Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biomedical, Electrical and Systems Engineering

Ciclo 30th

Settore Concorsuale: 09/G2-BIOINGEGNERIA

Settore Scientifico Disciplinare: ING-INF/06-BIOINGEGNERIA ELETTRONICA E INFORMATICA

PHYSICAL ACTIVITY CLASSIFICATION MEETING DAILY LIFE CONDITIONS FOR OLDER SUBJECTS

Presentata da: **Muhammad Awais**

Coordinatore Dottorato **Supervisore**

Prof. Daniele Vigo **Prof. Lorenzo Chiari**

Esame finale anno 2018

ABSTRACT

Physical inactivity can lead to several age-related issues such as falls, movement disorders and loss of independence in older adults. Therefore, promoting physical activity in daily life and tracking daily life activities are essential components for healthy aging and wellbeing. Recent advances in the MEMS devices make it happen to wirelessly integrate miniature motion capturing devices and use them in personal health care and physical activity monitoring systems in daily life conditions. Consequently, various systems have been developed to classify the activities of daily living. However, the scope and implementation of such systems are limited to laboratory-based investigations and they are mainly developed utilizing the sample population of younger adults. Therefore, this dissertation aims to develop innovative solutions for physical activity classification, with a specific focus on the elderly population in free-living conditions.

Firstly, we present an overview of the state of the art methodologies for physical activity classification. Then, we propose a fair and unbiased benchmark for the field-based validation of the existing state of the art systems for physical activity classification on the older-adults dataset. This benchmark study is particularly relevant since the existing systems for physical activity classification were developed mainly on younger adults' data in a laboratory-based environment. Furthermore, these systems are not directly comparable, due to the large diversity in their design (e.g., number of sensors, placement of sensors, data collection environments, data processing techniques, features set, classifiers, cross-validation methods). The finding concludes that the systems developed in controlled settings are not capable of performing well in real-life conditions where the activities are performed more naturally. Therefore, the newly-developed systems should be trained and tested on the dataset collected in the real-life conditions.

Secondly, we propose a wearable sensor-based physical activity classification system for older adults in free-living conditions as a continuity of our previous findings. We explore four sensor locations (thigh, lower back, chest, and wrist) 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 are implemented on the feature set obtained from the acceleration and angular velocity signals to classify the activities of daily living in free-living conditions. The findings show the potential of different solutions (single-sensor or multi-sensor) to correctly classify the activities of older people in free living conditions. Considering a minimal set-up of a single sensor, the sensor worn at the lower back achieved the best performance. A two-sensor solution (lower back and thigh) achieved a better performance with respect to a single-sensor solution. Then, we present a physical activity classification system to predict unlabeled activities of daily living. This objective is accomplished by training the single sensor based

system on the labeled dataset and testing it on the unlabeled dataset of older adults in free-living conditions.

Finally, we report on feasibility study aimed at developing a video-based method to automatically label the activities of daily living without the help of observers/raters. This system could be utilized alone or in combination with a wearable sensor-based physical activity classification system to validate its performance.

ABSTRACT (IN ITALIAN)

Nella popolazione anziana l'inattività fisica può portare a diverse affezioni legate all'età, quali cadute, disturbi motori e perdita di indipendenza. Pertanto la promozione e il monitoraggio dell'attività fisica nella vita quotidiana sono componenti essenziali per perseguire alti livelli di salute e benessere durante l'invecchiamento. Gli avanzamenti recenti raggiunti sui dispositivi MEMS rendono possibile l'integrazione wireless di sensori miniaturizzati di movimento e il loro impiego in sistemi di monitoraggio della salute personale e dell'attività fisica, in condizioni di vita quotidiana. Ciononostante, l'ambito di impiego di questi sistemi è limitato a ricerche di laboratorio e questi sistemi sono principalmente sviluppati utilizzando un campione di popolazione di adulti più giovani. Perciò scopo di questa dissertazione è sviluppare soluzioni innovative per la classificazione dell'attività fisica, con particolare riguardo alla popolazione anziana in ambiente non supervisionato.

In primo luogo, presentiamo una panoramica dello stato dell'arte delle metodologie per la classificazione dell'attività fisica. Quindi proponiamo un benchmark imparziale per la validazione basata sul campo per i sistemi esistenti in letteratura di classificazione di attività fisica su basi di dati di anziani. Questo studio di benchmark è particolarmente rilevante perché i sistemi esistenti di classificazione dell'attività fisica sono stati sviluppati principalmente su dati di giovani adulti in ambiente di laboratorio. Inoltre questi sistemi non sono direttamente confrontabili, avendo caratteristiche progettuali molto diverse (ad esempio numero e posizionamento di sensori, ambiente di raccolta dati, tecniche di elaborazione dati, insieme dei parametri, classificatori, metodi di validazione incrociata). Dai risultati si conclude che i sistemi sviluppati in ambienti controllati non sono capaci di avere buone prestazioni in condizioni di vita reale, dove le attività della vita quotidiana si svolgono più naturalmente.

In secondo luogo, a seguito di questi nostri risultati, proponiamo un sistema di classificazione di attività fisica basato su sensori indossabili per anziani in ambienti non supervisionati. Esploriamo quattro configurazioni di posizionamento per i sensori (coscia, schiena lombare, torace e polso) al fine di ottenere il numero e la combinazione ottimale di sensori come miglior compromesso tra prestazione e indossabilità del sistema. Sono state implementate diverse tecniche di selezione di parametri sull'insieme dei parametri ottenuti dai segnali di accelerazione e velocità angolare per la classificazione delle attività della vita quotidiana in ambiente non supervisionato. I risultati mostrano il potenziale di diverse soluzioni (a uno o più sensori) per classificare correttamente le attività di anziani in ambiente non supervisionato. Considerando una configurazione minima a sensore singolo, la migliore prestazione si è ottenuta col

sensore indossato all’altezza della quinta vertebra lombare (L5). Una configurazione a due sensori (L5 e coscia) ha raggiunto prestazioni migliori rispetto a quella a singolo sensore. Quindi presentiamo un sistema di classificazione dell’attività fisica per predire attività della vita quotidiana non verificate. Questo obiettivo è raggiunto addestrando il sistema a singolo sensore su un dataset verificato e testandolo su un altro dataset non verificato di anziani in ambiente non supervisionato.

Infine trattiamo di uno studio di fattibilità volto a sviluppare un metodo basato su video per verificare automaticamente le attività della vita quotidiana senza l’aiuto di osservatori/valutatori. Questo sistema potrebbe essere utilizzato da solo o in combinazione con un sistema di classificazione di attività fisica basato su sensori indossabili per validare la sua prestazione.

ACKNOWLEDGEMENTS

In the name of Almighty Allah, the most gracious the most merciful who gave me courage and strength to finish my project successfully. My prayers are upon his Messenger Prophet Mohammad (peace be upon him) who is our role model in every business of life and who urges to seek knowledge from cot to the grave. My foremost expression of submission and gratitude goes to Allah Almighty without whose grace and mercy; this work would never be possible even in the slightest.

I would like to express my sincere gratitude to my supervisor, Prof. Lorenzo Chiari, for giving me the opportunity to commence my PhD studies. It is his trust and guidance that always motivated me in shaping my research and also his encouragement and support during difficult times. I would also extend my appreciation to Luca Palmerini and Sabato Mellone for their guidance and support during my PhD. I am also obliged and thankful to Prof. Jorrun L. Helbostad for her guidance and advice during my period abroad at Norwegian University of Science and Technology (NTNU) Norway. I am also thankful to all the current and former members of my lab for their kind support and for providing a pleasant environment.

No acknowledgment would ever adequately express my obligation to my family for their endless support, love, prayers and good wishes to see me prospering. These are their sacrifices and encouragements for my better education and career which make me what I am today.

Finally, I would like to thank my friends Ehtasham Raja, Mohsin Raza, Junaid Ahmed, Usman Hadi and all others for their encouragement and cooperation to complete my research.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Some contents of this chapter are taken from [1], where we presented the review on existing methodologies for physical activity classification.

1.1 Introduction

Physical activity (PA) is one of the fundamental functionalities of human beings, and it is strongly linked with their physical and mental health. It is one of the key predictors of healthy ageing and well-being. Ageing is an extensive area of research due to the increase in the elderly population. A study conducted by the European Commission in 2012 [2] shows that, in Europe, the elderly population (above 65) is expected to increase from 87.5 million to 152.6 million during the period from 2010 to 2060. Therefore, healthcare systems need to be adaptive and robust to promote quality of life and active lifestyles for the growing elderly population. A report by the World Health Organization (WHO) in the 28 member states of the European Union (EU) suggests that physical inactivity in the elderly population correlates with a higher risk of falling, mobility disorders, low muscle strength and loss of independence [3]. It also shows that the proportion of falls per year is 30% among the elderly, which increases to 50% in those aged above 80 [3]. Adopting an active lifestyle can significantly minimize the development of many disabling conditions and chronic diseases [4]. The WHO recommends older adults to perform a moderate-to-intense physical activity for at least 30 minutes, five times per week in bouts (bout: uninterrupted period of any specific activity being considered) not shorter than 10 minutes, to achieve the health benefits [5]. Therefore, profiling the activities of daily living (ADLs) could provide better knowledge in designing the intervention to prevent inactivity and to improve health and functional capacity to achieve healthy aging and well-being.

