Physical Activity Classification for Elderly People in Free Living Conditions

(Article begins on next page)

20 April 2024

Physical Activity Classification for Elderly People in Free Living Conditions

Muhammad Awais, *Student Member, IEEE*, Lorenzo Chiari, *Member, IEEE*, Espen A.F. Ihlen, Jorunn L. Helbostad, and Luca Palmerini

Abstract—**Physical activity is strongly linked with mental and physical health in the elderly population and accurate monitoring of activities of daily living (ADLs) can help improve quality of life and well-being.**

This study presents and validates an inertial sensors-based physical activity classification system developed with older adults as the target population. The dataset was collected in free living conditions without placing constraints on the way and order of performing ADLs. Four sensor locations (chest, lower back, wrist, and thigh) were explored to obtain the optimal number and combination of sensors by finding the best tradeoff between the system's performance and wearability. Several feature selection techniques were implemented on the feature set obtained from acceleration and angular velocity signals to classify four major ADLs (sitting, standing, walking, and lying). Support vector machine was used for the classification of the ADLs.

The findings show the potential of different solutions (single-sensor or multi-sensor) to correctly classify the ADLs of older people in free living conditions. Considering a minimal set-up of a single sensor, the sensor worn at the L5 achieved the best performance. A two-sensor solution (L5 + thigh) achieved a better performance with respect to a single-sensor solution. On the other hand, considering more than two sensors did not provide further improvements. Finally, we evaluated the computational cost of different solutions and it was shown that a feature selection step can reduce the computational cost of the system and increase the system performance in most cases. This can be helpful for real-time applications.

Index Terms—**Physical activity classification, elderly people, free living conditions, inertial sensors, feature selection.**

I. INTRODUCTION

PHYSICAL activity has a strong influence on physical and mental health. Physical inactivity in older adults (above 65 years) is associated with a higher risk of falling, mobility disorders, low muscle strength and loss of independence [1]. An active lifestyle can minimize the development of many disabling conditions and chronic diseases [2] and can help to achieve healthy ageing and well-being. A report by the World Health Organization (WHO) shows that physically active older adults possess a higher level of muscular fitness and functional health, a lower risk of falling and better cognitive function [3]. WHO recommends older people to perform moderate-to-intense physical activity for at least 30 minutes, five times per week in bouts no shorter than 10 minutes to achieve health benefits [3]. Hence, profiling activities of daily living (ADLs) over an extended time period can provide better knowledge about active and sedentary periods in order to improve quality of life.

Monitoring ADLs has achieved major advancements in recent years due to the technological progressions in wearable devices in terms of processing power, battery life, miniaturization, and cost effectiveness. A substantial amount of work has been done during the past years to classify the ADLs using inertial sensors (accelerometers and, in some cases, gyroscopes) [4-16]. However, most of the existing systems are developed and validated on datasets acquired from young adults [4, 7, 8, 10, 12, 13, 15-20] and few systems have included older adults in their development process [6, 11, 21-23]. Rosario et al [11] trained the PAC system on younger adults and tested it on the older adults in the laboratory based conditions. Their findings showed that the performance of the system decreased significantly when trained on younger adults and tested on older adults as compared to system's performance when trained and tested on older adults. From this, we can infer that younger adults based trained system are not able to effectively classify the ADLs. Furthermore, the scope of the physical activity classification (PAC) systems developed in the literatures is often limited to laboratory-based environments and very little is known about the validity of such systems in free-living conditions [24]. The ADLs performed in controlled laboratory environments are supervised and predefined while, in real-life conditions, ADLs are of course performed in an unsupervised and more natural way.

Our earlier work [25] focused on this issue by highlighting the gaps and limitations imposed by free living conditions on existing PAC systems which have been developed in laboratory-based environments. To provide an unbiased and fair comparison we worked to have full control over the nature of the dataset (set of ADLs, studied population), sampling frequency, window size and cross validation procedure. Our findings suggested that the performance of laboratory-based systems is degraded when exposed to real-life conditions, emphasizing the need to design and develop PAC systems that

are natively fed by real-life data [25].

Therefore, the current study is in continuity with earlier works to propose a PAC system that can classify the ADLs of older adults in free-living conditions. The main objectives of this work are as follows.

1. To develop an inertial sensors-based PAC system trained and tested in free-living conditions for elderly people;
2. To analyze the impact of feature selection on system performance (F-measure and computational complexity);
3. To analyze multi-sensor versus single-sensor solutions, in order to highlight the optimal number of sensors that can achieve an acceptable level of performance.

The rest of the article is organized as follows. Section II describes the dataset and processing steps used for the development of the PAC system. In section III, results are presented with a detailed discussion on the findings. Finally, section IV concludes the study.

II. MATERIALS AND METHODS

A. Data Collection in Free Living Conditions

The dataset used in this study is a subset of a larger dataset collected by the Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences at the Norwegian University of Science and Technology (NTNU) under the ADAPT project (A Personalized Fall Risk Assessment System for promoting independent living). The ADAPT dataset was collected in free-living conditions, where the subjects were free to perform ADLs in an unsupervised way. The ways of performing activities were natural and unstructured. A total of 20 older adults (76.4 ± 5.6 years) participated in the protocol, performing a variety of ADLs. The subjects were instrumented in the lab and then they went home afterwards to perform the ADLs in free-living conditions. Performing the free-living protocol at home environment obtains more natural pattern and distributions of ADLs. Subjects were instructed to do their usual ADLs in a natural way, but in addition include defined activities as a part of free-living protocol (see Table A1) without any instruction or supervision on how to perform them. Therefore, they could choose whether to perform these tasks (Table A1) or not and chose how and when they wished to perform the activities. Predetermined categories of ADLs were; sitting (48.1%), standing (22.2%), walking (14.2%), transitions (5.1%), shuffling (4.7%), leaning (2.32%), lying (1.3%), ascending stairs (0.87%), descending stairs (0.65), picking (0.56%), kneeling (0.02). The values inside the parenthesis shows the percentage distribution of the respected ADL in the dataset. The total length of recording were 28.7 hours for the 20 subjects [26]. Data from the wrist sensor was missing for four subjects due to technical issues during recording, and these subjects were excluded from analysis. Consequently, the analyses have been performed on the remaining 16 subjects. The ADLs analyzed in this particular study were; sitting, standing, walking

and lying, and the detailed summary of these ADLs is provided in Table I. Various parameters were computed i.e. quantity (how many times a single ADL occurred in all subjects), mean (average duration of each ADL in sec), STD (standard deviation of each ADL in sec), min (minimum bout duration of each ADL in sec) and max (maximum bout duration of each ADL in sec).
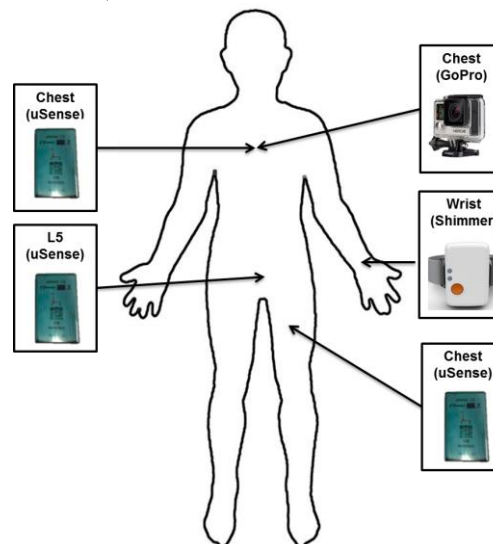
Fig. 1. Sensors' placements used in the free-living protocol

TABLE I
CHARACTERISTICS OF THE TOTAL DATASET OF ADLs ANALYZED FROM THE FREE-LIVING CONDITIONS (N=16)

