Meeting challenges of activity recognition for ageing population in real life settings

Aimilia Papagiannaki[1], Evangelia I. Zacharaki[1*], Konstantinos Deltouzos[1], Roberto Orselli[2], Anne Freminet[3], Sibora Cela[4] , Elena Aristodemou[5], Marina Polycarpou[5], Marina Kotsani[3], Athanase Benetos[3], John Ellul[4], Vasileios Megalooikonomou[1]

[1] Department of Computer Engineering and Informatics, University of Patras, Greece
[2]Smartex s.r.l , Prato, Italy
[3]Department of Geriatrics, CHRU Nancy, University Hospital of Nancy, Nancy, France
[4]Department of Neurology, University Hospital of Patras, Patras, Greece
[5] Care and Rehabilitation Unit MATERIA, Nicosia, Cyprus
* ezachar@ upatras.gr

***Abstract*—As the global community becomes more interested in improving the quality of life of older people and preventing undesired events related to their health status, the development of sophisticated devices and analysis algorithms for monitoring everyday activities is necessary more than ever. Wearable devices lie among the most popular solutions from a hardware point of view, while machine learning techniques have shown to be very powerful in behavioral monitoring. Nevertheless, creating robust models from data collected unobtrusively in home environments can be challenging, especially for the vulnerable ageing population. Under that premise, we propose an activity recognition scheme for older people along with heuristic computational solutions to address the challenges due to inconsistent measurements in non-standardized environments.**

Keywords; Activity Recognition; SVM Classification; Wearable devices; ADL

I. INTRODUCTION

A great amount of research work has focused on monitoring the physical behavior of ageing population, since the concept of frailty has proven to affect older people's lives and health radically. The operational definition of frailty, according to [1], refers to a clinical syndrome characterized by three or more of the following situations: unintentional weight loss, self-reported exhaustion, weak grip strength, slow walking speed, and low physical activity. It is considered to be a multi-parametric clinical state, in which physical, cognitive, or psychological factors seem to be affecting the appearance of the syndrome as well as its progression. The occurrence of such a syndrome is reflected to all aspects of older people's lives and increases the risk for falls, disability, hospitalization, loss of autonomy and mortality. Because of its complexity, multi-faced nature and not totally clarified pathophysiology, there is a great difficulty in defining, early identifying and preventing frailty. ICT (Information and Communications Technology) technologies tempt to address this unmet need.

A. Related Work

Sophisticated methods for the detection of distinct physical activities have previously been reported in the literature, using a variety of wearable and non-wearable sensors. In [2] the authors reported the use of a smart watch enclosing three different kinds of sensors, namely accelerometer, temperature sensors and altimeter. After performing some calibrating actions on the raw signals and feature selection, neural network and support vector machine (SVM) classifiers were used for classifying the activities of the elderly. In [3] an inertial measurement unit located near the sternum and a thigh wearable sensor were used to detect posture of the elderly with the deployment of a rule-based algorithm. A signature extraction methodology is proposed in [4] using a smartphone's accelerometer, placed at the subjects' pelvis and implementing a threshold-based, or a PCA-based classification algorithm. Extraordinary work on activity classification for the elderly is reported in [5], where the idea of instrumented shoes able to record movement is introduced, for the purpose of discriminating postural transitions, locomotion and walking activities using decision trees. In [6] the possibility of improving ADL (Activities of Daily Living) classification accuracy by applying feature ranking and selection is explored. Activity recognition was performed in [12] using a Hidden Markov Model on recordings from sensors placed in the house and on the body, whereas in [13] the aim was to identify high falls' risk related activities of older people based on a wearable passive Radio Frequency Identification sensor. Analysis was based on data from healthy adult volunteers.

Although many frameworks have been reported in the literature for behavioral monitoring of older people, most of them have been tested on data from young and healthy participants [4][6][13], or the experiments were performed on laboratory conditions, e.g. in [12] a scaled model of a house was used along with a simulated sequence of activities. Those works report high classification accuracy, but results are not directly comparable with uncontrolled

978-1-5386-4294-8/18/$31.00 ©2018 IEEE

monitoring systems in real home environments. In contrast to the aforementioned studies, our approach is tested on data recorded from wearable sensors incorporated in a vest designed for monitoring the physiological signals of older people.