Recent advances in the microelectromechanical systems (MEMS) has encouraged researchers and scientists to make use of IMU (inertial measurement unit) sensors in personal health care systems. This is mainly due to their low power consumption, lightweight, miniaturization, wearability, cost-effectiveness, and reliable data transfer capabilities [6]. A typical IMU sensor is composed of a tri-axial accelerometer and tri-axial gyroscope. The accelerometer measures linear acceleration while a gyroscope measures angular velocity. Hence, by utilizing the inertial sensor technology in conjunction with an appropriate signal processing algorithm aimed at the authentic recognition of activities, one can classify the ADLs in a laboratory-based environment for short-term recordings but also for long-term recordings in clinical and/or in free-living conditions.

1.2 Knowledge Gaps and Challenges in Existing Systems for Physical Activity Classification

A substantial amount of work is available in the literature regarding the development of physical activity classification (PAC) systems using inertial sensors [7-18]. However, there are knowledge gaps and open challenges which needs to be addressed in the current systems to make them suitable for usage in free-living conditions. A general overview of the common approach found in existing PAC system developed in laboratory-based environments and the challenges in implementing them in free-living conditions are illustrated in Figure 1. These are discussed in the following subsections.

1.2.1 Laboratory Controlled Environment Vs Free-Living Conditions

Majority of the existing PAC systems described in the literature are either developed in laboratory-based environments with predefined and structured ADLs [10, 15-18] or designed in a simulated real-world environment with predefined ADLs performed in indoor and outdoor environments [7-9, 11, 12, 14]. However, none of the existing systems, to the best of the authors knowledge, has been extensively validated in free-living conditions where the subjects are not instructed to perform ADLs in a predefined and structured manner. Furthermore, the ADLs performed in laboratory settings are more likely influenced by the experimenter and the surrounding environment which biases the true nature of performing ADLs in free-living conditions.

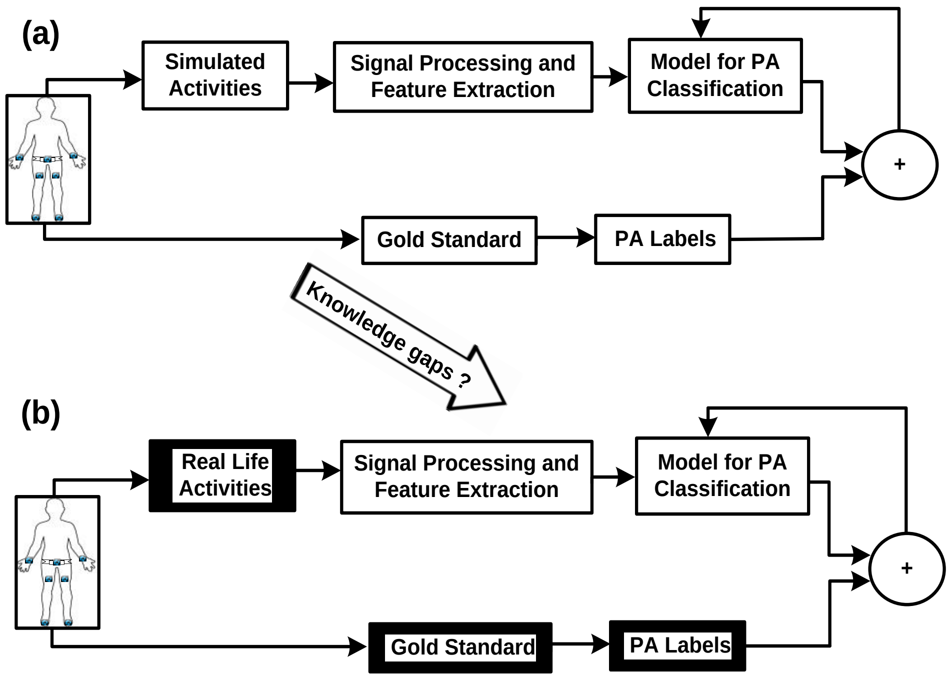


Figure 1.1: (a) PAC systems for controlled laboratory environment (b) PAC system in free-living conditions [1] and relevant knowledge gaps.

1.2.2 Gold Standard/Validation Procedure

Another issue highlighted in Figure 1 (with bold rectangle) is the gold standard or validation procedure for ADLs. The unavailability of a common gold standard is a critical issue in the literature. There are mainly two procedures used to mark the ADLs. i.e., structured/predefined protocol and observational methods. In the structured protocol, participants have to follow a certain protocol, and the sequence, type, and act of performing ADLs are predefined [10, 13, 16, 17]. In observational methods, the ground truth information is collected either by the help of video recording captured during the experimentation process or with observer [8, 11, 12, 18]. These videos are later processed offline by the trained/expert raters to classify/categorize the ADLs. Both of those methods have drawbacks. The structured protocol might work well in laboratory-controlled environments, but it is not applicable in real life conditions where ADL patterns are unforeseeable and are performed more naturally. On the contrary, observational methods are quite accurate in marking ADLs and can provide an excellent starting point in validating the PAC system developed in free-living conditions. However, these methods are hard to adopt as permanent validations procedure due to (i) privacy issues raised during the video capturing in real life conditions (ii) the human resources required to mark the ADLs in larger population [19, 20] (iii) observer bias during activity labelling. An attempt to counteract the observer bias requires that multiple independent raters do the marking offline.

1.2.3 PA Labels or Classes

There is no standardized way to define ADLs, which is another challenge encountered during the development of PAC systems. There are a variety of ADLs performed in free-living conditions such as: sitting, standing, walking, lying, stairs up, stairs down, shuffling, leaning, running, etc. Systems developed so far for PAC does not specify the definition of each class of activity and use the self-defined definitions, which make their performance ambiguous and incomparable. This is an important issue since a particular activity defined in one study can be marked with a different activity label in another study. For instance, walking can be easily confused with shuffling. Furthermore, each activity has certain aspects that can be considered while labeling e.g., in case of walking: how many steps can define walking, if the distance is covered in a straight path or in curved path, etc. These kinds of issues are relevant for each other ADLs mentioned earlier. The good practice would suggest defining the ADLs within the study not only to inform others or to allow replicability of the methods, but also to create homogeneity in the activity definition process.

1.2.4 Diversities in the Design Process of PAC Systems

The performance capabilities of existing PAC systems depend on many factors; dataset (sample population, type of ADLs performed, etc.), number of sensors, placement of sensors, feature-set, window-size for features computation and classifiers. Each of these factors contributes directly to the overall performance of the PAC system. Large diversities in the design process due to the aforementioned factors make the existing PAC systems largely incomparable. Furthermore, most of these systems are developed using the datasets of younger adults [10, 12-17, 21-24] and very few of them are developed using older adults [25-29]. Furthermore, the few systems developed on the older adults are not fully validated in free-living conditions.

1.3 Aims of the Thesis

The main aim of this thesis is to develop an innovative solution for PAC, particularly for the elderly population observed in free-living conditions. For this purpose, quantitative analysis of the wearable sensors signals (specifically accelerometers and gyroscope) is performed. This step is accomplished by using data mining techniques, which discover and interpret the relevant patterns obtained by processing the sensors' data and selecting only the relevant information to classify the ADLs objectively. The specific objectives of the thesis are listed below.

1. To provide a benchmark approach to compare the performance of the state of the art systems for physical activity classification (PAC) in a fair and unbiased way using a novel dataset collected from older adults in free-living conditions.

2. To develop a PAC System for older adults with an optimum number of sensors validated in ecological conditions, informed by the strengths and weaknesses of the current PAC systems.

3. To evaluate the feasibility of predicting the unlabeled activities of older adults in free-living conditions using a single sensor based PAC system.

1.4 Thesis Outline

This thesis is organized into six additional chapters, and its comprehensive overview is depicted in Figure 1.2.

- **Chapter 2** provides a detailed literature review of the state of the art methodologies (SOA) for PAC. It presents a concise overview of the PAC systems regarding the type of ADLs classified, data collection environment, sample population, type and number of sensors, feature-set, validation method and classification approaches used so far.

- **Chapter 3** describes the proposed benchmark study to evaluate the performance of SOA systems for PAC in laboratory-based environments as well as in free-living conditions. The data collection procedure is also explained briefly for the semi-structured protocol in the laboratory-based environment and for the unsupervised protocol in free-living conditions for older adults. Finally, the chapter highlights gaps and limitations within SOA systems when tested in free-living conditions and provides possible future directions to improve performance.

- **Chapter 4** develops the PAC system for older adults in free-living conditions, informed by the limitations and gaps highlighted in chapter 3. It explains various stages of the proposed PAC system, i.e., data processing, feature-extraction, feature selection, computational complexity analysis, classification model development and validation, and the performance analysis of single versus multi-sensors set-up.

- **Chapter 5** presents the PAC system to predict the unlabelled ADLs, performed by older adults during long-term recordings in free-living conditions. The details of the data collection procedure and the sensing devices are also provided. Then, it computes several statistical parameters obtained from the predicted labels to identify if there are patterns that can be associated with participants' lifestyle and general health. It also provides the correlation analysis between the acceleration based measure and the clinical measures.

- **Chapter 6** describes a pilot study to develop a video-based PAC system that can label the ADLs automatically with the help of image processing techniques. It explains the proposed methodology to record the video data and wearable sensors' data simultaneously. Then, it highlights the challenges encountered during the data processing stage. Future suggestions that can be implemented for the successful development of the video-based PAC system are made.

- **Chapter 7** concludes and discusses the overall findings of the thesis. It also highlights the extension of the work that can be the object of future research in order to further advance in the field of activity classification by promoting healthy ageing and well-being.