ADLs	Total (s)	Quantity (s)	mean bout (s)	STD (s)	Min. bout (s)	Max. bout (s)
sitting	48425.80	497	97.44	200.74	0.04	2075.64
standing	23462.72	4304	5.45	12.27	0.03	388.52
walking	14771.81	2617	5.64	8.75	0.28	139.56
lying	1280.32	12	106.69	154.02	3.48	583.84

TABLE II
DESCRIPTION OF THE SENSORS USED FOR DATA ANALYSIS [26]

Device	uSense	Shimmer3
Location	Thigh, L5, Chest	Non-dominant Wrist
Size	67 × 42 × 10 (mm)	51 × 34 × 14 (mm)
Weight	36 g	23.6 g
Sampling frequency	100 Hz	200 Hz
Battery Life	72 h	4.6 days @ 1 kHz
Sensor	3D accelerometer, gyroscope and magnetometer	3D accelerometer, gyroscope and magnetometer
Measurement range	±2 g, ±250_/s, ±1200 _T	±8 g, ±1000_/s, ±1900 µT
Company/ Institution	University of Bologna, Italy	Shimmer, DCU Alpha, Dublin 11, Ireland

The mean length of the analyzed data was 1.5 hours per subject. A total of nine inertial sensors were part of the ADAPT project [26] and a subset of these sensors were used in our analysis: chest (C), wrist (W), lower back (L5), and thigh (T) as shown in Fig. 1. The detailed description of the sensors used for data analysis is presented in Table II. The synchronization between the sensors and the camera unit (GoPro worn on the chest, pointing downwards) was accomplished by performing a series

of static and dynamic movements of the sensors in view of the camera unit, before attaching the sensors to the subjects. These movements were evident in the root-sum-of-squares of acceleration signal and in the video recording. By associating these movements, the synchronization between the camera and sensors was achieved. The wrist sensor from Shimmer was down sampled to 100 Hz to keep the same frequency for all sensing units. The detailed description of the ADAPT dataset is presented in the study protocol by Bourke *et al.* [26].

B. Ground Truth for ADLs

The ground truth information was captured using the video recordings of GoPro camera unit (Fig. 1). The original sampling frequency (25Hz) of the camera was up-sampled to 100 Hz to maintain the uniformity in the sampling frequencies of all sensors. Furthermore, a majority voting scheme was implemented to assign the window labels, i.e. if a window of 5 s (500 samples) contains 400 samples of standing and 100 samples of walking then the assigned label to this window would be standing [11]. The video recordings were annotated by five raters, which were instructed about the marking procedures and activity definitions. The overall agreement of video labelling assessed with Cohen's kappa was 90.05%. The inter-rater reliability statistics are provided in Table A2.

It should be noted that there were spurious bouts in the labelled data. For instance, the minimum duration of a walking bout was 0.28 s (see Table I). Such short bouts are not clinically relevant. However, the impact of these short bouts in the final labelling was limited since they provided only small percentages in the majority voting i.e., a bout of 0.28 s would correspond to less than 6% of a window of 5 seconds.

C. Features

Several features were extracted from acceleration and angular velocity (Table III) which are described in detail in the following subsections. Each of the features listed in Table III were computed across a time window of N samples (N=500, i.e. 5 s of data) with a 50% overlap. The letters x, y, and z in Table III represent the mediolateral, anteroposterior, and vertical axes, respectively. However, it is important to note that the sensor frame is moving, so the axes of the sensors are approximately aligned with this body-centric axes.

1) *Features Extracted from Acceleration*

The mean, variance, and correlation between axes were computed from the raw acceleration (Table III; features # 1-9). The gravitational acceleration (GA) components were obtained by low-pass filtering the signal with a third-order low-pass elliptic filter of infinite impulse response with a cutoff frequency at 0.25 Hz [27]. The mean of all three GA components [8] were used as separate features. The GA component was also used to compute the tilt angle [5, 27] from the expression below:

$$\text{tilt}_{\text{angle}} = \text{acos}(z) \qquad (1)$$

where z represents the gravitational component along the vertical axis computed by taking the mean of N samples,

resulting into a single value for the tilt angle obtained from each window of N samples.

The bodily motion components of acceleration (BA) were extracted by subtracting the raw acceleration from the GA component. The BA components were used to extract the signal magnitude area (SMA) [27, 28], energy [7], and the magnitude vector (MV) [9] from the expressions (2-4) and in Table III (features # 10-20):

$$SMA = \frac{1}{N} \sum_{i=1}^{N} (|x(i)| + |y(i)| + |z(i)|) \qquad (2)$$

$$Energy = \sum_{i=1}^{N} (|x(i)|^2) \qquad (3)$$

where the energy of the signal was computed by the sum of the time series samples squared.

$$MV = \sqrt{x^2 + y^2 + z^2} \qquad (4)$$

where x, y, z in (2-4) are from BA components. The mean, variance, and energy were then computed from the MV.
Note: The sqrt is monotonic and do not add any extra information. Thus, MV was computed without sqrt operation to reduce the computational time.

The jerk signal was derived by low-pass filtering the raw acceleration (4th order Butterworth infinite impulse response low-pass filter with a cutoff frequency at 20 Hz) and then taking the first derivative of acceleration. Features extracted from the jerk signal include the mean, variance, correlation between the axes, energy, and SMA (Table III; features # 21-33). Furthermore, the mean, variance, and energy were also computed (Table III: features #: 34-36) from the MV (4) of the jerk signal from acceleration.

2) *Features Extracted from Angular Velocity*

The mean, variance, correlation between axes, SMA, and energy (Table III; features # 37-49) were extracted from angular velocity and jerk signal of angular velocity (Table III; features # 53-65). The mean, variance, and energy of the MV from angular velocity (Table III; features # 50-52) and MV from jerk signal (Table III, features #: 66-68) were also derived. The jerk signal was obtained by low-pass filtering (4th order Butterworth low-pass filter with a cutoff frequency at 20 Hz) the angular velocity and then taking its second derivative.

3) *Features Extracted from the Sensor Combinations*

Apart from features extracted from signals of a specific sensor, there are features derived from sensor combinations (i.e. acceleration attenuation constant and correlation across each sensor combination). Both of these features were computed by filtering the raw acceleration with 4th order Butterworth low-pass filter with a cutoff frequency at 20 Hz [29]. The ability to attenuate the acceleration from the lower body segments (i) to the upper body segments (j) was described by the acceleration attenuation constant [29]:

$$C_{ij} = (1 - \frac{RMS_j}{RMS_i}) * 100 \qquad (4)$$

Therefore, a total of 6 sensor combinations ($C_{TW}, C_{TL}, C_{TC}, C_{WL}, C_{WC}, C_{LC}$) were formed from the four sensor locations (T, W, C and L5) resulting in 18 features (6×3). The correlation between each sensor combination was also analyzed resulting in 36 features (18 from acceleration, 18 from 18 from angular velocity) obtained from 6 sensor combinations ($\rho_{TW}, \rho_{TL}, \rho_{TC}, \rho_{WL}, \rho_{WC}, \rho_{LC}$). These features were considered only if a combination of sensors (see Table III) was available in the chosen sensor solution.

