Geriatric group analysis by clustering non-linearly embedded multi-sensor data

Spyridon Kalogiannis
Department of Computer Engineering and Informatics
University of Patras
Patras, Greece
kalogianni@ceid.upatras.gr

Evangelia I. Zacharaki
Department of Computer Engineering and Informatics
University of Patras
Patras, Greece
ezachar@upatras.gr

Konstantinos Deltouzos
Department of Computer Engineering and Informatics
University of Patras
Patras, Greece
deltouzos@upatras.gr

Marina Kotsani
Department of Geriatrics
CHRU Nancy, University Hospital of Nancy
Nancy, France
m.kotsani@chru-nancy.fr

John Ellul
Department of Neurology
University Hospital of Patras
University of Patras
Patras, Greece
ellul@upatras.gr

Athanase Benetos
Department of Geriatrics
CHRU Nancy, University Hospital of Nancy
Nancy, France
a.benetos@chru-nancy.fr

Vasileios Megalooikonomou
Department of Computer Engineering and Informatics
University of Patras
Patras, Greece
vasilis@ceid.upatras.gr

***Abstract*— The goal of this study is to extract new indicators that are descriptive of aging-associated decline in reserve and function (frailty) and perform this through an unobtrusive monitoring system aiming to augment the standard geriatric assessment. Extraction of such indicators was performed by fusing information from multiple devices, such as inertial measurement units (IMUs), a games platform and an outdoor monitoring system, and thereby creating a multi-parametric profile of the older person. Principal component analysis (PCA) was applied to remove correlations in the extracted features, followed by locally linear embedding (LLE) to embed the data in a lower dimensional space where unsupervised clustering is feasible. Exploration of the identified clusters revealed patterns of frailty manifestation that were in high accordance with geriatric indices from several domains (medical, cognitive, psychological, etc). Our results highlight the potential of the applied data mining methodology for the extraction of novel frailty indicators.**

Keywords—multi-sensor data, monitoring system, clustering, feature extraction, frailty, geriatric assessment

I. Introduction

With the increased interest in monitoring services and decentralized health care systems, there is a high demand in computational methods that summarize multi-sensor data in an automated and reproducible way. Precise quantification of the clinical status of selected population groups, which suffer from a medical or psychological condition, are under rehabilitation or are vulnerable due to normal aging, becomes a crucial aspect for assessment of the people's wellbeing and for evaluating the progression of their health status.

Such a quantification requires the extraction of novel biomarkers from a series of variables that are recorded in order to create a multi-parametric profile of each person. These variables correspond to "objective" measurements of health status and characterize people's behavior, both in an intermittent and a continuous unobtrusive way.

Focusing on the aging population as a case study, in this paper we present a methodology for extracting and analyzing multi-sensor measurements and use them to identify data clusters that might be linked to clinical profiles. Specifically, we aim to detect and measure aging-associated decline in reserve and function through the usage of novel high technology instruments and assess the recognition accuracy of those sensoring units by comparing the results with the classical clinical evaluation based on a comprehensive geriatric assessment (CGA) [1]. The CGA approaches quantification of frailty by evaluating a person's medical history and prescription, cognitive and emotional status, autonomy, pain, balance and gait patterns, sensory system performance, nutritional status, living conditions, social life, leisure activities and quality of life self-perception. The quantification of these aspects of a person's global health condition is done using questionnaires and standardized scales, as well as cumulative indices.

Using the CGA as guideline, we aim to evaluate the potential of a monitoring system that depicts a subject's health status in a more sophisticated and precise way. The proposed architecture for monitoring and analysis was developed as part of the FrailSafe project [2], which aims at the development of an end-to-end system for sensing and predicting treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions. Using this large-scale data collection methodology, it is possible to make comparisons between the clinical expression and the performance of various measurements and tools to identify and even predict evolution of frailty status.