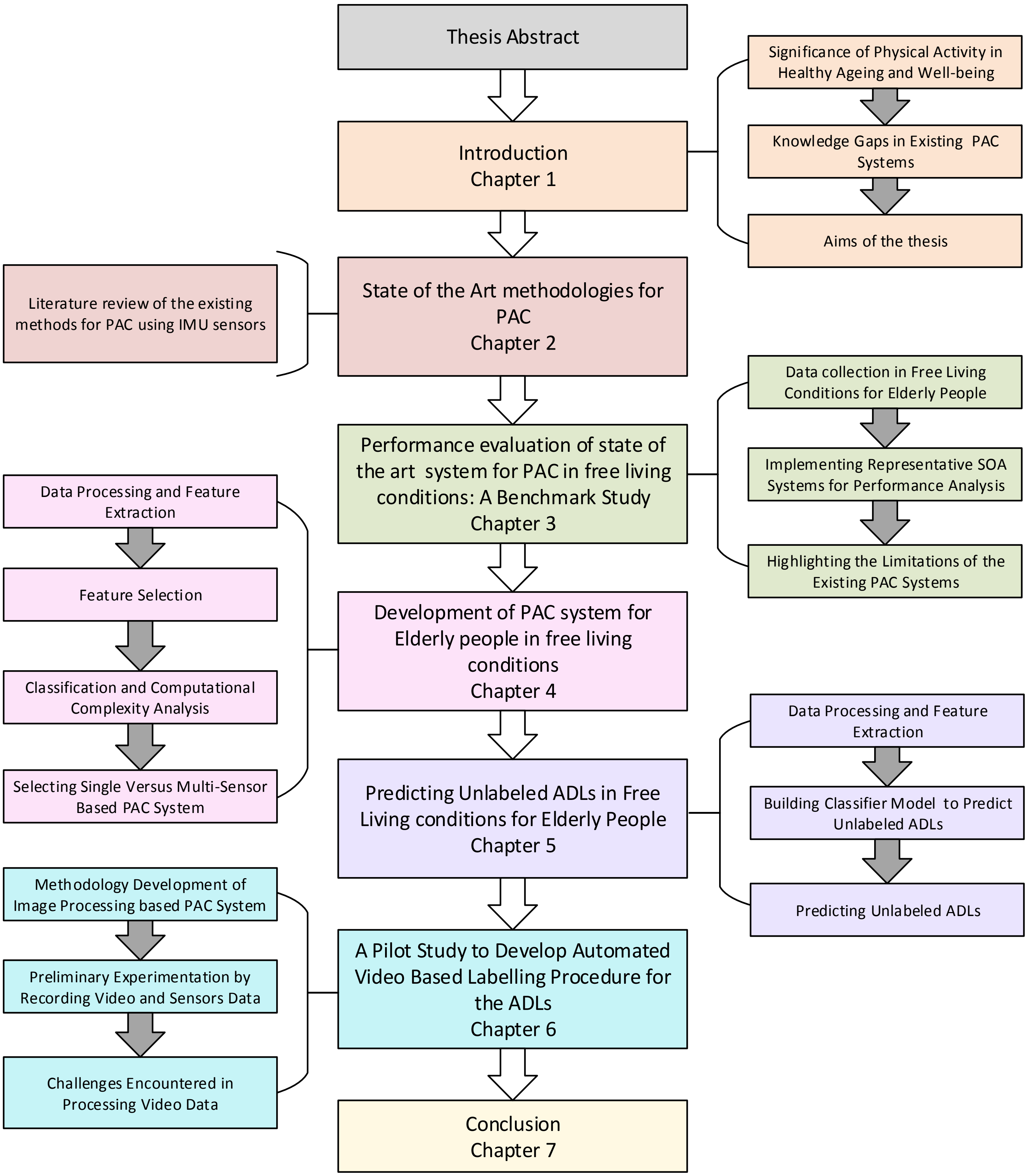


Figure 1.2: Comprehensive overview of the thesis

Chapter 2

State of the Art Methodologies for Physical Activity Classification

Technological advancements in wearable inertial sensors have made them an appealing and unmissable component of health monitoring systems. In particular, various systems for PAC have been developed and are described in the literature. This chapter provides a detailed overview of these systems and reviews the main distinctive factors of PAC systems developed so far.

2.1 Factors Contributing Toward the Overall Performance of a PAC System

A synthetic overview of the state of the art systems for PAC is presented in Tables 2.1 and 2.2. Different systems have been proposed for different target groups; our main focus in this thesis is physical activity performed by older adults. For this reason we present separately state of the art systems developed for young (Table 2.1) and older adults (Table 2.2). It is evident from these tabular representations that several distinctive factors are different across studies and contribute to the overall performance of PAC systems.

Table 2.1: Overview of the PAC systems developed for young adults

Sr. No.	Authors	Data collection protocol	Sample Population (age in years)	Gold standard/ labeling	ADLs Classified	Sensor Placement (no. of sensors; sampling frequency)	Features used (window size, analyzed signals)	Classifier, validation procedure (accuracy)
1	Bao et al [9]	Structured protocol in home environment	20 young adults (21.8 ± 6.59)	Self-labelling by the subject	Walking, sitting, standing, watching TV, running, folding laundry brushing teeth, riding elevator, bicycling, lying, etc.	Hip, wrist, ankle, arm and thigh (5; 76.25 Hz)	mean, energy, frequency domain entropy, correlation (6.7s, **acceleration**)	Decision tree, leave-one-subject-out-cross validation (LOSOCV) (84.0%)
2	Leutheuser et al [14]	Structured protocol in university campus	23 young adults (27 ± 7)	Labelling performed by the experimenter	Sitting, lying, standing, washing dishes, vacuuming, sweeping. Walking, bicycling, ascending / descending stairs,	Wrist, hip, chest, ankle (4; 204.8 Hz)	minimum, maximum mean, variance, spectral centroid, bandwidth, energy (5s, **acceleration**, **angular velocity**)	Hierarchical classification, LOSOCV (89.9%)
3	Cleland et al [22]	Structured protocol in laboratory settings	8 young adults (26.25 ± 2.86)	Predefined sequences of ADLs	Sitting, lying, standing, walking, jogging, ascending /descending stairs,	Chest, lower back, wrist, hip, thigh, foot (6; 51.2 Hz)	mean, standard deviation, skewness, kurtosis, energy and correlation of axes separately and average over 3 axes (10s, **acceleration**)	Support vector machine (SVM) 10-fold cross validation (97.26%)

4	Ravi et al [10]	Structured protocol in laboratory environment	2 adults (NA)	Self-labelling by the subject	Standing, walking, running, ascending /descending stairs, sit-ups, vacuuming, brushing teeth	Pelvic region (1; 50 Hz)	mean, standard deviation, Energy, correlation (5.12 s, **acceleration**)	Plurality voting, 10-fold CV (88.8%)
5	Preece et al [12]	Structured protocol in university campus	20 young adults (31 ± 7)	Offline video labeling	walking, jogging, walking upstairs / downstairs, running, hopping on the left and right leg, jumping	Waist, thigh, ankle (3; 64 Hz)	mean, standard deviation, energy, mean, correlation, entropy, percentile, FFT magnitude, wavelet components (2 s, **acceleration**)	K-nearest neighbor (KNN) LOSOCV (95.0%)
6	Altun et al [13]	Structured protocol in university campus	8 young adults (20-30)	Predefined sequences of ADLs	sitting, standing, lying, walking, cycling, jumping ascending/descending stairs, running etc	Chest, right wrist, left wrist, right leg, left leg (5; 25 Hz)	mean, variance, standard deviation, kurtosis, autocorrelation, DFT (5s, **acceleration, angular velocity, magnetometer signal**)	Bayesian decision making 10-fold CV (99.2%)
7	Fida et al [23]	Structured protocol in university campus	9 young adults (22-34)	Predefined sequences of ADLs	standing, sitting, walking, ascending/descending stairs	Waist (1; 100 Hz)	Mean, standard deviation, skewness, kurtosis, correlation between each axis and magnitude signal (1.5s, **acceleration**)	SVM LOSOCV (90%)
8	Trabelsi et al [15]	Structured protocol in laboratory environment	6 young adults (25-30)	Predefined sequences of ADLs	standing, sitting, transitions, walking ascending/descending stairs,	Chest, right thigh, left ankle (3; 25 Hz)	raw signal (window size is equal to each activity's duration, **acceleration**)	Multiple Hidden Markov Model Regression 10-fold CV (91.4%)

9	Guiry et al [17]	Structured protocol	6 young adults (30.6 ± 6)	Predefined sequences of ADLs	sitting, standing, lying, walking, running, cycling	Chest, thigh (2; 120 Hz)	counts per minute, device angle, DFT (1s, **acceleration**)	Naïve Bayes 10-fold CV (93.0%)
10	Khan et al. [21]	Structured protocol in home environment	6 young adults (mean age of 27)	Self-labelling by the subject	Sitting, standing, lying, transitions, walking, walking upstairs/downstairs, running	Chest (1; 20 Hz)	Autoregressive coefficients, signal magnitude area, tilt angle (3.2s, **acceleration**)	Artificial neural network (ANN) 6-fold CV (97.65 %)
11	Karantonis et al. [30]	Structured protocol in laboratory settings	6 young adults (22-60)	Predefined sequences of ADLs	Sitting, standing, lying, transitions, falls	Waist (1; 45 Hz)	Tilt angle, signal magnitude area, signal vector magnitude (2s, **acceleration**)	Thresholding (90.8%)
12	Hickey et al. [31]	ADLs in free-living conditions	10 young adults (27.5 ± 4.7)	Offline video labeling	Gait analysis	Lower back (1; 100 Hz)	Mean, standard deviation, wavelet (0.1s, **acceleration**)	Intra-class correlation ≥0.941 for walking (N.A.)
13	Torres et al. [18]	Semi-structured protocol in laboratory environment	6 young adults (27.5 ± 4.7)	Offline video labeling	walking, running, stair ascent, stair descent, brushing teeth, drinking, writing, cutting and peeling food.	Chest, right and left wrist, wrist, right and left ankle (5; 50 Hz)	mean, standard deviation, variance, inter quartile range, signal magnitude area, correlation, time and frequency domain kurtosis, entropy, energy, maximum frequency component, RMS value, percentile (2.56s, **acceleration, angular velocity, barometric pressure**)	KNN LOSOCV (95.0%)