TABLE III
FEATURES COMPUTED FROM EACH SIGNAL

Feature #	Feature description
1-3	Mean of acceleration (x, y, z) ^{a}
4-6	Variance of acceleration (x, y, z)
7-9	Correlation between axes of acceleration (x, y, z)
10-12	Energy of BA component (x, y, z)
13	Signal magnitude area (SMA) of BA component
14	Tilt angle obtained from GA component in vertical direction
15-17	Mean of GA components (x, y, z)
18	Mean of MV of BA component
19	Variance of MV of BA component
20	Energy of MV of BA component
21-23	Mean of jerk signal from acceleration (x, y, z)
24-26	Variance of jerk signal from acceleration (x, y, z)
27-29	Correlation between the axes of jerk signal from acceleration (x, y, z)
30-32	Energy of the jerk signal from acceleration (x, y, z)
33	SMA of the jerk signal from acceleration
34	Mean of MV of jerk signal from acceleration
35	Variance of MV of jerk signal from acceleration
36	Energy of MV of jerk signal from acceleration
37-39	Mean of angular velocity (x, y, z)
40-42	Variance of angular velocity (x, y, z)
43-45	Correlation between axes of angular velocity (x, y, z)
46-48	Energy of angular velocity (x, y, z)
49	SMA of the angular velocity
50	Mean of MV of angular velocity
51	Variance of MV of angular velocity
52	Energy of MV of angular velocity
53-55	Mean of jerk signal from angular velocity (x, y, z)
56-58	Variance of jerk signal from angular velocity (x, y, z)
59-61	Correlation between the axes of the jerk signal from angular velocity (x, y, z)
62-64	Energy of jerk signal from angular velocity (x, y, z)
65	SMA of the jerk signal from angular velocity
66	Mean of MV of jerk signal from angular velocity
67	Variance of MV of jerk signal from angular velocity
68	Energy of MV of jerk signal from angular velocity
69-71 ^{b}	Attenuation constant between sensor combinations of acceleration (x, y, z)
72-74 ^{b}	Correlation between sensor combinations of acceleration (x, y, z)
75-77 ^{b}	Correlation between sensor combinations of angular velocity signal (x, y, z)

^{a} *x, y, z shows that all three axes of the signal (can be raw acceleration, BA component, angular velocity, jerk etc.) are used to compute the respective features.*
^{b} *Features from 69-74 were considered only if a sensor combination was analyzed.*

For example, if the performance of the single sensor on L5 was analyzed then none of the across sensor features were considered. Then, if the performance of a sensor combination is being analyzed (e.g. thigh and L5), then 3 features are obtained from attenuation constant and 6 features from correlation (3 form acceleration, 3 from angular velocity) resulting into 9 additional features (Table III; features # 69-77). Therefore, the total number of features in a sensor combination will be 145 (i.e. 68 features from the thigh sensor, 68 features from L5 sensor, 3 features from attenuation constant C_{TL}, and 6 features from correlation ρ_{TC}). Similar comparisons were done for other multi-sensor solutions (231 features from three sensors, 326 features from four sensors).

D. Class Distribution in the Dataset

The dataset [26] originally contains eleven ADLs. We considered only four ADLs (standing, walking, sitting, and lying) for analysis. The choice behind the selection of 4 classes is motivated by the fact that these are the most commonly performed activities in the elderly population and to keep consistency with our previous work [25]. The pie chart in Fig. 2 shows the percentage distribution of the four ADLs of the 16 subjects. The values inside the legend show the number of instances belonging to each class (an individual instance corresponds to 5 s or 500 samples of data).
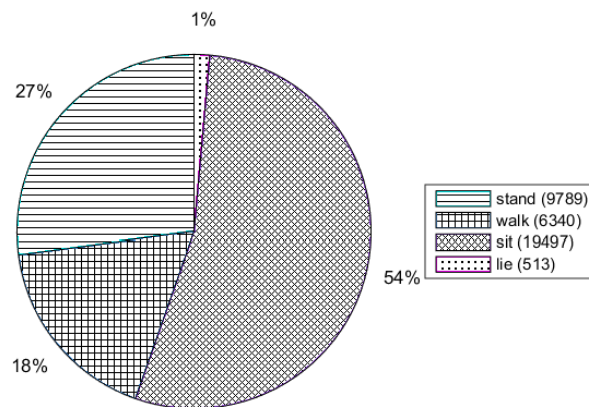
Fig. 2. Percentage distribution of the four ADLs (sitting, standing, walking, and lying) for the 16 subjects in the dataset.

E. Feature Selection

The selection of a subset of features is an important step as the feature vector may contain redundant features. This procedure not only reduces the computational complexity of the system but also reduces the feature extraction time and classification time of the machine learning algorithm. Therefore, in order to eliminate redundant and irrelevant features, we implemented and compared the following feature selection methods: correlation-based feature selection (CFS), fast correlation based filter (FCBF), and ReliefF.

In CFS, correlation between features and class labels are computed along with inter-correlation between features to find the redundancy between them. The final feature subset consists of features exhibiting high correlation with the classes and very low intercorrelation between features. A feature subset is determined by computing linear correlation [30].

The FCBF method computes the predominant correlation among features and classes, and selects predominant features by eliminating redundant features. Predominant correlation uses the concept of symmetrical uncertainty to select the feature subset. This method effectively handles the feature redundancy resulting in fast selection of a small subset of features [31].

The third method used for feature selection is ReliefF [32]. This algorithm statistically assigns weights to each feature by estimating its relevance in terms of how well it can differentiate the data points of same and different classes. The features with higher weights are more important than others. Since this method only ranks the features according to their weights and does not select a subset of features, a user-defined threshold is necessary to produce the final subset. The threshold in our case was calculated by averaging all of the positive weights in the feature-ranked list and selecting only the features with weights equal to or higher than the average threshold value [14].

TABLE IV
SENSOR COMBINATIONS ANALYZED FOR PERFORMANCE COMPARISON

Sensor Combinations	Thigh	Wrist	L5	Chest
1	✕			
2		✕		
3			✕	
4				✕
5	✕	✕		
6	✕		✕	
7	✕			✕
8		✕	✕	
9		✕		✕
10			✕	✕
11	✕	✕	✕	
12	✕	✕		✕
13	✕		✕	✕
14		✕	✕	✕
15	✕	✕	✕	✕

F. Classification and Cross Validation

A support vector machine (SVM) classifier was implemented to analyze the performance of the PAC system using the LibSVM library with RBF kernel [33]. To overcome any bias in the training process, the leave-one-subject-out cross validation procedure was used to split the training and testing datasets. In this way, features from all but one subject were used in the training process while the remaining subject was tested. This process was repeated until all subjects had been tested. The effect of class imbalance was compensated by using the weighted SVM. The classifier weighting was implemented using the process described by Huang *et al.* [34] by setting the weights of the different classes to the inverse ratio of the training classes sizes. In this way, the class with largest samples size will have the lowest weight and the class with lowest data samples will have the highest weight. The weights were calculated using the training samples and the calculation was repeated for each fold. The training and testing samples were normalized using the z-score normalization process. The z-

score parameters (mean, standard deviation) obtained from the normalization of the training data were used to normalize the testing data. The z-score normalization was followed by the feature selection process where the feature selection techniques were implemented only using the training data. This process was repeated across all the iterations (folds) of the cross-validation procedure.

The accuracy measure is not the best metric to evaluate the performance in our dataset because of the unbalanced class sizes. Thus, F-measure was analyzed and is interchangeably used with the term "performance" throughout the remainder of this paper. The expressions to compute the performance metrics are described in Appendix B.

The standard error (SE) is also computed for F-measure and accuracy across each sensor combination as shown in (5).

$$SE = \frac{SD}{\sqrt{16}} \tag{5}$$

where SD is the standard deviation across 16 folds (i.e. the total number of subjects analyzed).

G. Single-Sensor vs Multi-Sensor Solution

One of the objectives of this study was to identify the optimal number of sensors by analyzing the performance of all possible sensors combinations. Therefore, performance of 15 sensor combinations listed in Table IV were analyzed and compared.