In the literature, many works focus on the extraction of geriatric indices that asses the older persons' wellbeing [3,4,5,6]. The majority of these works however use raw clinical variables which are acquired on a single time point (during clinical evaluation) and usually with the guidance of a medical personnel. These approaches are time consuming and require skills and expertise, therefore are not accessible to a large percentage of the targeted population. Other works combine the aforementioned measurements into simplified models [7,8,9,10], often without accounting for correlations in the extracted variables or by assuming that the high-dimensional

The research reported in the present paper was partially supported by the FrailSafe Project (H2020-PHC-21-2015-690140) "Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions", co-funded by the European Commission under the Horizon 2020 research and innovation programme.

data fall onto a linear distribution. In other cases, sensor monitoring has been used [11,12,13,14], but these studies don't share the same objective with this work, as they target specific conditions arising from aging-associated decline, such as dementia, and they mainly rely on monitoring the older people in their home environment. On the contrary, we aim to extract geriatric indices data considering several aspects of the older persons' life. The idea of the combined use of sensor data and indicators of the scientific literature is also adopted by Ricevuti et al. [15] who focus on the prevention of mild cognitive impairment and frailty in the aging population supposing that the older person lives in a "smart city".

Summarizing, the proposed approach has several advantages: (i) it monitors the subject's condition without the need of human (medical expert) intervention, (ii) it quantifies decline in multiple clinical domains (physiological, behavioral, cognitive, etc) by fusing data from different devices (iii) it extracts more informative features by calculating local patterns of activities after activity classification and (iv) it reduces data dimensionality by manifold learning, which intrinsically takes into consideration the inherent geometry representation, and allows relevant comparison of individuals to the studied population through a low-dimensional Euclidean space map. The obtained results highlight the potential of our unsupervised framework to extract objective indicators of health status in an automatic way and add to the main contributions of the current work.

The proposed methodology is presented with more details in section II which includes the multi-sensor feature extraction and fusion along with the utilized devices, the dimensionality reduction and clustering techniques. Subsequently the data and evaluation protocol are described followed by the obtained results and some discussion.

II. Methods

FrailSafe's monitoring system [2] consists of various sensors with the ability to record a large amount of data corresponding to several parameters like weight, blood pressure, heart and respiratory rate, physical strength and activity, localization and cognitive performance. Most of the parameters' measurements take place unobtrusively in real time circumstances. More specifically, participants are requested to use a set of devices in their natural environment while performing their usual activities, for a period of several days. This procedure is repeated in predetermined intervals. The proposed method first builds a multi-dimensional profile of each participant by processing the multiple physiological signals to extract meaningful secondary measurements (e.g. heart rate from raw ECG). Statistical features are extracted from the raw or secondary measurements representing physiological and cognitive state, as well as indoor and outdoor mobility behavior. The multi-parametric features are subsequently fused into a long feature vector and introduced into a linear and non-linear dimensionality reduction technique for extracting a small number of distinct patterns.

Clustering is then performed in this low dimensional embedding space in order to discover coherent and well separated groups. The results are evaluated by clustering validity criteria and the identified clusters are also compared with the groups determined by CGA in respect to several clinical metrics from multiple domains. A schematic diagram of the proposed methodology is illustrated in Fig. 1, while more details on data and algorithms are provided next.

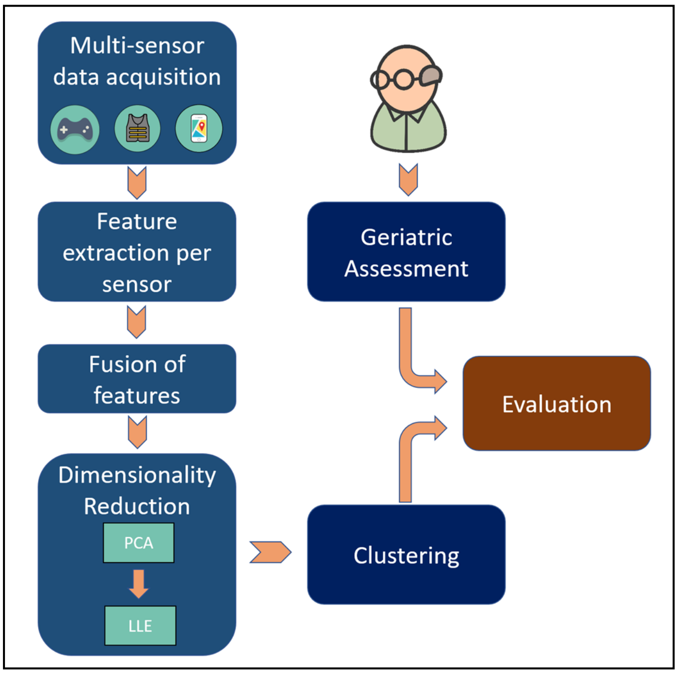

Fig. 1. Proposed methodology for finding clusters of the aging population

A. *Devices and feature extraction*

The monitoring system consists of an aggregation of sensors, devices, and developed software that capture several aspects of the participants' health status (physiological, behavioral, cognitive etc.). In this work we extract features from a wearable device for physiological and kinetic monitoring (WWBS), a dynamometer for strength measurement, a game platform and a global positioning system (GPS) for outdoor monitoring. The devices and the extracted parameters are briefly described next.