B. Challenges in activity recognition using wearable sensors

Moving from laboratory environment to real-life experiments, researchers are dealing with numerous obstacles that they must overcome, concerning mostly the devices used to monitor older people. A first challenge appears in case the device is not placed with standard orientation, causing rotation of axes. For sensors such as the accelerometer, orientation plays a significant role in recognizing the subject's activity. Thus, a misplacement of a wearable device could easily disorientate the data analysis.

Another issue occurs when a different type of sensors is used across individuals (e.g. accelerometers with different technical characteristics), or the sensors are placed in different location at the body. This is possible when data from different clinical centers are combined or for example after updating hardware and software components, since it causes lack of uniformity in the dataset adding barriers on inference and modeling. When activity annotations are used for model construction (e.g. in supervised learning settings), additional challenges come from the inter- and intra-rater variability, the rater's subjectivity and the interactive nature in the annotation process, especially when frail individuals are instructed and monitored.

The list of challenges could be real long, but we focus on the aforementioned issues since they are likely to occur in studies such as the one reported here. Retrospective solutions to address these issues during data analysis are proposed and discussed in the next section.

C. Aim of current study

In this paper we present an ADL classification scheme based only on accelerometer for the purpose of detecting everyday-life activities of older people and focus on the necessary model reconfigurations for resolving challenges imposed by inconsistent data. The main contributions of the current work are summarized as follows:

- The classification model is trained and tested on older people's recordings exclusively, in real-life conditions.
- An optimized classification model is selected with reduced number of features.
- Variations of the initial model are proposed to address device-relative issues. Specifically, the recordings were acquired using two wearable devices developed for monitoring older people's physiological signals.

The remaining of this article is structured as follows: in section II the methodology applied to build the classification model is described, data acquisition and experimental details are reported in section III, while the results of the current study and related work are presented in sections IV and V correspondingly, followed by the study's conclusions in section VI.

II. MATERIALS AND METHODS

The main device used to record the physiological signals of older people is a wearable solution (vest) that takes its origin from a previously developed product of Smartex [7], 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 also records data from the heart and respiration. It has been developed for the purpose of long term monitoring of several physiological parameters, together with data from some IMUs to be used to better classify and parameterize user's movements and physical activity, in the context of the European research project FrailSafe [8].

The classification scheme includes the same steps, as reported in [6]:

- Data acquisition
- Split in training and test sets
- Preprocessing
- Feature extraction
- Model building with training samples
- Classification of test samples

The purpose of the classification procedure is to discriminate the following activities: sit/stand, laying, walking, walking upstairs/downstairs, transition between activities. Only accelerometer-generated recordings are selected as data input for classification, although the gyroscope and magnetometer recordings could be used as well. This decision was made due to the fact that the current approach focuses on a minimal use of resources, for computational and performance-related reasons. The accelerometer has a sampling rate at 25Hz with unit value 0.97 e10^{-3}g.

A. Preprocessing, Feature extraction, Classification

The classification scheme is based on previously reported work [6]. The preprocessing procedure involves the separation of body acceleration from gravity acceleration, as reported in [9]. Specifically, the raw 3-axial signals from the accelerometer were initially preprocessed using low-pass filtering to separate the body and gravity acceleration components. The accelerometer Jerk signals were also calculated, as well as the magnitude of the tri-axial signals. The recordings were then split into fix-width sliding time windows and a set of statistical features were calculated for each time window. These were a subset of the proposed features reported in [9], namely: mean value, standard deviation, median absolute deviation, largest value in array, smallest value in array, signal magnitude area, energy measure, interquartile range, signal entropy, autoregression coefficients, correlation coefficient between two signals, index of the frequency component with largest magnitude, weighted average of the frequency components to obtain a mean frequency, skewness of the frequency domain signal, kurtosis of the frequency domain signal and energy of a frequency interval within the 64 bins of the FFT of each window. The resulting feature vector contained 254 features.