14	Mannini et al. [32]	Structured protocol in home environment	13 young adults (21.8 ± 6.59)	Self-labelling by the subject	Sitting, lying, standing, walking, stairs climbing, running, cycling	Hip, wrist, ankle, arm and thigh (5; 76.25 Hz)	DC component, energy, frequency domain entropy, correlation (6.7s, **acceleration**)	Continuous emissions hidden markov model (99.1%)
15	Lee et al. [33]	ADLs in free living conditions	2 young adults (NA)	GPS tracker trajectory	Jogging, walking, sitting, and	Thigh (1; 21 Hz)	Mean, standard deviation, binned range, min-max, peak duration, peak count, mean dominant frequency, mean energy of frequency (10s, **acceleration**)	Random forest Predicted results in free living conditions (95%)
16	Zhang et al. [34]	Structured protocol in outdoor environment	14 young adults (30.1 ± 7.2)	Using sparse representation	Walk forward, walk left, walk right, go upstairs, go downstairs, jump up, run, stand, and sit	right front hip (1; 100 Hz)	Mean, median, standard deviation, variance, RMS, interquartile range, first and second order derivate, skewness, kurtosis, zero and mean crossing rate, correlation, energy, dominant frequency, spectral entropy, movement intensity, signal magnitude area, acceleration correlation, average acceleration energy, average velocity, eigenvalues of dominant direction, average rotation angle and energy (4s, **acceleration, angular velocity**)	Sparse representation LOSOCV (96.1%)

17	Veiga et al. [35]	Structured protocol in laboratory environment	82 young adults (24.68 ± 4.91)	Predefined sequences of ADLs	squat, lunge, deadlift, single-leg squat, and tuck jump data	Lower back (1; 51.2 Hz)	Raw low pass filtered signal of IMU sensor (activity based, **acceleration, angular velocity**)	Convolutional neural networks LOSOCV (95.9%)
18	Ordóñez et al. [36]	Structured protocol in laboratory environment	3 adults (NA)	Offline video labelling	Sitting, standing, walking, lying, closing and opening of the doors, drawer, dishwashers and fridge, cleaning table, drinking, switch toggling. (opportunity dataset [37])	Upper and lower body (17; 30 Hz)	Raw signal of IMU sensors (500ms, **acceleration, angular velocity, magnetometer**)	Deep coevolution long term short memory (DeepConvLSTM) neural networks Manual division for training/testing (93.0%)
19	Hammerla et al. [38]	Structured protocol in laboratory environment	9 young adults (27.22 ± 3.31)	Predefined sequences of ADLs	lie, sit, stand, walk, run, cycle, Nordic walk, iron, vacuum clean, rope jump, ascend and descend stairs (PAMAP dataset [39])	Chest, wrist, ankle (3; 100 Hz)	Raw time series signal and feature selection using machine learning approaches, e.g. restricted Boltzmann machines (5.12s, 33.3 Hz, **acceleration, angular velocity, magnetometer, temperature,** and **heart rate**)	Convolutional neural network Manual division for training/testing (93.7% for PAMAP)
20	Altini et al. [40]	Structured protocol in laboratory environment	15 young adults (29.8 ± 5.2)	Labelling performed by the experimenter	Lying, sitting, reading, writing, working on a PC, watching TV),	Ankle, thigh, wrist, waist, hip (5; 60 Hz)	Mean, interquartile range, mean distance between axes, median, variance, standard deviation, zero crossing rate, main frequency peak,	SVM LOSOCV (98%)

					standing, walking, biking, running etc.		low and high-frequency band signal power (4s, **acceleration**)	
21	Lester et al. [41]	Structured protocol in university campus	12 young adults (20-30 years)	Labelling performed by the experimenter	sitting, standing, walking, jogging, walking up/down stairs, riding a bicycle, driving a car, and riding an elevator up/down.	Accelerometer, digital compass, light, temperature, IR, pressure and microphone sensors on shoulder, waist, wrist (7; 2Hz-16 kHz)	linear and log-scale FFT frequency coefficients, cepstral coefficients, spectral entropy, band-pass filter coefficients, correlations, integrals, means, and variances. (0.25s, **acceleration, pressure, temperature, IR sensor, microphone**)	Static and HMM 4-fold CV (90%)
22	Chowdhury et al. [42]	Structured protocol in laboratory environment	9 young adults (27.22 ± 3.31)	Predefined sequences of ADLs	lie, sit, stand, walk, run, cycle, ascend and descend stairs (from PAMAP)	Chest, wrist, ankle (3; 100 Hz)	Standard deviation, minimum, maximum, variance, median, skewness, kurtosis, energy, correlation, principal frequency, magnitude of principal frequency, median crossing, 21th and 75^{th} percentiles (2s, **acceleration**)	Posterior-adapted class-based fusion with random forest classifier LOSOCV (92.32%)
23	Gupta et al. [43]	Structured protocol in laboratory environment	7 young adults (22-28 years)	Predefined sequences of ADLs	Walking, jumping, running, sit-to-stand /stand-to-sit, sitting, sit- to-kneel-to-stand	Waist (1; 126 Hz)	Energy, entropy, mean, variance, mean trend, windowed mean, variance trend, windowed variance, detrended fluctuation analysis, spectral energy, max. difference acceleration (6s, **acceleration**)	KNN classifier LOSOCV (97.8%)

Table 2.2: Overview of the PAC systems developed for older adults

Sr. No.	Authors	Data collection protocol	Sample Population (age in years)	Gold standard/ labeling	ADLs Classified	Sensor Placement (no. of sensors; sampling frequency)	Features used (window size, analyzed signals)	Classifier, validation procedure, (accuracy)
1	Najafi et al. [25]	Semi-structured protocol in home environment	9 older adults (66 ± 14)	Labelling performed by the experimenter	Sitting, standing, walking, lying	Chest (1; 60 Hz)	features derived from discrete wavelet transform i.e. tilt angle vertical acceleration and displacement (60s, **acceleration, angular velocity**)	thresholding (mean sensitivity of 93.6%)
2	Godfrey et al. [27]	Structured protocol in home environment	10 young adults (23.7 ± 2.2) 10 older adults (77.2 ± 4.3)	Predefined sequences of ADLs	Sitting, standing and lying on various objects, transitions	Waist (1; 1 kHz)	Velocity estimate, tilt angle (activity based, **acceleration**)	thresholding (mean sensitivity of 92.5% for older subjects)
3	Rosario et al [29]	Structured protocol in university campus	20 young adults (21.9 ± 1.65) 37 older adults (83.9 ± 3.4)	Offline video labelling	standing, sitting, lying, walking on stairs up/down, riding an elevator up and down, transitions	Smartphone in front pocket of the trouser (1; 100 Hz)	cumulative sum of various angular velocity component, acceleration component due to gravity, differential pressure (2.5s, **acceleration, angular velocity, barometric pressure**)	Decision tree classifier LOSOCV (82.0% for older subjects)

4	Gao et al [16]	Structured protocol in home environment	8 older adults (76.50 ± 4.41)	Predefined sequences of ADLs	Sitting, standing and lying on various objects, transitions	chest, thigh, waist and left under arm (4; 20 Hz)	mean, standard deviation, variance, zero crossing rate, RMS, peak count, spectral energy, entropy, centroid, signal magnitude area, correlation, tilt angle, angle velocity (1s, **acceleration**)	Decision tree classifier 10-fold CV (96.4%)
5	Khan et al. [44]	Structured protocol in home environment	6 older adults (65 ± 3)	Self-labelling by the subject	Walking, resting, running, running, cycling vacuuming, walking up/down stairs	Single sensor tested on chest, front and back trouser pocket, inner jacket (1; 90 Hz)	Spectral entropy, autoregressive coefficients, signal magnitude area (1s, **acceleration**)	ANN LOSOCV (94.4%)
6	Lyons et al. [26]	Semi-structured protocol in rehab-center	1 older adult (NA)	Labelling performed by the experimenter	Sitting, standing, lying, moving	Thigh, trunk (2; 50 Hz)	Mean, standard deviation, tilt angle, (1s, **acceleration**)	Thresholding (90%)
7	Kamada et al. [45]	ADLs in free-living conditions	94 older adults (71.9 ± 6)	Self-labelling by the subject	Walking	Waist, wrist (2; 30 Hz)	Vector magnitude (60s, **acceleration**)	Manual (71%)
8	Ayachi et al. [46]	Structured protocol in laboratory environment	7 older adults (73 ± 4)	Labelling performed by the experimenter	Sitting, standing, walking, reaching ground, step over obstacle, reach up/down, release mid/down, turn right/left	Various body locations (17; 60 Hz)	Discrete wavelet transforms, (160ms, **acceleration, angular velocity**)	Thresholding (97.5%)

In summary, a graphical representation of the distinctive factors of existing PAC systems is depicted in Fig. 2.1. These factors consist of: datasets, number of sensors, placement of sensors, feature set, window size used for feature computation, and the classification approach used for the development of the model for PAC. We shall review them one by one in the following.