1) Wearable WBAN System: The WWBS is a wearable system developed by SMARTEX [16] and composed by a sensorized garment, an electronic device and a software tool. It allows to monitor older adults at home during standard day-time activities, collecting data from their heart (a full ECG lead), respiration, posture and physical activity (IMUs with nine degrees of freedom). The following seven (numeric or categoric) variables were extracted from the plethora of recordings using dedicated software algorithms and represented as time series:

- Acceleration
- Respiration Rate
- Breathing Rate
- Breathing Amplitude
- Heart Rate
- Heart Rate Variability
- R-R interval in ECG

After time synchronization of all channels, the acquired or calculated signals were broken down into segments of main physical activities including moving and not moving conditions [17], such as sitting/standing, laying down, walking forward, walking upstairs or downstairs, and transition of activity. Such a partitioning of the signal (clustering of features) makes interpretation more consistent with human perception and allows extracting features within more uniform clusters. Previous work has also hinted at the utility of clustering to improve prediction accuracy. This has been attributed to its potential to exploit structure in the data and perform compression [18].

A set of statistical features was then calculated for each activity, resulting to an augmented feature vector containing features that correspond to all the activities. Nine statistical properties were calculated: average, standard deviation, 5% and 95% percentiles, mode, skewness, kurtosis, energy, entropy. The mode corresponds to the peak of the histogram, indicating the most frequently encountered value. Kurtosis characterizes the relative peakedness or flatness of the histogram, skewness is a measure of the distribution asymmetry and indicates the direction towards which the distribution is shifted, while energy and entropy are statistical measures of randomness and uncertainty. This feature extraction process resulted to 315 (9 statistical properties × 7 channels × 5 activities) variables for WWBS. The statistical properties were calculated within each *session*, defined as the time span of one day. That means that recordings acquired in consecutive days are treated as independent samples.

2) Game platform and dynamometer: Brainstorm [19] has developed a serious game for older people in which the user is navigating a small plane by applying force to a dynamometer. The game recordings include a set of 13 variables tracked over time:

- Max force
- Average max force
- Average endurance
- Max endurance
- Average score
- Max score
- Average game duration
- Max game duration
- Height over game duration
- Distance over game duration
- Speed over game duration
- Lives over game duration
- Force over game duration

The first eight features were summarized variables exported by the game application characterizing game performance. Endurance is a measure of muscle fatigue in older people and it is measured by calculating the integral of the instantaneous force function during game time [20]. The last five dynamic measurements, concerning the height and the speed of the plane, the distance it covered, the remaining lives of the user and the force applied over the game, were aggregated by extracting their statistical properties similarly to the analysis of the WWBS recordings. Thus, the total number of features from the games was 53 (8 variables + 9 statistical properties × 5 dynamic variables).

3) GPS data: Finally, a number of features was extracted from the GPS data collected through the outdoor monitoring application (GPS logger). These features capture the participants' outdoor mobility behavior and are shown next:

- total distance
- total duration
- total number of steps
- radius covered
- area covered
- average walk speed
- total walk time
- total stop time
- total vehicle time
- walk time percentage
- vehicle time percentage
- stop time percentage
- track number
- track average distance
- track average duration
- track maximum distance
- track maximum duration

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

B. Dimensionality reduction

When dimensionality increases, the volume of the observations' space grows so much that the concept of similarity, distance or nearest neighbor may not even be qualitatively meaningful, thus impeding clustering or classification. Therefore, to facilitate clustering, we first perform dimensionality reduction based on a linear and a non-linear transformation of the data. Specifically, PCA [21] is first applied to extract a set of uncorrelated factors and then LLE [22] is used to find a lower dimensional embedding of the PCA scores.