Subsequently, the features were normalized, in order to avoid skewing the analysis by specific features' scale. This was achieved by applying the widely used z-score normalization (standardization), that centers the features at zero and scales them to have unit variance. The parameters that used to scale the training features, were stored in a data structure to be subsequently used for the standardization of the test features.

The normalized feature vectors of the training samples were introduced to an SVM classifier. SVMs find the hyperplane that discriminates the classes by maximizing the margin in between. Once the model is calculated, it can be used to classify new instances. To evaluate the model's performance on unknown data, the classification accuracy is calculated according to the test samples' labels.

B. Reducing differences across devices

As discussed previously, performing experiments with sensors of slightly different technology, or with a different sensor placement prohibits the use of a uniform classification model. In particular for the data used in this study, we explored whether any differences occurred between recordings from the two incorporated devices, by asking two young volunteers to perform daily activities while simultaneously wearing the two devices used in the study placed at the center and laterally of the chest, respectively. It was revealed that a scaling difference occurred between the two types of measurements. To address this issue, we define a reference space and normalize all recordings with regards to the reference measurements using a baseline correction technique. The baseline was defined as the mean value of time segments with small standard deviation for each axis. Correction was performed by aligning the baseline of each new recording to the reference baseline of the corresponding channel. After baseline correction, the classification model to be used is selected according to the rules identifying errors in sensor's orientation, as described in the next section.

C. Resolving the rotation of axes issue

To overcome the challenge of axes rotation, we investigated the possibility to automatically identify mis-oriented device's data and map them back to a reference space. Although some heuristic rules helped us to recognize mis-orientations in most of the cases, no robust automatic technique was found to be always successful. Thus, we decided to extract also rotation-invariant-features and build a substitute classification model. To achieve this, each triplet of features extracted from the three axes, X, Y and Z, was reduced to only one feature computed as the mean of the three axes' corresponding features. These features are insensitive to the orientation, and thus misplacement, of sensors.

Although the rotation-invariant features are preferable in the case of inconsistent data, they are expected to have lower predictive value, because some activities, such as walking, are strongly related to a particular axis. Therefore, it would not be the best practice to replace the detailed model by the less complex one overall. Accordingly, we had to deal with the problem of automatically recognizing whether the wearable device had been misplaced or not and select the corresponding model. The problem was addressed by learning the distribution of measurements in the reference space (correct orientation) defined by the training set. A two-steps heuristic approach was introduced for that purpose. The first rule determines whether the vertical axis (axis X), which is the most prominent axis due to gravity, coincides with the one used during training. This was achieved by outlier detection assuming that the largest amount of measurements (80^{th} percentile) of test and reference data should be in the same range. If so, then the axis-dependent model is selected for classification, otherwise the second rule is examined.

The 2^{nd} rule is used to examine more thoroughly cases in which the signal has a shifted baseline value, therefore might fall outside the predefined range. The signature for the recording is based in this case on the overall distribution of measurements along all three axes and specifically on the 3D histogram of raw values (within ~1-hour interval). Outlier detection is performed by computing the pairwise distance between the test signal and each of the reference signals (from the training subjects). If the minimum distance is smaller than a threshold (θ), the axis-dependent model is selected, otherwise classification is based on the axis-invariant scheme. The threshold θ was determined by examining the normal variation, i.e. it was equal to the maximum within-training-subjects' distance. The chosen distance metric was a 3-D version of the Kolmogorov-Smirov distance [14].

D. Enforcing temporal coherency of activities

In order for the classification results to be physically probable, we made the assumption that the minimum duration of an activity could not be less than a threshold. This assumption was made to ensure that the predictions describe coherent transitions between activities. To that end, we "smoothed" the activity signal using a moving majority voting filter of size equal to the predefined minimum duration of each activity. We used a value of 4 sec for the older people.

A schematic representation of the complete methodology is illustrated in Fig. 1.