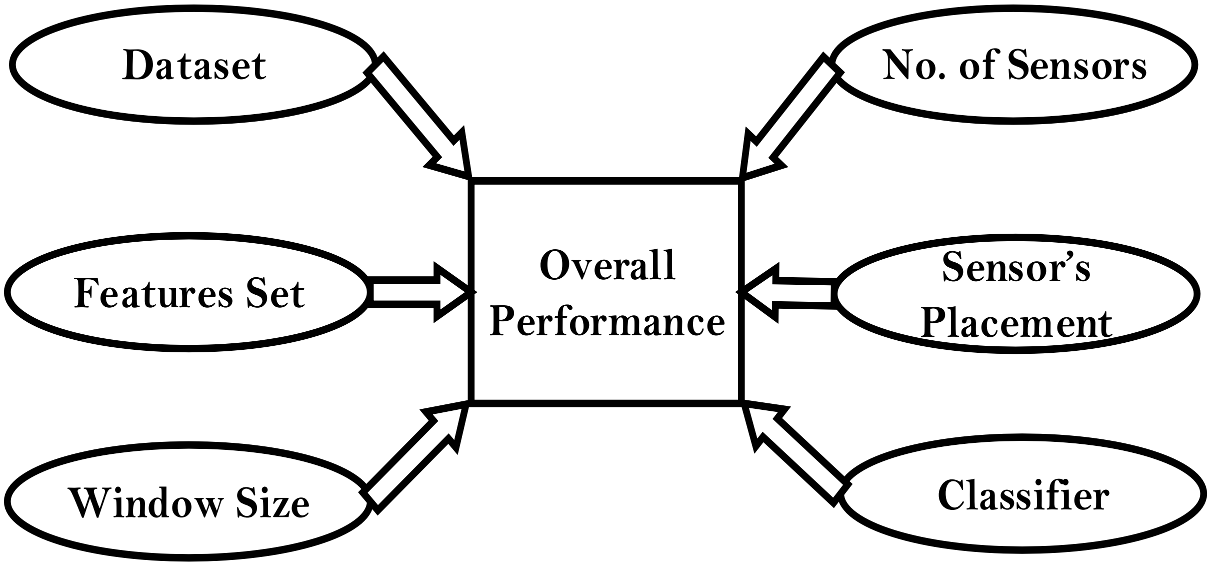


Figure 2.1: Factors that contribute to the overall performance of the PAC systems presented in the literature [47]

2.1.1 Dataset or Nature of the Dataset

Nature of the datasets differs regarding the sample population (younger or older), how and where the ADLs are performed (laboratory controlled environment or free-living conditions), and the type of ADLs (sitting, standing, walking, etc.) included in the dataset.

2.1.1.1 Younger vs Older Adults

The concise summary presented in Tables 2.1 and 2.2 shows that majority of the PAC systems (three fourths) have been developed and tested on younger adults' dataset and only a few PAC systems (one fourth) have collected and analyzed data of older adults. Therefore, it is important to investigate the peculiarities and assess the performances of such systems in order to develop PAC systems that are suited for the ever-increasing elderly population, with a higher overall quality.

The systems developed for younger adults' data cannot be directly transferred to older subjects' data without a loss in accuracy, since the quantity and quality of ADLs may be very different between the two age groups. This is because relevant characteristics of the ADLs performed by younger subjects are possibly different than the ones performed by the older subjects, even if the environmental conditions are comparable. In fact, a recent study by Rosario et al. [29] found that misclassification rate of the several ADLs (walking, sitting, stairs ascend/descend) was much higher for older adults as compared to younger adults considering the same system design (feature set, classifier, cross-validation procedure) and data collection environment. This study [29] also investigated if the same PAC system developed for the younger population is transferable to the elderly population. For this purpose, they trained the PAC system on younger adults' activity data and tested it on the older adults in a laboratory environment. Their findings showed that the performance of the system trained on younger adults and tested on older adults degraded significantly as compared to performance when the PAC system is trained and tested on older adults.

Another important factor to emphasize is the environmental conditions where the data collection is performed. It is worth noting that 21 out of 23 PAC systems developed for younger adults (Table 2.1) are using structured protocols where the ADLs are sequenced and mostly performed in a laboratory environment [10, 12, 14, 17, 21-23, 34, 36]. There are only two PAC systems developed in free-living conditions (Table 2.1): Hickey *et al.* [31] and Lee *et al.* [33]. The system by Hickey *et al.* only classify the walking bouts and did not explore other commonly performed ADLs (e.g. sitting, standing, lying). The system by Lee *et al.* uses the GPS trajectories to keep the ground truth information, which might not be much reliable to fully validate the PAC system in free-living conditions, as compared to the validation performed by Hickey *et al.* using offline video marking. Furthermore, the system by Lee *et al.* is developed using only two subjects and its use is limited to outdoor environment because of the erroneous behavior of the GPS signal in indoor environments.

Similarly, majority of the PAC systems (5 out of 8) developed for older adults in Table 2.2 uses structured protocols for data collection. Only the PAC system developed by Kamada *et al.* [45] use the dataset collected in free-living conditions. However, their system mainly focuses on the detection of walking activity, and other ADLs were not classified. Moreover, in the system by Kamada *et al.*, ADLs were self-annotated, which certainly biased the quality of the assessment procedure. The PAC system developed by Najafi et al. [25] and Lyons et al. [26] (Table 2.2) utilizes semi-structured protocols, where the subjects were instructed to perform certain ADLs at their usual pace. However, the ADLs were performed with the presence of an observer, who labeled the ADLs, which might influence the natural behavior of performing ADLs, as performed in free-living conditions where subjects have more freedom in performing their ADLs more naturally and without any sequence. Secondly, these two systems use a few number of subjects (below 10), which are insufficent for the generalisibility of the findings.

2.1.2 Number of sensors

The number of sensors in PAC systems varies from a single sensor setup [10, 27] to multiple sensors setup [9, 12, 13]. The multi-sensor sensor set-up ranges from two sensors [17, 34] to as many as 17 sensors [36, 46].

These sensors also differ in terms of sampling rates, ranging from 20 Hz [16] to 1 kHz [27]. Certainly, the larger the sampling frequency the higher the power consumption of the sensors as well as the computational complexity of the system. Furthermore, high sampling frequencies of movement signals has no significant contribution as all body movements can be captured below 20 Hz [48]. The most commonly analyzed signal for activity classification is acceleration and the second is angular velocity (Table 2.1 and 2.2). This could be because of the low power requirement of the acceleration signal as compared to angular velocity signal, which increases the data recording time significantly.

2.1.3 Placement of sensors

As shown in Tables 2.1 and 2.2, the sensors' placement may be very different, covering several body locations to record the upper and lower body movements. The sensors locations used for the development of the PAC systems in the literature (Table 2.1 and 2.2) are presented in Figure 2.2 and the number indicates how many studies used such locations. The most commonly used sensor locations are: chest, wrist, waist, thigh and feet.

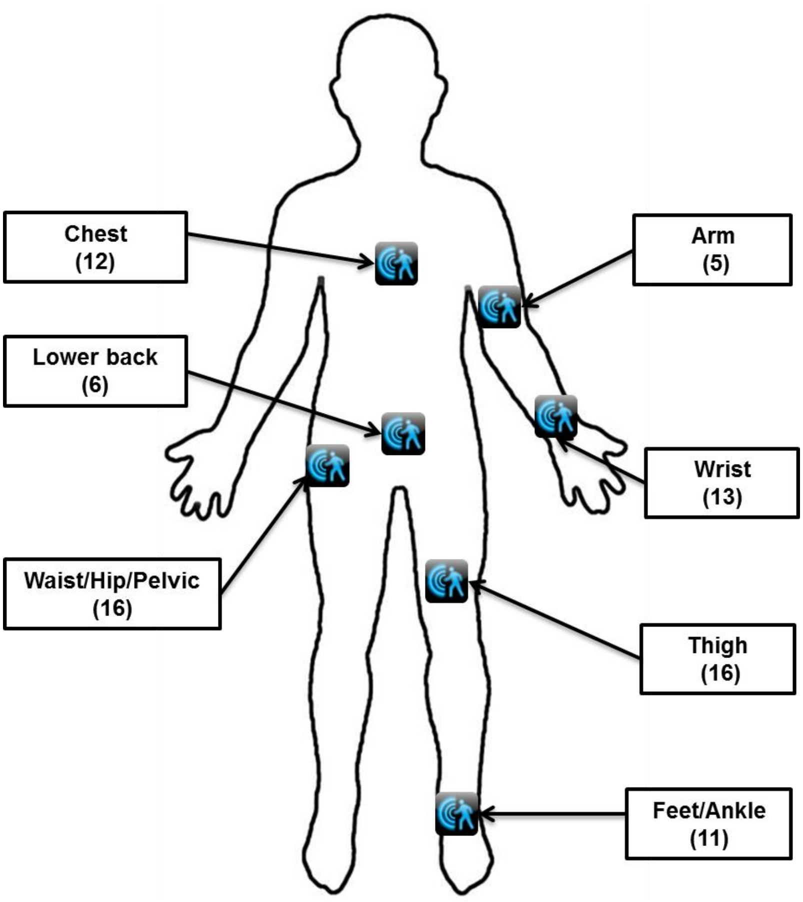


Figure 2.2: Sensor locations used by the PAC systems in the literature and a number indicating how many systems used such locations

2.1.4 Features set

Existing PAC systems have used numerous time and frequency domain features, statistical features and biomechanical features [49, 50] as shown in Table 2.1 and 2.2.