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

In contrast to PCA which performs only translation and rotation of the data, LLE recovers global non-linear structure from locally linear fits [22]. Let us consider a set of N data samples of dimensionality D in the ambient space R^D. If the data lie on or near a smooth non-linear manifold M of lower dimensionality $d \ll D$ then we can calculate a neighborhood-preserving mapping from the high-dimensional coordinates of each neighborhood to global internal coordinates on the manifold. The intrinsic dimensionality (d) is unknown but we used a small number, such as $d = 2$, to facilitate the subsequent cluster analysis.

C. *Unsupervised clustering*

Clustering performs partitioning of the data space into disjoint parts aiming to find hidden patterns in the data and gain insight. Since it is an unsupervised learning technique it allows an unbiased interpretation of the results, however an inherent difficulty is how to determine the best number of clusters (K). This is usually accomplished by employing a second criterion that measures the robustness of the clustering, but we tested only the values $K = 2$ and 3, to be in accordance with the number of categories of the clinical metrics (see section III.D).

We investigated four popular clustering algorithms [23], namely the agglomerative (Agg), Birch (Bir), spectral clustering (Spec), k-means (KM), and also the combination of their results by majority voting (Comb). The algorithms were executed mostly with the default parameter values [23] which are described in more details next.

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

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

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

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

D. *Mapping of a new (unseen) data sample*

The analysis of the currently available dataset allows to examine non-linear relations in the extracted features and identify data clusters. For new data it might be desired to classify them into the previously extracted clusters without rebuilding the models. For this purpose, the same feature extraction process should be applied followed by projection on the previously calculated principal components to reduce dimensionality and extract the scores in the PCA space (considered as ambient space R^D in the next step). Subsequently, in order to obtain the low-dimensional manifold position of the new sample from the ambient space (scores after PCA), the intrinsic coordinates based on the samples’ neighborhood representation in the high-dimensional space have to be inferred. We can assume an explicit mapping from the ambient space R^D to the manifold space M following the strategy described in [24], i.e. perform the forward mapping by estimating the relationship between R^D and M as a joint distribution, such that there exists a smooth functional which belongs to a local neighborhood. This will result to a small feature vector (with $d = 2$ in our case) that can subsequently be assigned to the closest cluster.

III. Evaluation and Results

A. *Data*

The clinical study was performed in 3 clinical centers: University of Patras, Greece; INSERM-Nancy, France; and MATERIA- Nicosia, Cyprus. Each center has recruited 120 individuals resulting to a total of 360 community living participants aged 70 years and older. All participants provided informed consent and all relevant approvals have been obtained by the competent ethical committees. In order to be able to comply better with the requirements of the study (usage of technical material, cooperation for the testing of novel equipment, need for participant’s feedback and relatively long follow up period), subjects with highly debilitating conditions, such as inability to walk, presence of clinically significant cognitive impairment, or active psychiatric disorder were excluded from the study.

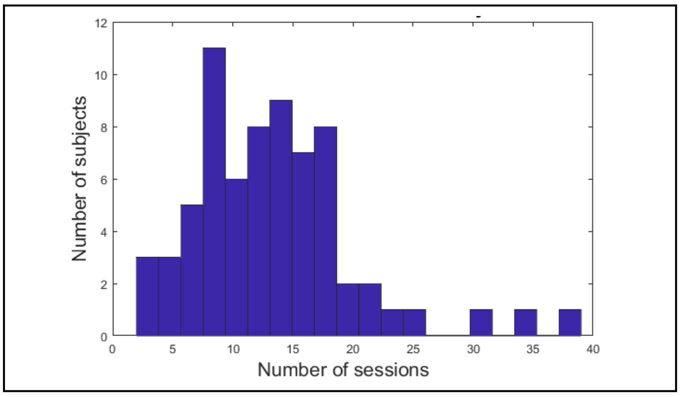

Fig. 2. Histogram of the number of sessions for each participant

Similarly, subjects with serious medical conditions that convey a guarded prognosis (estimated life expectancy of less than 12 months) were excluded as well. Out of all participants, 69 of them had used all the devices (one or more times) by the time the data were extracted for analysis and didn't have missing values in the CGA, thus the rest were excluded from the analysis. The number of sessions was not the same for each participant and is illustrated in the form of a histogram in Fig. 2. The total number of sessions for the 69 participants was 924.