H. Computational Complexity Analysis

Computational complexity was also evaluated consisting of two measures; 1) feature extraction time - the total time required to extract (compute or calculate) the features; and, 2) classifier testing time - the total time it takes to test the classifier. This process was completed for both categories: the whole feature set (without feature selection) and the subsets obtained from all feature selection approaches. Our earlier work [25] reported the total classification time by computing the classifier training time and testing time. However, the current work presents only the classifier testing time (excluding the classifier training time) as this can give a better idea of how much time is needed by the system to classify an instance in real-life conditions.

All feature selection methods were implemented in MATLAB (Release 2014b, The Math Works, Inc., Natick, MA, USA) using the feature selection repository [35]. The SVM classifier was implemented using the LibSVM library [33] for MATLAB. The analysis was performed on a Dell laptop (Model # M3800, Intel® Core™ i7-4712HQ, CPU @2.30Gz, 16GB RAM, 64-bit operating system).

III. RESULTS AND DISCUSSION

A. Performance Analysis of Single-Sensor vs Multi Sensor Solution Using All Features

The results obtained from the performance analysis of all 15 sensor combinations are presented in Table B1 for the F-measure and for accuracy. The F-measure for all single-sensor solutions and the best multi-sensor solutions (with 2, 3, and 4 sensors) are presented in Fig. 3.
For every sensing solution, each of the four columns in Fig. 3 presents the respective performance measure obtained from a

given feature selection approach (i.e. column 1: All features without using any feature selection method, column 2: using CFS, column 3: using FCBF, column 4: using ReliefF) and the values above each column show the associated standard error, as computed in (5).

Among all single-sensor solutions, the best performance was accomplished by the sensor at the lower back (L5), with an F-measure of above 80% using the subset selected by CFS (Fig. 3). Sensors at the chest and the thigh also performed considerably well (above 75%) as compared to the sensor on the wrist, which performed worst among all single-sensors with performance below 60%.

Comparing the best solutions in Fig. 3, we observed a noticeable improvement in the performance of 7.3% from single sensor solution (L5) to two-sensor solution (C, L5). Furthermore, improvement in the performance is almost negligible by increasing number of sensors from two to four. These results are relevant as this suggests that a plateau is reached at a two-sensor solution, beyond which the performance cannot be improved further even by increasing the number of sensors.

B. Comparison with State-of-the-Art Systems

The performance of three representative systems for PAC tested in our earlier work [25] is also presented in Fig. 3 (solid lines) to provide a direct comparison with the newly proposed system. All the three systems by Bao *et al.* [17], Cleland *et al.* [7] and Leutheuser *et al.* [8] were implemented using the same dataset, type of ADLs, windowing approach, and cross validation procedure. The performances (F-measure, Fig. 3) obtained by these systems are: 83.7% (SE: 1.8%) by Leutheuser

et al. which uses three sensors (chest, wrist, L5), 83.3% (SE: 1.4%) by Cleland *et al.* which uses five sensors (chest, L5, wrist, waist, thigh) and 78.4% (SE: 2.8%) by Bao *et al.* which uses three sensors (L5, wrist, thigh).

The performance of our single-sensor based solution at L5 was better (increase of 2.4%) than the system by Bao et al. Furthermore, its performance is also comparable with the systems by Cleland and Leutheuseur with a slight decrease (less than 2%) in the performance. Therefore, these findings show the potential of using our single-sensor-based solution in real-life conditions instead of such multi-sensor solution. Additionally, the performance of our two-sensor system (T+L5) was much better than the state of art systems and still uses less number of sensors than these systems (3 or more).

C. Effect of Feature Selection on System Performance

Three feature selection methods were implemented on the whole feature set and the respective performances obtained from each method have been shown in Fig. 3. The number of features obtained through all single sensor based systems and from the best (in terms of performance) multi-sensor based systems are presented in Table V. These results were computed across 16 folds and the corresponding mean and standard deviation reported for each of the seven systems. The highlighted text in Table V corresponds to the best feature selection method. The type of features selected by the best feature selection method are listed in Appendix (Table B2).

The performance of the single-sensor systems using L5 or chest increased using the CFS method as compared to the performance obtained without using feature selection. This improvement was larger (7.4%) in chest based PAC system and
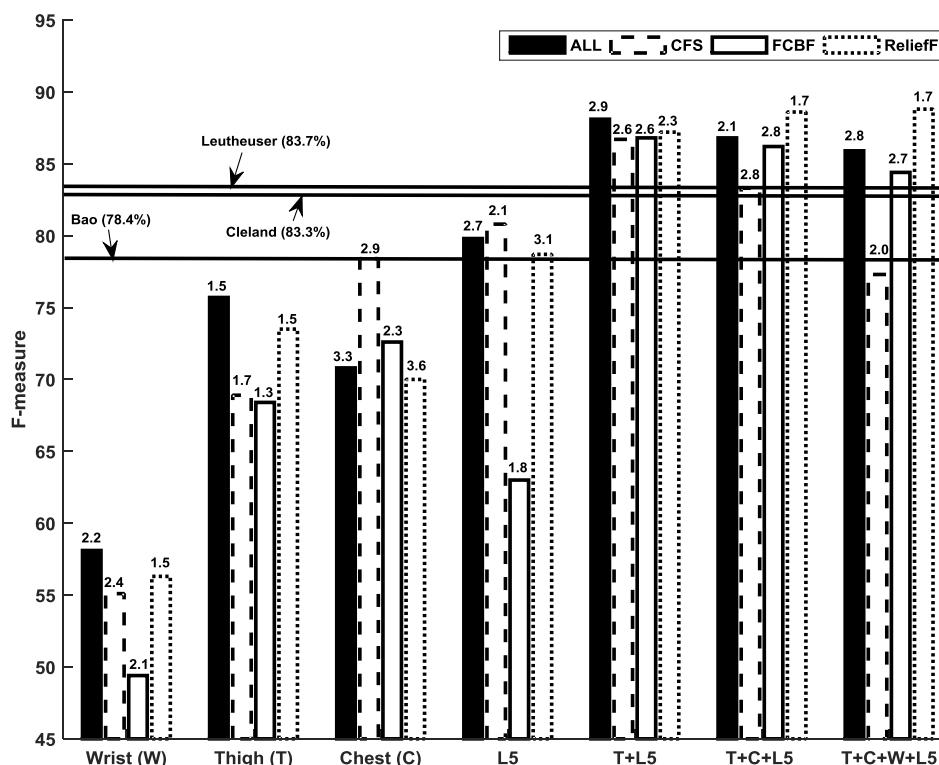

Fig. 3. F-measure analysis using SVM Classifier with and without feature selection methods across various sensors combinations.

smaller (1%) in L5 based PAC system. For wrist and thigh based single-sensor systems, the feature subset of ReliefF performed better than other feature selection approaches but the performance was much lower than the one obtained using all feature set.

On the contrary, the performance using FCBF was the poorest within this dataset using single-sensor solutions. This might be due to the fact that FCBF is an aggressive method of selecting features and selected less features (Table V) as compared to other methods and resulted in losing important features. These findings are in line with the work in [36], where the subset of features chosen by FCBF was smaller than the subset chosen by CFS using single-sensor based system.

For multi-sensors based systems, the feature subset selected by ReliefF performed better than the whole feature set (without feature selection) for two out of three systems (Fig. 3). The improvement in the performance was between 2-3%. The performances of all three feature selection approaches were quite close to each other in multi-sensors based systems (Fig. 3). It is worth noting that there is not a single feature selection method that performed better, both for single-sensor based solutions and multi-sensor solutions.