These features include summary statistics measures (i.e. mean, standard deviation, minimum, maximum, variance, median, skewness, kurtosis, root means square, etc.), signal magnitude area, energy, spectral measures (i.e. centroidal frequency, principal frequency, power at the principal ftrequency, etc.), non-linear measures (i.e. entropy), etc. Thus, depending upon the intended objectives, various kind of features can be considered for the development of PAC system.

2.1.5 Window size

Window size and overlapping intervals used for the feature computation are also very different across studies, and these may affect the performance of machine learning algorithms. The window size largely differs across the PAC systems proposed in the literature: 2 sec [12], 2.5 sec [29], 5 sec [13], 5.12 sec [10], 6.7 sec [9], 10 sec [22]. The overlapping interval used in most of the PAC systems is 50% of the window size [49]. Therefore, large diversities exist in the literature in choosing the window size and overlapping intervals.

2.1.6 Classifiers/ Machine Learning Approaches

In most of the PAC systems, a single classifier is used to differentiate between all the different ADLs in the dataset. A common choice for such classifiers may include a decision tree classifier [9], Support Vector Machine (SVM) [51], Artificial Neural Network (ANN) [16], random forest [52], Naïve Bayes [53] and K- Nearest Neighbors (KNN) [54]. However, some PAC systems are developed by integrating base level classifiers either by plurality voting [10] or by defining a hierarchical classification process which uses different classifiers for each subset of ADLs [14, 21, 55].

There are also some newly developed systems [35, 36, 38] which skip the feature computation stage and implement machine learning approaches (e.g. convolutional neural networks (CNN), deep convolution long-term-short memory (DeepConvLSTM) neural networks) directly to the raw dataset. However, these systems are not yet fully exploited in the free-living conditions and on the older adults. Therefore, these methods and also the feature-based methods require further investigation, in order to objectively classify the ADLs of the elderly population in free-living conditions.

Another important factor to highlight is the cross-validation procedure to compute the performance of PAC systems. The more common validation procedures are the 10-fold cross-validation procedure and the leave-one-subject-out cross-validation (LOSOCV) procedure. In 10-fold cross validation, 90% of the data samples are used to train the classifier, and the remaining 10% of data is used to test the performance of the classifier. This process is repeated 10 times to test all data samples. In LOSOCV, the classifier is trained on all subjects except the one that is being tested. This process is repeated until all subjects get tested.

2.2 Conclusions

The review suggests that majority of existing systems are developed using the younger subjects' data and a few systems are developed using the older subjects' data. The scope of such systems is limited to stimulated and structured activities which differ from real life activities, where the activities are more naturally performed and in an unstructured way. Consequently, there is a need to develop PAC system for the elderly population, validated in free-living conditions. A group of researchers [56] recently proposed a set of recommendations about the standardization of validation procedures for PAC systems in older people. It emphasizes the need to develop and validate the PAC using a semi-structured protocol where ADLs are pesrformed in real life conditions, in addition to the validation performed in the laboratory-based settings.

Moreover, the choice of every single factor discussed above (Fig. 2.1) is crucial in the development of a robust PAC system since all these factors contribute directly to overall performance. Due to the large diversity in the design process, the existing PAC systems are not easily comparable which hinders the development of new techniques informed by the strengths and the gaps of current systems.

Chapter 3

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

This chapter is largely taken from our published work [47], which is about the performance evaluation of the state of the art methodologies for physical activity classification of the older subjects.

3.1 Introduction

The conclusion drawn from the review of the existing methodologies for PAC has shown that the current PAC systems are not directly comparable, due to the large diversity in their design (e.g., number of sensors, placement of sensors, data collection environments, data processing techniques, features set, classifiers, cross-validation methods). In the past, some researchers [14, 51, 57] have tried to compare the performance of their proposed PAC systems with existing systems. However, in our opinion, they failed to provide a fair comparison, since they did not consider that the factors reported in Figure 2.1 were not comparable.

Therefore, this analysis aims to propose a fair and unbiased benchmark for the field-based validation of existing state of the art (SOA) systems for PAC of older subjects highlighting the gap between the laboratory and real-life conditions. The specific aims are as follows:

- To compare the performance of existing PAC systems in a common dataset of activities of older subjects in an unbiased way (i.e., with the same subjects, sensors, sampling frequency, window size and cross-validation procedure) and to investigate the effect of varying window size on system's performance.

- To validate and compare the performance of the PAC systems in real life scenarios compared to an in-lab setting to check if these systems are transferable to real-life settings.

- To evaluate the impact of the number of sensors on the performance in the analyses in 1) and 2) using a reductionist approach (i.e., analyzing only the sensing unit worn at the lower back instead of the multi-sensor setup). The lower back location is chosen since it is commonly used for elderly population and does not show any major drawbacks for long term activity monitoring in terms of feasibility.

For the presented aims, we selected three representatives SOA systems for PAC [9, 14, 22] motivated by the following reasons: i) diversity in the number of sensors used; ranging from four sensing units by Leutheuser *et al.* [14] up to six sensing units by Cleland *et al.*[22]; ii) use of different time intervals for windowing (ranging from 5 sec [14] to 10 sec [22]); iii) different classification techniques i.e. decision tree classifier by Bao *et al.* [9], SVM by Cleland *et al.* [22], and hierarchical classification by Leutheuser *et al.* [14]. Four ADLs (sitting, standing, walking, and lying) are studied in this work to provide a fair comparison. These ADLs are chosen as they are the most common in this kind of studies and due to these four activities being present in all of the selected systems.

3.2 Materials and Methods

3.2.1 Data Collection in Real-Life Scenarios

The data collection was performed at the Department of Neuroscience, Faculty of Medicine, at the Norwegian University of Science and Technology (NTNU) Norway, by the research group on Geriatrics, Movement, and Stroke, as part of the ADAPT project (A Personalized Fall Risk Assessment System for promoting independent living). A detailed description of the ADAPT dataset and the video annotation process is presented in the study protocol by Bourke et al. [58]. The data collection protocol was composed of two sessions: a semi-structured supervised protocol (in-lab) and a free-living unsupervised protocol (out-of-lab). Twenty older subjects (76.4 ± 5.6 years) participated in the study. For both data protocol sessions, video recording was used as a gold standard. Various inertial sensing units were placed on different body locations, and a subset of these sensors was used in our analysis: chest, lower back (L5- 5th lumbar vertebrae), dominant wrist, waist, left thigh, and right foot. The details of the sensors used, and their respective placements are presented in Table 3.1.

All sensors were part of in-lab and out-of-lab protocols except the sensor on the feet which was excluded from out-of-lab data recording for usability issues. Each subject performed a variety of ADLs in both sessions with the ADLs analyzed in our study being sitting, standing, walking, and lying. The in-lab session was performed in a smart home environment where subjects were supervised and instructed to perform ADLs. Video recording was performed using the ceiling-mounted cameras at 25 fps. The in-lab session was followed by an out-of-lab session on the same day where subjects performed their daily routine activities in an unsupervised way. They were instructed to perform as much ADLs as possible and to incorporate certain tasks into their daily routine. A GoPro camera unit with a frame rate of 29 fps (fixed to the chest pointing downward towards the feet) was used to video record the gold-standard information of the ADLs performed in free living protocol.

Table 3.1: Description of the sensors used from the ADAPT dataset.

Sensor Type	Location	Sampling Frequency	Measured Signals
uSense	Thigh	100 Hz	3D Accelerometer, 3D Gyroscope
uSense	L5	100 Hz	3D Accelerometer, 3D Gyroscope
ActiGraph	Waist	100 Hz	3D Accelerometer
uSense	Chest	100 Hz	3D Accelerometer, 3D Gyroscope
Shimmer	Wrist†	200 Hz	3D Accelerometer, 3D Gyroscope
uSense	Feet *	100 Hz	3D Accelerometer, 3D Gyroscope

† Initially collected at 200Hz but later down sampled.

* Sensor on the feet were not included in out-of-lab data collection.

Video annotation of the camera units used in the in-lab and out-of-lab protocols was performed by the recruited raters. Raters were instructed on the marking procedures and activity definitions. For both sessions, video annotation agreement was around 90%. The original sampling frequency (25 Hz) of the annotations was up-sampled to 100 Hz [58]. Due to technical issues with the wrist sensor, 16 subjects were used for analysis purposes as the authenticity of sensed data was compromised in rest of the cases due to missing data at the time of recording. Therefore, four subjects were excluded from the analysis as all selected PAC systems make use of the wrist sensor data.

A summary of the ADLs from 16 subjects analyzed from the in-lab and the out-of-lab protocols is presented in Tables 3.2 and 3.3, respectively. Statistical analysis was performed and various parameters were computed: occurrences (how many times a single ADL occurred in all subjects), mean (average duration of each ADL in seconds), STD (standard deviation of each ADL in seconds), min (minimum duration of each ADL in seconds), max (maximum duration of each ADL in seconds), and range (difference between min and max in seconds).

Table 3.2: In-lab ADLs.