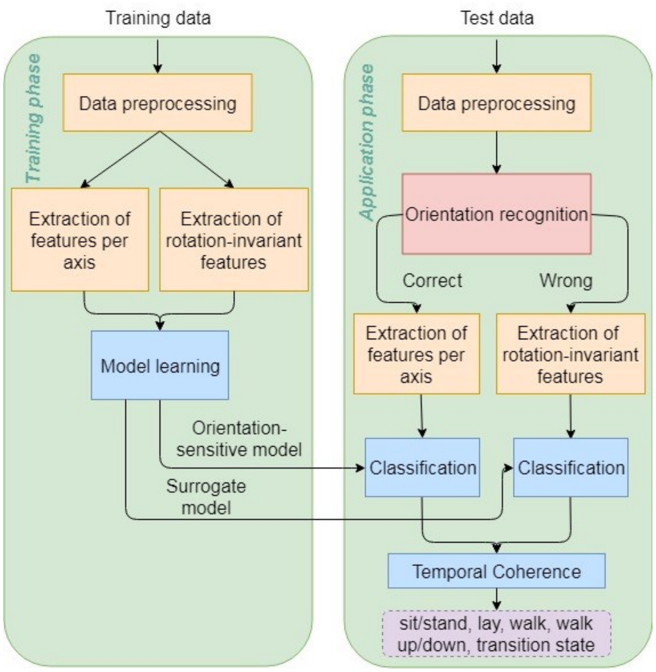

Figure 1: Pipeline of activity recognition methodology

III. Experimental Procedure

A. Data acquisition and annotation

The recordings were obtained by twenty subjects (17 females and 3 males; age: 70-92 years) who participated in the FrailSafe project [8]. Considering their health condition, they were categorized according to Fried's criteria [1], resulting in 10 non-frail and 10 pre-frail subjects. All participants were instructed to perform a set of activities while wearing one of the two devices. The protocol was performed in clinical centers of three different countries, and included the following actions:

- Standing for 1 minute
- Sitting for 1 minute
- Walking for 1 minute
- Walking upstairs for 30 seconds
- Walking downstairs for 30 seconds
- Laying for 30 seconds

Upstairs and downstairs walking was only performed if stairs were accessible by the subjects in their residence. Times were recorded by the medical instructors while the participants performed the ADL protocol. The annotations of 8 subjects were used to train the classifiers, while the rest of the subjects were left for testing. Four-fold cross-validation on the training set was used to optimize the parameters for both classification models. The recordings of the subjects selected for training had the same orientation, which was used as reference space (with -X denoting the vertical axis). An example of the raw acceleration signals during different activities is illustrated in Fig. 2.

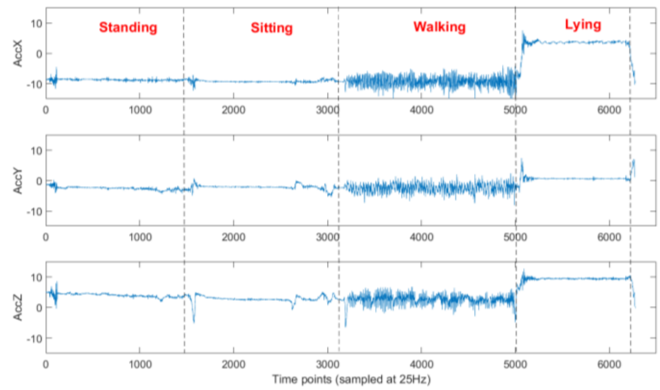

Figure 2: Acceleration signals while performing ADLs

Although six activity classes were initially defined, the classes *sitting* and *standing* were merged into one class, as well as *walking upstairs* and *walking downstairs*. This was performed based on previously reported work that suggests that these classes are not easily separable [6]. Additionally, the time windows corresponding to the first five seconds of the beginning of each activity, were automatically annotated as "transition state", to indicate the time required to switch between two different activities. To that end, the final classes were: *sitting/standing, laying, walking, walking upstairs/downstairs, transition state.* The number of samples that corresponded to each class was different. To deal with the imbalanced classes, we used weights (inverse proportional to the class size) in the classification function.