In addition to the improvement in performance, a substantial decrease in the number of features (above 70%) was observed in both systems i.e. single-sensor and multi-sensor. Reduction in the feature set is quite important since it is directly related to the computational complexity of the system.

In this study, we focused on filter-based methods to select the feature subset by looking at the general characteristics of the data, without involving a specific classifier. In this way, the selected feature subset will be more generalized and can be used to compute and analyze the performance of different classifiers. It is possible that other features selection approaches (wrapper methods, embedded methods) may lead to different results. However, these approaches involve a specific classifier to find the feature subset, which may not be useful to compute the performances of other classifiers.

TABLE V
STATISTICS OF THE FEATURES SELECTED BY THE THREE FEATURE SELECTION
APPROACHES FOR THE SENSOR COMBINATIONS PRESENTED IN FIG.3

No.	Sensors	CFS (mean ± std)*	FCBF (mean ± std)	ReliefF (mean ± std)
1	W	28.9 ± 1.8	2.8 ± 0.8	**22.1 ± 0.3**
2	T	8.7 ± 1.1	5.1±0.9	**12.7 ± 0.7**
3	C	**21.9 ± 1.5**	4.3±0.9	26.3 ± 0.9
4	L5	**17.9 ± 0.7**	4.1±1.1	22.6 ± 0.8
5	T+L5	10.8 ± 0.9	12±1.8	**39.8 ± 0.8**
6	T+C+L5	16.8 ± 1.8	17.6±1.8	**70.9 ± 1.0**
7	T+C+W+L5	19.9 ± 1.4	21±2.5	**104.9 ± 1.2**

* *Mean and standard deviations were obtained from the number of features selected by each of the feature selection algorithm across 16 folds.*

D. Computational Complexity of the System

Computational complexity of the best single-sensor solution was analyzed for a subject (all window instances) and a single window instance (consisting of 5 s or 500 samples) of the same subject. The subject was chosen in such a way that it contained

enough instances of each class (standing: 449 instances; walking: 237 instances; sitting: 1001 instances; lying: 54 instances; resulting into 1741 instances). Computational costs obtained from a single window instance and from all 1741 instances are shown in Fig. 4 and 5 respectively. Such computational costs were estimated as the mean and standard deviation of 10 runs in order to account for computer performance variability.

As expected, the feature extraction (computation) time for single window instance (Fig. 5 (a)) was low in the selected feature subsets compared to the time taken to compute the whole feature set.
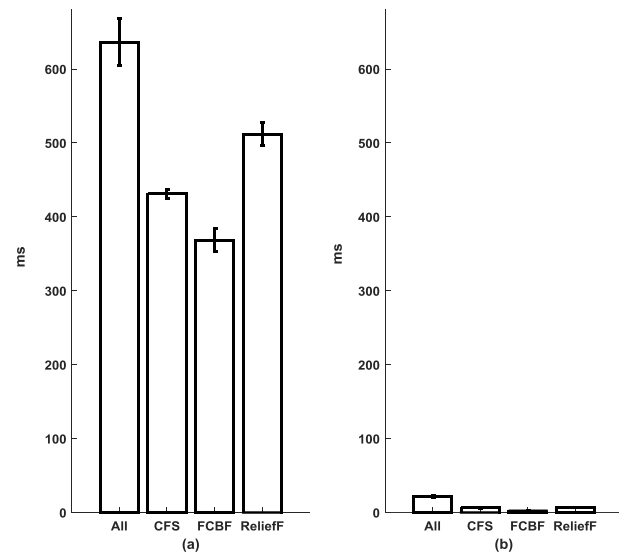
Fig. 4. Computational complexity analysis of single window instance: a) feature extraction time, b) classifier testing.
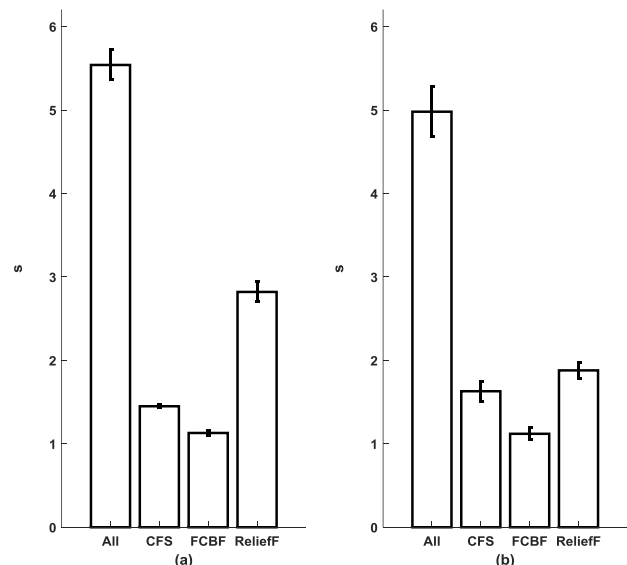
Fig. 5. Computational complexity analysis of single subject's data: a) feature extraction time, b) classifier testing time.

The total number of features for the L5 sensor for the chosen subject are: 68 (no feature selection), 19 (CFS subset), 6 (FCBF subset) and 23 (ReliefF subset). Among the three feature selection methods, the feature subset selected by FCBF took shorter time to extract (compute), a possible reason being the smaller subset of features chosen by FCBF than the other two

subsets. Moreover, the feature extraction time taken by the subset of CFS was smaller than the time taken by the subset of ReliefF. The reason behind this behavior is the lower number of features selected by CFS as compared to ReliefF.

The analysis of classifier testing time shows that the feature selection approaches have improved the time consumption by taking less time to classify the single instance with respect to the whole feature set (Fig. 4 (b), Table B3). Among the three feature subsets, the feature subset of FCBF took less time to classify the instance than the feature subsets of CFS and ReliefF. These results are also coherent showing that larger subset of features takes more time to classify the data instance as compared to the subset with small number of features.

The time taken by the PAC system in real-life conditions is the sum of the feature extraction time and the classifier testing time. Therefore, feature selection can play an important role in reducing the time required to classify any window instance. The overall behavior of computational complexity analysis of a single subject (Fig. 5, Table B4) was quite similar to the one obtained from single window instance. Also in this case feature selection reduces the computational cost of the system. The single subject analysis gives a broader picture of computational complexity which can be helpful in building a personalized (subject-dependent) PAC system for older adults in real-life conditions.

E. Single-Sensor vs Multi-Sensor Solution, What to Choose?

There is no *a priori* definite rule to select a single sensor or multi-sensor solution as it depends on many factors. For instance, the aim of a certain PAC system might not be to achieve an acceptable level of performance for all classes (ADLs) but to obtain high performances for certain ADLs at the expense of others. However, it is important to note that selection of sensor solutions not only depends on the ADLs of interest but also on the environmental conditions. In applications intended for home environments or monitoring of outdoor activities, wearability (in terms of comfort and number of sensors) and battery life can play a different role and raise different concerns than in clinical and laboratory-based settings.