ADL	Total (hours)	Occurrences	Mean *	STD *	Min *	Max *	Range *
sitting	1.67	708	8.50	18.90	0.03	267.36	267.33
standing	2.67	1319	7.28	16.40	0.03	296.97	296.94
walking	0.90	613	5.29	2.79	0.96	20.07	19.11
lying	0.28	187	5.47	9.87	0.13	113.23	113.10

* The values are in seconds.

Table 3.3: Out-of-lab ADLs.

ADL	Total (hours)	Occurrences	Mean *	STD*	Min *	Max *	Range *
sitting	13.45	497	97.44	200.74	0.04	2075.64	2075.60
standing	6.52	4304	5.45	12.27	0.03	388.52	388.49
walking	4.10	2617	5.64	8.75	0.28	139.56	139.28
lying	0.36	12	106.69	154.02	3.48	583.84	580.36

* The values are in seconds.

3.2.2 Implementation of the SOA Systems for PACs Using Their Original Framework

The set of sensors used in our work for the in-lab (S_{IN}) and out-of-lab (S_{OUT}) analysis performed on the ADAPT dataset is shown in Table 3.4. The brief description of the three PAC systems, selected for the comparative analysis is presented in Table 3.5. It is much evident from Table 3.5 that all PAC systems possess different solutions for a number of sensors, sensor locations, set of features, classifiers, and time window used for feature computation.

To investigate the sensitivity of the classification accuracy to window size (first specific objective), all systems are trained and tested in the in-lab data with a window size ranging from w = 1 s to w = 10 s in steps of 1 s. 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.

Table 3.4: Sensors used from ADAPT dataset to assess the performance on three PAC systems.

Author	$\mathbf{S_{IN}}$	$\mathbf{S_{OUT}}$
Cleland et al. [9]	Chest, L5, Wrist, Waist, Thigh, Foot	Chest, L5, Wrist, Waist, Thigh
Bao et al. [2]	L5, Wrist, Thigh, Foot	L5, Wrist, Thigh
Leutheuser et al.	Wrist, L5, Chest, Foot	Wrist, L5, Chest

$\mathbf{S_{IN}}$ —Sensors used in our data analysis from In-lab protocol of ADAPT dataset; and $\mathbf{S_{OUT}}$ —Sensors used in our data analysis from out-of- lab protocol of ADAPT dataset.

Analysis of the out-of-lab data is performed by training and testing all systems with the real-life data. The window size of 5s is used with the sensor set S_{OUT} (Table 3.4) and leave-one-subject-out cross-validation is performed. The window size of 5 s is chosen, since it is closer to the window size used by two out of three PAC systems (Table 3.5).

To address the second specific objective, each PAC system is trained with the in-lab data and tested on the out-of-lab data. To overcome any bias in the training process, the in-lab data of all subjects except one is included in the training stage. The left-out subject is tested in free-living conditions (i.e., with the out-of-lab data). In this way, all participants are tested in free-living condition using this leave-one-subject-out strategy. The sensor set S_{OUT} is used with the window size of 5 s.

Table 3.5: Overview of the three SOA systems for PACs implemented in this study for performance analysis.

Author	Fs (W)	S_O	Experiment Setting (Population)	Features	Activities	Accuracy Reported
Cleland et al. [22]	51.2 (10 s)	Chest, lower back, wrist, hip, thigh, foot	Laboratory setting (8 young adults) (26.25 ± 2.86 years)	Mean, standard deviation, skewness, kurtosis, energy and correlation of axes (separately and average over 3 axes)	Walking, jogging on a treadmill, sitting, lying, standing, walking up stairs, walking downstairs	SVM: 97.26%
Bao et al. [9]	76.25 (6.7 s)	Hip, wrist, arm, thigh, ankle	Semi-naturalistic conditions (20 subjects) age group not reported	Mean, energy, frequency domain entropy, correlation between the acceleration signals	Walking, sitting, standing, eating or drinking, watching tv, reading, running, bicycling, stretching, strength-training, scrubbing, vacuuming, folding laundry, lying, brushing, climbing stairs, riding elevator, riding escalator	Decision tree: 84%
Leutheuser et al. [14]	204.8 (5 s)	Wrist, hip, chest, ankle	Laboratory setting (23 young adults) (27 ± 7 years)	Minimum, maximum, mean and variance, spectral centroid, bandwidth, energy, gravitational component	Sitting, lying, standing, washing dishes, vacuuming, sweeping, walking, running, stairs climbing, bicycling, rope jumping	Hierarchical classifier : 89.6%

Fs—Sampling Frequency in Hz, W = Window Size, $\mathbf{S_O}$—Original set of sensors used by the authors to develop PAC system, Activities—Set of Activities used by authors to develop their PAC system.

The overlap is set to 50% of the window size for all the analysis. Furthermore, a majority voting scheme is implemented to assign the window labels, i.e., if a window of 5 s (500 samples) contains 400 samples of sitting and 100 samples of standing then the assigned label to this window would be sitting.

All of the PAC systems are implemented in MATLAB (Release 2014b, The MathWorks, Inc., Natick, MA, USA) and respective classifiers are implemented using the libraries of Weka data mining software (University of Waikato, Version 3.6.12 [59]). The analysis is performed on a Dell laptop (Model # M3800, Intel® Core™ i7-4712HQ, CPU @2.30Gz, 16GB RAM, 64-bit operating system). For all systems, overall accuracy, accuracy by class, and sensitivity by class of all activities is computed in the in-lab training/out-lab testing scenario. The overall accuracy term will be used interchangeably as accuracy or performance in the upcoming sections. The formulas used for the computation of performance metrics are reported in Appendix A, and the respective classification methods implemented for each PAC system are described in Appendix B.

3.2.3 Implementation of the SOA Systems for PAC Using a Reductionist Framework

The performance of all systems is also computed in the reductionist framework implemented using only the sensor data collected at waist-level in L5 (third specific objective). The steps in the analysis are the same as described in Section 3.2.2.

3.3 Results and Discussion

3.3.1 Performance Comparison of the PAC Systems in the In-Lab Setting Using Their Original Framework and Sensitivity Analysis to the Window Size

Overall accuracy computed for the sensitivity analysis of the in-lab data to different window sizes (w = 1 s to 10 s) is presented in Figure 3.1.

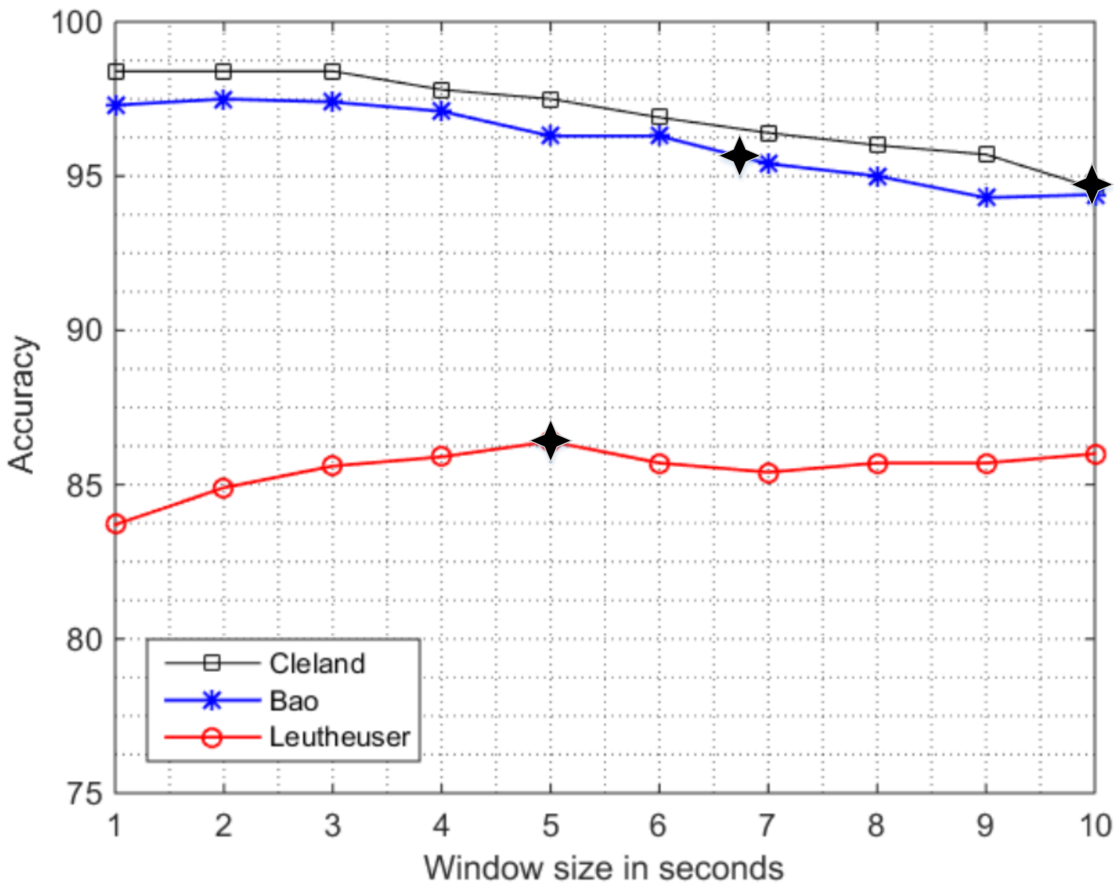


Figure 3.1: Sensitivity analysis of the overall accuracy of in-lab data when window size is increased from w = 1 s to w = 10 s using sensor set S_{IN} (Table 4). The symbol (✦) specifies the window size used in the original PAC system by the authors.