B. Clustering validity

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

TABLE I. SILHOUETTE INDEX USING SQUARE EUCLIDEAN DISTANCE

Num. Clusters	**ALGORITHMS**				
	Agg	***Bir***	***Spec***	***KM***	***Comb***
2	0.96	0.96	0.96	0.96	0.96
3	0.94	0.90	0.94	0.94	0.94

C. Clinical profile

Additionally, to clustering validation with internal criteria, we want to empirically investigate what is the predictive accuracy of the obtained clusters. For this purpose, we used clinical metrics acquired during CGA (shown in TABLE II). Those metrics are selected under the prism of their operational function to quantify frailty, and categorized into domains taking under consideration the interrelationships that run through the implication of each variable in the various aspects of frailty. Before proceeding with classification assessment, we present the clinical characteristics of the participants, evaluated in each session, within each of the identified clusters. We distinguish two types of clinical metrics: the ones quantified in a numeric scale, such as the *TUG (Timed Up and Go) test*, and the categorical ones, such as the frailty status according to Fried [3]. Since CGA does not always coincide temporally with the use of the monitoring devices, the value of the clinical metric at the time of the session was estimated by linear interpolation if the variables were numeric, or by nearest neighbor interpolation if the variables were categorical.

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

TABLE II. CLINICAL PROFILE IN EACH CLUSTER PRODUCED BY THE COMBINED CLUSTERING ALGORITHM WITH K=2.

DOMAIN	**CLINICAL METRIC**	**Cluster 1**[a]	**Cluster 2**[a]
Medical	Orthostatic hypotension	No: 96 Yes: 4	No: 100 Yes: 0
	Hearing	Poor: 27 Moderate/Good: 73	Poor: 100 Moderate/Good: 0
	Vision	Poor: 3 Moderate: 19 Good: 78	Poor: 0 Moderate: 0 Good: 100
General Condition	Unintentional weight loss	No: 100 Yes: .0	No: 0 Yes: 100
Physical Condition	Single foot standing	< 5 sec: 33 > 5 sec: 67	< 5 sec: 0 > 5 sec: 100
	Time get up and go test	8.48 sec	12.57 sec
	Low physical activity	No: 90 Yes: 10	No: 100 Yes: 0
	Low grip strength	No: 61 Yes: 39	No: 100 Yes: 0
	Gait speed (4m)	0.82 m/sec	0.86 m/sec
Cognitive	MoCA score [25]	26.85	21.67
	Subjective memory complain	No: 92 Yes: 8	No: 0 Yes: 100
Psycholo-gical	Self-rated anxiety	3.34	2.39
	GDS-15 score [26]	1.24	5.51
Social	Leisure club participation	No: 16 Yes: 84	No: 100 Yes: 0
	Leisure activities	7.55	7.0
	Telephone calls per week	15.1	4.0
	Visits / social interactions per week	4.45	1.0
Wellness	Self-rated quality of life	8.4	5.06
	Self-rated health status	Bad/Medium: 23 Good: 65 Excellent: 12	Bad/Medium: 100 Good: 0 Excellent: 0
	Self-rated pain	2.39	8.63
Lifestyle	Smoking	Never: 61 Past: 30 Current: 9	Never: 0 Past: 100 Current: 0
	Physical activity	<2h per week: 23 >2h and <5h per week: 26 >5h per week: 51	<2h per week: 0 >2h and <5h per week: 0 >5h per week: 100
Nutrition	MNA screening score [27]	13.72	10.51
Frailty Status	Fried status	Non frail: 0.5 Pre-Frail/Frail: 0.5	Non frail: 0.0 Pre-Frail/Frail: 1.0

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

D. Predictive accuracy

We can assume that the clustering algorithm is good for prediction if it agrees well with a set of hidden labels using a small number of clusters [18]. We selected the previous clinical

TABLE III. CLUSTERING ACCURACY PER CLINICAL METRIC

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

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

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

Classification algorithms could also have been exploited in order to build prediction models for the different clinical metrics. In supervised learning however a different prediction model would have been trained for each target variable, thus such an approach does not provide a unified view of the clinical profile of the subject. In this work our goal was to cluster the participants according to the extracted variables from the monitoring devices without any supervision in learning, and therefore create an unbiased clinical profile modeling frailty. In our future work we will further investigate the prognostic capacity of the monitoring system and also compare the current clustering framework with the use of different machine learning algorithms for supervised classification. Moreover, the analysis of data from a larger cohort is necessary in order to evaluate the generalization ability of the system.