To get more insight, let us consider, as an example, the performances obtained by three sensing solutions: a) chest, b) L5, c) thigh and L5. The respective confusion matrices are presented in Table VI along-with F-measure for each case. In the first solution, a sensor at the chest successfully classified walking and sitting but did not performed well in classifying standing and lying (Fig. 6). The true positives of lying class are quite high but the large number of false positives (432) has reduced the performance. Still, if we are interested in improving the classification of the standing and sitting class, the single-sensor system using L5 is the appropriate choice with an additional improvement in the overall performance (80.8%). Furthermore, the overall performance and the performance of certain classes can be improved by adapting a multi-sensor based solution, i.e. combining the thigh and L5 sensors with performance of 87.2% and a significant improvement in the performance of walking, sitting and standing class (Fig. 6).
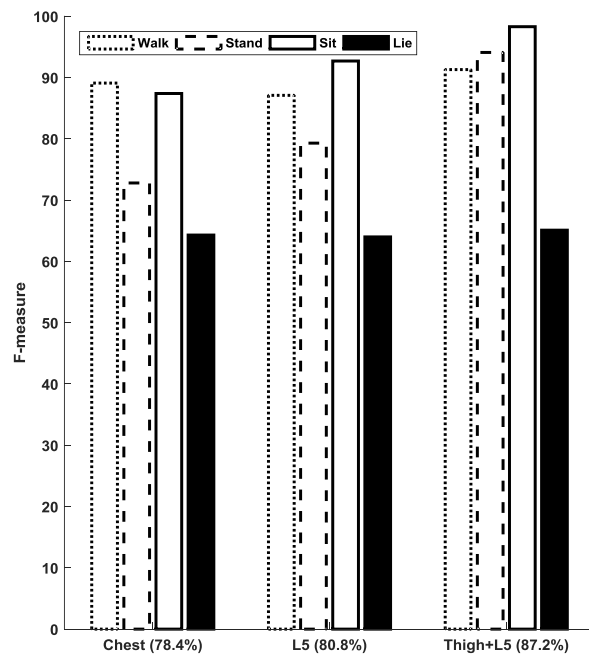

Fig. 6. F-measure by class for three sensing solutions: a) Chest, b) L5, c) Chest, d) Thigh + L5. Value in parenthesis show the averaged F-measure.

TABLE VI
CONFUSION MATRIX USING SVM CLASSIFIER FOR THE SENSORS AT (a) CHEST (b) L5 (c) THIGH + L5

F-measure 78.4%	(a) Chest Sensor (CFS subset)			
	Predicted Class			
	classified as →	walk	stand	sit	lie
Actual Class walk	5796	519	25	0
stand	730	7291	1768	0
sit	110	2421	16534	432
lie	39	0	26	448

F-measure 80.8%	(b) L5 Sensor (CFS subset)			
	Predicted Class			
	classified as →	walk	stand	sit	lie
Actual Class walk	5573	754	13	0
stand	776	7673	1337	3
sit	103	1124	18197	73
lie	0	1	235	277

F-measure 87.2%	(c) Thigh + L5 (ReliefF subset)			
	Predicted Class			
	classified as →	walk	stand	sit	lie
Actual Class walk	5688	498	154	0
stand	421	9151	217	0
sit	5	2	19470	20
lie	0	0	256	257

The performance of lying was not good both for the single-sensor based system and the two-sensor based system, suggesting that the number of samples in the lying class are probably too small for an effective learning by the classifier (even though a weighted version was implemented). Further improvements could be made by combining the classifier with simple algorithms specific to the lying class. For example, to

avoid misclassifications between the lying and the sitting classes, simple algorithms using the chest or L5 sensor and a threshold on the tilt angle could further improve performance. These findings have shown the potential of using various modalities (single-sensor or multi-sensor based solutions) to classify the ADLs of elderly people in free-living conditions. Certainly, there is not a one-fits-all solution that offers a global optimum, regardless specific objectives. Considering the comfort level of the user, sensor at the L5 is the best option to achieve the highest performance among all single-sensor solutions. Moreover, a multi-sensor PAC system may be the desired option to obtain better overall performance as well as performance by class, while compromising the comfort level of user as well as the computational cost of the system.

It must be noted that the obtained results depend on the available data that is used for training the classifiers and the ADLs chosen. The dataset analyzed in this study was collected in free-living conditions. Participants were unsupervised and able to perform their tasks freely thus resulting in unbalanced data samples of ADLs, where certain ADLs (lying) were less frequent than others (sitting, standing). This unbalanced class distribution also creates classification bias when PAC systems are developed using machine learning approach (e.g., if there are few instances of lying it is difficult for the classifier to learn the lying pattern). However, the unbalanced data samples are a true reflection of real world conditions where frequency and act of performing ADLs cannot be controlled and supervised.

To the best of our knowledge, none of the existing activity classification systems developed for older adults using inertial sensors have been fully validated in free-living conditions. The study outcomes suggest the potential benefits of incorporating inertial sensors to monitor the mobility patterns of elderly people in home environments, which can be helpful in determining quality of life and promoting healthy ageing.

IV. CONCLUSION

This study presents a new PAC system that can accurately classify the ADLs of elderly people performed in free-living conditions. The analysis shows very encouraging results, where a single sensor's overall performance is close to that obtained by multiple sensors based state of the art systems, disclosing the potential of using a single sensor for activity classification. In addition, our proposed two-sensor based system improved the system's performance further while still using less sensors than start of the art systems.

Based on presented results a single sensor-based PAC system is highly recommended for real-life conditions when the objective is to have a good overall performance. Some classes may have lower performance than others, but the system would be less computationally complex and more comfortable to wear. On the other hand, the multi-sensor solutions may be recommended when, e.g. designing a surveillance system for fragile older adults, higher performance are desired, even at the cost of reducing the wearability of the system.

The use of feature selection approaches can not only enhance the system's performance but also reduce the computational cost of the system, with the payoff of reducing power consumption and lengthening battery life in real-life conditions.

The main limitation of the current study is the small number of subjects involved. However, the dataset analyzed is among the largest of its kind so far being collected in free-living conditions for elderly people and annotated manually with a very high frequency of 25Hz (annotation every 0.04 s) [26].

APPENDIX

A. Dataset Characteristics

The Table A1 describes the characteristics of the unsupervised free-living protocol. Table A2 represents the inter-rater statistics of video labelling procedure.

TABLE A1
FREE LIVING UNSUPERVISED AND UNSTRUCTURED TASK BASED PROTOCOL [26]

Free-Living Protocol
Sit at a table and write a letter/list or read
Sit on an armchair watch TV/video, or read a magazine Sit on a low stool or toilet seat (lid down clothes on, simulation only)
Lie on a bed, clothes on
Get in and out of a car or sit on a bed
Prepare and consume a drink or food while standing
Set a table for dinner or move from one counter to another many times (up to 10) (shuffling)
Simulate unloading a washing machine for 10 s or prepare a fireplace
Pick an object off the floor then replace or tie/untie shoe laces
Ascending/descending stairs or walking up and down an inclined path
Remove clothes from washing machine and hang on clothes rack or remove rubbish from bin and dispose
Sit and prepare and eat something
Clean mirror or clean a window
Wash and dry hands
Sit at a table and read

TABLE A2
INTER-RATER RELIABILITY STATISTICS [26]

Inter-Rater Reliability Statistics	Average	Maximum	Minimum
Category agreement (%)	90.05	93.31	87.93
Cohen's kappa	0.86	0.91	0.83
Corrected kappa	0.89	0.93	0.87
Krippendorff's alpha	0.86	0.91	0.83

B. Computation of the Performance Metrics

This section provides the details regarding the computation of the performance metrics used in this study. The expressions to calculate the accuracy and F-measure are described below:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100 \qquad (A.1)$$

$$Fc - measure = \frac{2 * TPc}{2 * TPc + FPc + FNc} \times 100 \qquad (A.2)$$

where TP= True Positive, TN = True Negative, FN = False Negative, FP = False Positive. The subscript "c" is used with TP, TN, etc., to represent the metrics by class.