B. Implementation details

A SVM classifier with radial basis function (*rbf*) kernel from the LIBSVM library [10] was used to train the orientation-sensitive as well as the rotation-invariant classification models. Grid search on parameters *C* and *gamma* of the *rbf* kernel was performed, in order to achieve optimum cross-validation accuracy on the training set.

Furthermore, dimensionality reduction and removal of irrelevant features was performed using the Relief-F algorithm [11], since this algorithm is widely known to perform well on multiclass problems. The best cross-validation accuracy was achieved using only the 10 higher ranked features (out of the initial 254 features). In the case of the rotation-invariant model, a total of 40 out of initially 90 features were selected using the same feature selection technique.

IV. RESULTS

Feature selection before classification revealed the most important variables. We observed that features from the Y-axis did not contribute to the final classification model.

The standard classification model achieved 89.46% cross-validation accuracy (computed as the number of correctly classified instances to the total number of instances on each fold and averaging across folds). The model was then applied to 8 test subjects from different clinical centers and whose data were acquired by placing the sensors according to the reference orientation. The classification accuracy of the independent test set reached 81.7%. The mean confusion matrix is illustrated in Table II.

TABLE I: MEAN CONFUSION MATRIX

	Predicted					
	Classes	Sit/Stand	Laying	Walking	Walking up/down	Transition
Actual	Sit/Stand	96.08	0	0.76	0	3.16
	Laying	0	86.75	1.65	0	11.60
	Walking	8.26	0	74.33	1,56	15.85
	Walking up/down	0	0	100	0	0
	Transition	36.07	2.73	18.03	0	43.17

It appears that *sit/stand* is the most easily predictable activity by the model, since it is correctly classified in almost all cases. The *walking up/down* class has zeros everywhere, since there were no test subjects with annotations on this class (to be precise, only one subject had performed this activity, and it was decided to include it in the training set). Class *walking* has some false negatives identified mostly as transition between activities, which is not surprising since a transition state is not a class with a standard pattern. Similarly, the class *transition* is intermixed with all other classes. This might be also attributed to imperfect annotations of the transition samples, extracted within constant time windows (located at the beginning of each activity) without any visual inspection or individualized correction.

The rotation-invariant model reached 75.56% cross-validation accuracy. For the evaluation phase we used recordings of 8 test subjects, two of which had rotated axes, while the rest six were manually rotated in several ways, in order to assess the model's performance. The mean classification accuracy of these subjects was 70.31%. To better interpret the models' performance, we juxtapose the classification accuracy of the rotated subjects' recordings when applying the orientation sensitive model, with the accuracy of the surrogate model. The results are depicted in Table III. It is evident that the surrogate model boosts the accuracy significantly, thus being effective in addressing the rotation of axes issue.

TABLE II: MODELS' PERFORMANCE IN CASE OF AXES ROTATION

Subject	**Classification Accuracy %**		**Increased by %**
	Orientation sensitive model	**Surrogate model**	
1	19.43	**60.19**	40.76
2	19.60	**60.40**	40.80
3	23.08	**72.60**	49.52
4	29.84	**78.01**	48.17
5	25.27	**64.52**	39.25
6	7.03	**69.17**	62.14
7	7.19	**87.05**	79.86
8	25.37	**78.54**	53.17

Regarding the evaluation of the baseline correction approach for data acquired by different devices, we used recordings acquired with the older device from 4 volunteers (2 young and 2 elderly) and evaluated the classification. The mean accuracy was initially 37.5% with the orientation sensitive model (trained with data from the newer device), while after performing baseline correction it was increased to 61.7%.

Finally, we assessed the overall accuracy of the end-to-end pipeline on 12 test subjects (not used during training) whose recordings were acquired by wearing either of the two devices in real settings, i.e. with the sensors placed correctly or rotated, and by letting the algorithm decide which model to use. The average accuracy was 71.8%.

V. DISCUSSION

A direct comparison with other studies is not feasible due to differences in the experimental setup including the type of activity, as well as the use of different classification performance metrics. Nevertheless, the different approaches reported in this paper are compared in Table IV, in respect to incorporated sensors, classification technique, and performance.