IV. DISCUSSION AND CONCLUSIONS

The paper aims at investigating and specifying appropriate physiological and behavioral characteristics that can be used for defining biomarkers of frailty that can be of a significant predictive value. Cluster analysis helped finding groups of data that were in good accordance with the outcome of clinical tests, indicating that there is high potential in the proposed monitoring system and data analysis framework. This encourages further investigation of the prognostic capacity in terms of predicting frailty transition and subsequent risk factors. The translation of such results in clinical practice could contribute to the organization of strategies for early intervention and to prevent loss of autonomy both in individual and in population scale.

ACKNOWLEDGMENT

The authors want to thank all ICT partners from the FrailSafe Project (Brainstorm, CERTH, Gruppo Sigla, Smartex) for software and hardware support and M. Polykarpou, from Materia Group, Nicosia, Cyprus, for her help with data acquisition and useful discussions on the clinical study.

REFERENCES

[1] L. Z. Rubenstein, "An overview of comprehensive geriatric assessment: rationale, history, program models, basic components," *Geriatric Assessment Technology: The State of the Art. New York, NY: Springer*, 1995.

[2] FrailSafe project. [Online]. https://frailsafe-project.eu/

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

[4] Kenneth Rockwood et al., "A global clinical measure of fitness and frailty in elderly people," *Canadian Medical Association Journal*, vol. 173, pp. 489-495, 2005.

[5] Andrew Clegg et al., "Development and validation of an electronic frailty index using routine primary care electronic health record data," *Age and ageing*, vol. 45, pp. 353-360, 2016.

[6] Fernando G. Peña et al., "Comparison of alternate scoring of variables on the performance of the frailty index," *BMC geriatrics*, vol. 14, p. 25, 2014.

[7] Ingmar Schäfer et al., "Multimorbidity patterns in the elderly: a new approach of disease clustering identifies complex interrelations between chronic conditions," *PloS one*, vol. 5, p. e15941, 2010.

[8] David A. Gold, "An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment," *Journal of clinical and experimental neuropsychology*, vol. 34, pp. 11-34, 2012.

[9] Cécilia Samieri et al., "Dietary patterns derived by hybrid clustering method in older people: association with cognition, mood, and self-rated health," *Journal of the Academy of Nutrition and Dietetics*, vol. 108, pp. 1461-1471, 2008.

[10] N. Cable et al., "Identifying Frail-Related Biomarkers among Community-Dwelling Older Adults in Japan: A Research Example from the Japanese Gerontological Evaluation Study," *BioMed Research International*, vol. 2018, 2018.

[11] Ahmad Lotfi, Caroline Langensiepen, Sawsan M. Mahmoud, and Mohammad Javad Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour," *Journal of ambient intelligence and humanized computing*, vol. 3, pp. 205-218, 2012.

[12] Margriet C. Pol et al., "Effectiveness of sensor monitoring in an occupational therapy rehabilitation program for older individuals after hip fracture, the SO-HIP trial: study protocol of a three-arm stepped wedge cluster randomized trial," *BMC health services research*, vol. 17, p. 3, 2017.

[13] Marek Novák, Miroslav Biňas, and František Jakab, "Unobtrusive anomaly detection in presence of elderly in a smart-home environment," in *ELEKTRO, 2012*, 2012, pp. 341-344.

[14] Marilyn J. Rantz et al., "Using sensor networks to detect urinary tract infections in older adults," in *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, 2011, pp. 142-149.

[15] Giovanni Ricevuti, Letizia Venturini, Sergio Copelli, Franco Mercalli, and Giuseppe Nicolardi, "Data driven MCI and frailty prevention: Geriatric modelling in the City4Age project," in *Research and Technologies for Society and Industry (RTSI), 2017 IEEE 3rd International Forum on*, 2017, pp. 1-6.

[16] Smartex. [Online]. http://www.smartex.it/en/

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

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

[19] Brainstorm. [Online]. http://www.brainstorm3d.com/

[20] Ivan Bautmans, Ellen Gorus, Rose Njemini, and Tony Mets, "Handgrip performance in relation to self-perceived fatigue, physical functioning and circulating IL-6 in elderly persons without inflammation," *BMC geriatrics*, vol. 7, p. 5, 2007.

[21] Ian T. Jolliffe, "Principal component analysis and factor analysis," in *Principal component analysis*.: Springer, 1986, pp. 115-128.

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

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

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

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

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

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