TABLE B1
PERFORMANCE ANALYSIS OF MULTIPLE-SENSOR COMBINATIONS

No.	Sensors	F-measure (SE)*				Accuracy (SE)*			
		ALL	CFS	FCBF	ReliefF	ALL	CFS	FCBF	ReliefF
1	T	75.7 (1.5)	68.9 (1.7)	68.4 (1.3)	73.5 (1.5)	92.9 (1.1)	82.7 (2.8)	82.2 (2.8)	91.5 (1.1)
2	W	58.1 (2.2)	55.1 (2.4)	49.4 (2.1)	56.3(1.5)	75.8 (2.5)	71.2 (3.2)	61.1 (2.3)	75.7 (2.3)
3	L5	79.8 (2.7)	80.8 (2.1)	63.0 (1.8)	78.7 (3.1)	88.3 (1.1)	87.8 (0.6)	77.7 (1.0)	85.5 (2.3)
4	C	70.8 (3.3)	78.4 (2.9)	72.6 (2.3)	70.0 (3.6)	81.7 (1.6)	83.2 (1.5)	77.7 (1.6)	79.8 (2.2)
5	T+W	73.0 (1.5)	69.5 (1.2)	68.7 (0.8)	72.1 (0.4)	93.9 (0.6)	87.8 (2.3)	87.0 (2.2)	94.8 (0.5)
6	T+L5	88.1 (2.9)	86.7 (2.6)	86.8 (2.6)	87.2 (2.3)	96.8 (0.5)	95.5 (0.7)	95.4 (0.6)	95.6 (1.1)
7	T+C	83.5 (3.0)	80.1 (2.6)	79.5 (2.5)	81.2 (2.8)	96.0 (0.6)	94.5 (0.7)	94.3 (0.7)	95.6 (0.9)
8	W+L5	82.3 (2.7)	82.5 (2.2)	73.8 (2.8)	81.3 (3.1)	88.0 (2.6)	88.4 (1.6)	82.2 (1.7)	87.2 (2.7)
9	W+C	72.8 (3.2)	78.3 (3.1)	74.9 (2.7)	73.5 (2.9)	84.3 (2.0)	84.6 (1.8)	79.8 (1.7)	84.3 (1.8)
10	L5+C	83.2 (2.5)	80.1 (2.0)	68.4 (3.1)	79.7 (3.7)	89.0 (1.4)	88.2 (1.1)	79.2 (1.6)	86.5 (2.3)
11	T+W+L5	87.8 (2.6)	81.6 (2.8)	84.3 (2.6)	87.9 (2.3)	96.2 (0.5)	95.3 (0.7)	95.3 (0.6)	96.0 (0.8)
12	T+W+C	80.6 (2.8)	71.1 (1.1)	73.4 (1.5)	81.8 (2.8)	95.4 (0.6)	92.8 (1.6)	94.1 (0.6)	95.8 (0.7)
13	T+C+L5	86.8 (2.1)	83.3 (2.8)	86.2 (2.8)	88.6 (1.7)	96.5 (0.5)	95.3 (0.6)	95.3 (0.6)	96.1 (0.7)
14	W+L5+C	83.2 (2.5)	80.7 (2.1)	75.0 (2.7)	82.2 (2.6)	89.6 (1.6)	89.2 (1.3)	82.6 (1.5)	89.4 (1.5)
15	T+W+L5+C	85.9 (2.8)	77.3 (2.0)	84.4 (2.7)	88.8 (1.7)	96.1 (0.5)	95.2 (0.6)	95.2 (0.6)	96.4 (0.6)

Accuracy, Standard Error (SE) and F-measure are in percentage, Highlighted rows are the seven combinations presented in Fig. 3.

TABLE B2
BEST FEATURE SELECTION METHOD FOR THE SENSOR COMBINATIONS PRESENTED IN FIG.3

Sensors	Best FS* (# features)	Selected Features (as per Table IV)
W	ReliefF (22)	1, 2, 3, 7, 8, 9, 13,14,15,16,17, 27, 28, 29, 33,43,44,45,49,59,60,61
T	ReliefF (11)	1, 2, 3, 8, 9, 13, 14, 15, 16, 43, 49
C	CFS† (11)	3, 10, 39, 41, 42, 45, 47, 48, 49, 54, 55
L5	CFS† (12)	2, 10, 15, 18, 29, 42, 45, 47, 48, 49, 54, 55
T+L5	ReliefF (37)	1, 2, 3, 7, 8, 9, 13, 14, 15, 16, 17, 18, 43, 44, 45, 49, 59 **(T)**, 1, 2, 3, 7, 9, 13, 14, 15 16, 18, 29, 43, 44, 45, 49, 61 **(L5),** 73, 75, 76, 77 **(T, L5)**
T+C+L5	ReliefF † (66)	1, 2, 3, 7, 8, 9, 13, 14, 15, 16, 17, 18, 28, 43, 44, 45, 49, 59, 61 **(T)**, 2, 3, 9, 13, 15, 16, 17, 18, 29, 43, 45, 49, 61 **(L5)**, 1, 2, 3, 9, 13, 14, 15, 16, 17, 18, 29, 43, 44, 45, 49, 61 **(C)**, 72,73,74 **(T, C)**, 72,73,74 **(T, L)**, 72,73,74 **(C, L)**, 75,76,77 **(T, C)**, 75,76,77 **(T, L)**, 75,76,77 **(C, L)**
T+C+W+L5	ReliefF† (97)	1, 2, 3, 7, 8, 9, 13, 14, 15, 16, 43, 44,45, 60, 61 **(W)**, 1, 2, 3, 6, 7, 8, 9, 13, 14, 15, 16, 17, 18, 28, 29, 33, 41, 43, 44, 45, 47, 49, 50, 59, 61, 65 **(T),** 1, 2, 3, 9, 13, 14, 15, 16, 17, 18, 29, 33, 43, 44, 45, 49, 61 **(C)**, 1, 2, 3, 7, 9, 13, 14, 15, 16, 17, 18, 29, 33, 43, 44, 45, 49, 60, 61, **(L5)** 73, 74 **(W, T)**, 72,73,74 **(T, C)**, 72,73,74 **(T, L)**, 72,73,74 **(C, L)**, 75,76,77 **(T, C)**, 75,76,77 **(T, L)**, 75,76,77 **(C, L)**

shows the best feature selection method based on performance. Values inside the parenthesis (Column 2) shows the number of features that were common across 16 folds of the respective feature selection method.

† *show that the performance of the feature selection method is higher than the performance using all feature-set without using feature selection.*

For instance, if we are interested in calculating the performance metrics for standing class using sensor at L5 (Table V (b)): $TP_c = 7673$, $FN_c = 2116$, $FP_c = 1879$, $TN_c = 24471$.

$$Accuracy = \frac{9468 + 5939 + 19494 + 1}{36139 \ (sum \ of \ all)} \times 100 = 96.6\%$$

$$Fc - measure = \frac{2 * 7673}{2 * 7673 + 1879 + 2116} \times 100 = 79.3\%$$

The F-measures and accuracy values obtained from 15 sensor combinations are presented in Table B1 with and without feature selection approaches. The F-score and accuracy are computed from the confusion matrix as presented in Table VI. The standard error is computed from the metrics obtained through all the 16 folds.

The number and type of features selected by the best feature selection method are listed in Table A2. Table B3 and B4 present the computational complexity analysis of single window instance and single subject, respectively.

