensor factorizations.* **Acar, Evrim, et al.** s.l. : Elsevier, 2013, Chemometrics and Intelligent Laboratory Systems, Vol. 129, pp. 53-63.

28. *A flexible modeling framework for coupled matrix and tensor factorizations.* **Acar, Evrim, Nilsson, Mathias and Saunders, Michael.** 2014. 2014 22nd European Signal Processing Conference (EUSIPCO). pp. 111-115.

29. *Structure-revealing data fusion.* **Acar, Evrim, et al.** s.l. : BioMed Central Ltd, 2014, BMC bioinformatics, Vol. 15, p. 239.

30. *Discovery of spatial association rules in geographic information databases.* **Koperski, Krzysztof and Han, Jiawei.** 1995. International Symposium on Spatial Databases. pp. 47-66.

31. **Srikant, Ramakrishnan and Agrawal, Rakesh.***Mining generalized association rules.* s.l. : IBM Research Division, 1995.

32. *Rough set based rule induction in decision making using credible classification and preference from medical application perspective.* **Tseng, Tzu-Liang Bill, et al.** s.l. : Elsevier, 2016, Computer methods and programs in biomedicine, Vol. 127, pp. 273-289.

33. *Data mining using clouds: An experimental implementation of apriori over mapreduce.* **Li, Juan, et al.** 2012. Proc. 12th International Conference on Scalable Computing and Communications (ScalCom'13). pp. 1-8.

34. *Apriori-Map/Reduce Algorithm.* **Woo, Jongwook.** 2012. Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA). p. 1.

35. *MapReduce as a programming model for association rules algorithm on Hadoop.* **Yang, Xin Yue, Liu, Zhen and Fu, Yan.** 2010. Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on. pp. 99-102.

36. *Data fusion for paroxysmal events' classification from EEG.* **Pippa, Evangelia, et al.** s.l. : Elsevier, 2017, Journal of Neuroscience Methods, Vol. 275, pp. 55-65.

37. *Frailty in elderly people.* **Clegg, Andrew, et al.** s.l. : Elsevier, 2013, The Lancet, Vol. 381, pp. 752-762.