TABLE III: COMPARISON WITH OTHER STUDIES

Study	**Sensor**	**Method**	**Evaluation metric**	**Performance**
Proposed	IMU at sternum	SVM	Accuracy	81.7%
[2]	Smart watch, accelerometer	NNs, SVM	Accuracy	90.23%
[3]	IMU at sternum and thigh sensor	Rule-based	ROC analysis	97.2%
[4]	Smartphone	Threshold & PCA	Sensitivity/ Specificity per class	PCA high specificity and sensitivity
[5]	Instrumented shoes	Decision Tree	Accuracy	97.41%

Other works report higher accuracy than the current work, but there can be many reasons for it. First of all, most studies use data from younger participants, selected to be in good physical health. This results to a more homogeneous group and allows the collection of a larger number of samples. Since our work is targeted to the ageing population, functional status and age variability are confounding factors when building a monitoring system for older people. Second, we use a single sensor (accelerometer) in a single location, whereas the combination of more sensors could affect the classification performance, especially if located in different parts of the body. Third, the data of this study were acquired from everyday life in home environments whereas most of the studies report accuracies in controlled, simulated or laboratory settings.

Finally and more importantly, the reported performance is not always calculated with respect to an independent test set, but may refer to accuracy of subject-specific models (where train and test samples come from the same subject) [6].

VI. CONCLUSIONS

The current study aimed at developing an activity recognition scheme for the ageing population with respect to real-life challenges arising from device-generated or human-related parameters. Data from older participants with different levels of frailty or functional conditions were used to train and test an end-to-end modeling framework developed to address those challenges Promising results support the use of the proposed activity recognition scheme for unobtrusive monitoring of older people.

ACKNOWLEDGMENT

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 program.

REFERENCES

[1] Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., ... & McBurnie, M. A. (2001). Frailty in older adults: evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *56*(3), M146-M157.

[2] Chernbumroong, S., Cang, S., Atkins, A., & Yu, H. (2013). Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications*, *40*(5), 1662-1674.

[3] Liu, J., Sohn, J., & Kim, S. (2017). Classification of Daily Activities for the Elderly Using Wearable Sensors. *Journal of healthcare engineering*, *2017*.

[4] Andò, B., Baglio, S., Lombardo, C. O., Marletta, V., Pergolizzi, E. A., Pistorio, A., & Valastro, A. (2015). ADL Detection for the Active Ageing of Elderly People. In *Ambient Assisted Living* (pp. 287-294). Springer, Cham.

[5] el Achkar, C. M., Lenoble-Hoskovec, C., Paraschiv-Ionescu, A., Major, K., Büla, C., & Aminian, K. (2016). Instrumented shoes for activity classification in the elderly. *Gait & posture*, *44*, 12-17.

[6] Pippa, E., Mporas, I., & Megalooikonomou, V. (2016). Feature Selection Evaluation for Light Human Motion Identification in Frailty Monitoring System. In *ICT4AgeingWell* (pp. 88-95).

[7] http://www.smartex.it/en/

[8] http://frailsafe-project.eu/

[9] Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2012, December). Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International workshop on ambient assisted living* (pp. 216-223). Springer, Berlin, Heidelberg.

[10] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 27.

[11] Kononenko, I. (1994, April). Estimating attributes: analysis and extensions of RELIEF. In *European conference on machine learning* (pp. 171-182). Springer, Berlin, Heidelberg.

[12] Sebestyen, G., Stoica, I., & Hangan, A. (2016, September). Human activity recognition and monitoring for elderly people. In *Intelligent Computer Communication and Processing (ICCP), 2016 IEEE 12th International Conference on* (pp. 341-347).

[13] Ranasinghe, D. C., Torres, R. L. S., & Wickramasinghe, A. (2013, June). Automated activity recognition and monitoring of elderly using wireless sensors: Research challenges. In *Advances in Sensors and Interfaces (IWASI), 2013 5th IEEE International Workshop on* (pp. 224-227).

[14] Chakravarty, I. M., Roy, J. D., & Laha, R. G. (1967). Handbook of methods of applied statistics.