TABLE B3
COMPUTATIONAL COMPLEXITY ANALYSIS OF SINGLE WINDOW INSTANCE

Measure	All Features	CFS	FCBF	ReliefF
feature extraction time (msec)	636.40 ±31.96	430.90 ±5.44	368.64 ±15.47	511.88 ±15.43
classifier testing time (msec)	21.92 ±1.14	6.15 ±0.19	2.36 ±0.13	6.65 ±0.28

TABLE B4
COMPUTATIONAL COMPLEXITY ANALYSIS FOR SINGLE SUBJECT

Measure	All Features	CFS	FCBF	ReliefF
feature extraction time (sec)	5.54 ±0.18	1.45±0.02	1.13±0.03	2.82±0.12
classifier testing time (sec)	4.98±0.30	1.63±0.12	1.12±0.07	1.88±0.10

REFERENCES

[1] WHO for Europe. (2017, March 20), *Policies and Priority Interventions for Healthy Ageing* [online]. Available: http://www.euro.who.int/__data/assets/pdf_file/0006/161637/WHD-Policies-and-Priority-Interventions-for-Healthy-Ageing.pdf?ua=1.

[2] W. J. C. Zajko *et al.*, " Exercise and physical activity for older adults," *Med. Sci. Sports Exerc,* vol. 41, pp. 1510-30, 2009.

[3] WHO. (2017, March 20). *Global Recommendations on Physical Activity for Health* [online]. Available: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf.

[4] K. Altun *et al.*, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, vol. 43, pp. 3605-3620, 2010.

[5] A. M. Khan *et al.,* "A triaxial accelerometer based physical activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inform. Technol. in Biomed.,* vol. 14, no.5, pp. 1166-1172, Sep. 2010.

[6] R. Ganea, A. Paraschiv-Ionescu, and K. Aminian, "Detection and classification of postural transitions in real-world conditions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 688-696, Sep. 2012.

[7] I. Cleland *et al.*, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors*, vol. 13, pp. 9183-9200, 2013.

[8] H. Leutheuser, D. Schuldhaus, B. M. Eskofier., "Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset," *PloS one,* vol. 8, pp. e75196, 2013.

[9] A. Mannini *et al.*, " Activity recognition using a single accelerometer placed at the wrist or ankle," *Med. Sci. Sports Exerc.,*Vol. 45, pp. 2193–2203, 2013.

[10] M. Arif *et al.*, "Better physical activity classification using smartphone acceleration sensor," *J. Med. Syst.*, vol 38, pp. 1–10, 2014.

[11] M.B.D. Rosario *et al.*, "A comparison of activity classification in younger and older cohorts using a smartphone," *Physiol. Meas.,* vol. 35, pp. 2269, 2014.

[12] L. Gao, A. Bourke, and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Med. Eng. Phys.*, vol. 36, pp. 779-785, 2014.

[13] J. J. Guiry, P. V. D. Ven, J. Nelson, L. Warmerdam, and H. Riper, "Activity recognition with smartphone support," *Med. Eng. Phys.*, vol. 36, pp. 670-675, 2014.

[14] A. M. Torres *et al.*, " Activity classification based on inertial and barometric pressure sensors at different anatomical locations". *Physiol. Meas.* vol. 35, pp. 1245–1263, 2014.

[15] B. Fida, I. Bernabucci, D. Bibbo, S. Conforto, and M. Schmid, "Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer," *Med. Eng. Phys,* vol. 37, pp. 705-711, 2015.

[16] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors," *Sensors,* vol. 16, p. 426, 2016.

[17] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, Berlin/Heidelberg, Germany: Springer-Verlag, 2004, pp. 158-175.

[18] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," i*n Proc. 20th Nat. Conf. Artif. Intell.,* 2005, pp. 1541-1546.

[19] S. J. Preece, J. Y. Goulermas, L. P. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Trans. Biomed. Eng.,* vol. 56, no. 3, pp. 871-879, Mar. 2009.

[20] D. Trabelsi *et al.*, "An unsupervised approach for automatic activity recognition based on hidden Markov model regression," *IEEE Trans. Autom. Sci. Eng.,* vol. 10, no. 3, pp. 829-835, July 2013.

[21] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly," *IEEE Trans. Bio-Med. Eng.*, vol. 50, no. 6, pp. 711-723, Jun. 2003.

[22] G. Lyons, K. Culhane, D. Hilton, P. Grace, and D. Lyons, "A description of an accelerometer-based mobility monitoring technique," *Med. Eng. Phys,* vol. 27, pp. 497-504, 2005.

[23] A. Godfrey, A. Bourke, G. Olaighin, P. V. D. Ven, and J. Nelson, "Activity classification using a single chest mounted tri-axial accelerometer," *Med. Eng. Phys,* vol. 33, pp. 1127-1135, 2011.

[24] M. Awais, S. Mellone, and L. Chiari, "Physical activity classification meets daily life: Review on existing methodologies and open challenges," *in Proc. 37th Annu. IEEE Int. Conf. Eng. Med. Biol. Soc.,* 2015, pp. 5050-5053.

[25] M. Awais, L. Palmerini, A. K. Bourke, E. A. Ihlen, J. L. Helbostad, and L. Chiari, "Performance Evaluation of State of the Art Systems for Physical Activity Classification of Older Subjects Using Inertial Sensors in a Real Life Scenario: A Benchmark Study," *Sensors,* vol. 16, pp. 2105, 2016.

[26] A. K. Bourke, E. A. F. Ihlen, R. Bergquist, P. B. Wik, B. Vereijken, and J. L. Helbostad, "A Physical Activity Reference Data-Set Recorded from Older Adults Using Body-Worn Inertial Sensors and Video Technology—The ADAPT Study Data-Set," *Sensors,* vol. 17, pp. 559, 2017.

[27] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-

time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. Inf. Technol. in Biomed.,* vol. 10, no. 1, pp. 156-167, Jan. 2006.

[28] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Pers. Ubiquitous Comput.,* vol. 14, pp. 645-662, 2010.

[29] C. Mazzà, M. Iosa, P. Picerno, and A. Cappozzo, "Gender differences in the control of the upper body accelerations during level walking," *Gait and Posture,* vol. 29, pp. 300-303, 2009.

[30] M. A. Hall and L. A. Smith, "Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper," *in Proc. 12th Int. FLAIRS Conf.,* 1999, pp. 235-239.

[31] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc 20th Int. Conf. Mach. Learning,* 2003, pp. 856-863.

[32] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. 9th Int. Workshop Mac. Learning*, 1992, pp. 249-256.

[33] C.C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Iintell. Syst. Technology (TIST),* vol. 2, pp. 27, 2011.

[34] Y. M. Huang and S.X. Du, "Weighted support vector machine for classification with uneven training class sizes," *in Proc. Int. Conf. Mach. Learning Cybern.,* 2005, pp. 4365-4369.

[35] Z. Zhao *et al.* "Advancing feature selection research asu feature selection repository". Technical Rep., 2013.

[36] N. A. Capela, E. D. Lemaire, and N. Baddour, "Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients," *PLoS One,* vol. 10, pp. e0124414, 2015.

Muhammad Awais received the B.S. electronic engineering degree from Mohammad Ali Jinnah University, Islamabad, Pakistan and the M.S. degree in electrical and electronic engineering from Universiti Teknologi PETRONAS, Malaysia. He is currently pursuing the Ph.D. degree with the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Italy. His research interests include biomedical signal processing, wearable sensors, data mining and the application of machine learning to biomedical signals in the field of human movement analysis.

Lorenzo Chiari, PhD, is a Full Professor of Biomedical Engineering at the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Italy. He is also the director of the Health Sciences and Technologies Interdepartmental Center for Industrial Research and leads the eHealth Action Group at the University of Bologna. His main research activity is in the area of digital biomarkers, active and healthy ageing, fall prevention, neurobiomechanics of posture and movement in physiological and pathological conditions, rehabilitation technologies.

Luca Palmerini, PhD, is a Postdoctoral Researcher in Biomedical Engineering at the Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna. His main research interests include machine learning applications to biomedical data, quantification of motor impairments in Parkinson's disease, fall risk assessment, fall detection, active and healthy ageing, and wearable sensors.