The system by Cleland et al. [22] performs best in our framework, with an overall accuracy ranging from 98.4% for w = 1 s to 94.6% for w = 10 s. Hence, it shows a degradation by 3.8% when increasing the window size. Our result for in-lab data compares well with the original paper that, for w = 10 s, reported an overall accuracy of 97.3%. The second-best performance we obtained is with the system proposed by Bao et al. [9]. It also shows a decreasing trend in the overall accuracy from 97.3% (for w = 1 s) to 94.4% (for w = 10 s) with a difference of 2.9%. The original system was implemented with w = 6.7s and had an overall accuracy of 84%; our closest term of comparison is the window with w = 7 s, which produces an accuracy of 95.4%. The accuracy of the system by Leutheuser et al. [14] is below the aforementioned. In the system by Leutheuser et al., we obtain an overall accuracy which, unlike previous systems, increases by 2.3%, from 83.7% (w = 1 s) to 86.0% (w = 10 s). Results obtained in our framework (overall accuracy of 86.4%) fits well with the original one at w = 5 s (overall mean classification rate of 89.6%). A possible reason for the increase in the performance (although the performance is the worst of the three) for increasing window sizes of the system by Leutheuser et al. is the difference in the classifier design. Their work is the only one that uses a hierarchical classification approach.

The systems by Bao et al. [9] and Cleland et al. [22] achieved very high accuracies, at the cost of using a large number of sensors, which is a practical issue in real-life conditions. The system developed by Bao et al. uses four sensors and the system proposed by Cleland et al. uses six sensors, which raise feasibility and computational complexity issues for these systems, which could make them less practical in real life conditions.

The probable cause in the overall lower performance of the system by Leutheuser et al. could be the fact that in their original implementation, six subsets of ADLs were considered (1: HOUSE (vacuuming, sweeping); 2: REST (sitting, standing, and lying); 3: WALK (walking, running, ascending stairs, descending stairs); 4: bicycling; 5: rope jumping; 6: washing dishes). Instead, in our analysis, only two sub-systems are used i.e., REST (sitting, standing, lying) and WALK (walking). The subdivision of ADLs which characterizes this hierarchical classification can be a limitation in implementing the original work when choosing only a subset of activities, as in our case. It could also be an issue if a hierarchical classification approach is implemented on a set of activities which is not the same as the original PAC system.

Our findings regarding the decrease in performance are in line with the recent work by Fida et al. [23] who analyzed the effect of varying window size from w = 1 s to 3 s and suggests that 1 s to 2 s window size gives a better tradeoff when analyzing static and dynamic activities. On the contrary, more recently Shoaib et al. [24] proposed a system for complex human activity recognition by varying window sizes from 1 s to 30 s and found that increasing window size improves the recognition rate of complex activities. However, our analysis is novel due to the demographics of the studied population. Our work investigates the activities of older adults, whose ADLs differ from those analyzed by Fida et al. and Shoaib et al. on the younger subjects.

It is possible that the performances of all the three PAC systems would decrease if the number of the ADLs are scaled up. This is because more robust set of features would be required to build the system model instead of using the same feature set.

3.3.2 Performance of the PAC Systems in Real-Life Scenarios

3.3.2.1 In-Lab vs. Out-of-Lab

The results of the out-of-lab analysis show a decreased accuracy concerning the in-lab across all systems. Figure 3.2 (first and last point on time axis), shows the overall accuracy of the three systems in the in-lab and out-of-lab with w = 5 s, chosen as a representative window size. A slight decrease of 1% (96.4%–95.4%) in work by Cleland et al. and 1.3% (94.7%–93.4%) in work by Bao et al., is observed.

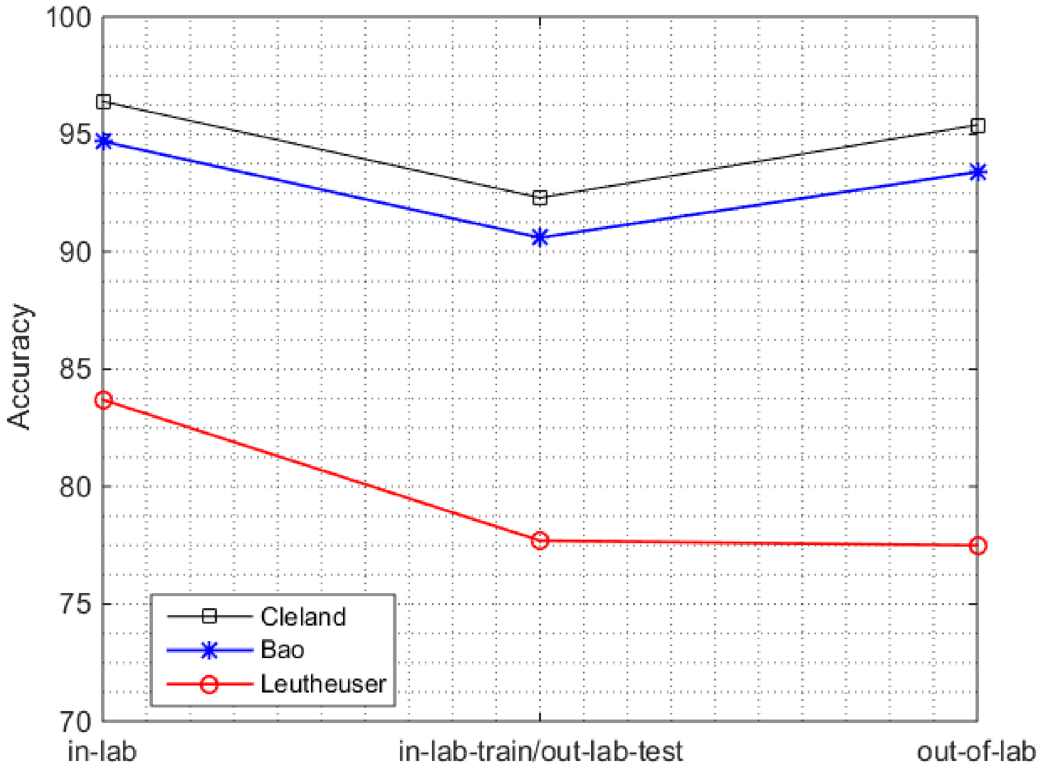


Figure 3.2: Performance analysis of in-lab, out-of-lab, and in-lab training/out-lab testing scenario for all PAC systems using sensor set S_{OUT} (Table 3.4).

However, such degradation is larger in work by Leutheuser et al. with a decline of 6.2% (83.7%–77.5%). The best performance of 95.4% is obtained (when trained and tested on the real-life data) by the system of Cleland et al. which is quite encouraging, but at the cost of using five sensors and large features set, which may not be feasible in real-life conditions.

3.3.2.2 In-Lab Training/Out-Lab Testing

We then evaluated the performance of in-lab trained systems in the real-life setting. In the in-lab training/out-lab testing scenario, the performance of all the SOA systems decreased between 4–6% when compared to the in-lab results (Figure 3.2).

Table 3.6: Confusion matrix for the systems; (a) Bao et al.; (b) Cleland et al.; and (c) Leutheuser et al.; in the in-lab training/out-lab testing scenario.

	(a) Bao et al.				
	Predicted Class				
Actual Class	**stand**	**walk**	**sit**	**lie**	**←classified as**
	9214	571	4	0	stand
	2329	4000	2	9	walk
	24	16	19,260	197	sit
	233	0	2	278	lie
	(b) Cleland et al.				
	Predicted Class				
Actual Class	**stand**	**walk**	**sit**	**lie**	**←classified as**
	9712	73	4	0	stand
	2474	3857	9	0	walk
	1	1	19,492	3	sit
	0	0	234	279	lie
	(c) Leutheuser et al.				
	Predicted Class				
Actual Class	**stand**	**walk**	**sit**	**lie**	**←classified as**
	7423	350	1572	16	stand
	395	5397	94	0	walk
	5289	107	13,950	0	sit
	0	0	15	480	lie

Each individual instance in the table corresponds to 5 s or 500 samples of data

Table 3.7: Accuracy and sensitivity by class for all SOA systems for PAC in the in-lab training/out-lab testing scenario.

Authors	Accuracy	Accuracy by Class				Sensitivity by Class			
		Stand	Walk	Sit	Lie	Stand	Walk	Sit	Lie
Bao et al.	90.6	91.3	91.9	99.3	98.8	94.1	63.1	98.8	54.2
Cleland et al.	92.3	92.9	92.9	99.3	99.3	99.2	60.8	100.0	54.4
Leutheuser et al.	77.7	78.3	97.3	79.8	99.9	79.3	91.7	72.1	97.0

The respective confusion matrix for each SOA system for PAC is shown in Table 3.6, where sensor set S_{OUT} (Table 3.4) is used for implementation of all systems. Each sample of the confusion matrix corresponds to a 5s window.

Moreover, the accuracies by class and the sensitivities by class for all PAC systems in the in-lab training/out-lab testing scenario are listed in the Table 3.7. The decreases in accuracy are: from 96.4% to 92.3% (4.1%) in the work by Cleland et al., from 94.7% to 90.6% (4.1%) in the work by Bao et al., and from 83.7% to 77.7% (6.0%) in the work by Leutheuser et al.

The degradation of performance in all the systems in this scenario reflects the lack of field-based validity as highlighted more recently by Lindemann et al. [56]. The reason for this degradation is due to the fact that:

- Most of the existing PAC systems are developed using a standardized protocol which does not include the ADLs performed under real-life conditions.

- The order and way of performing these activities in a more natural and quite different environment to the one performed in a laboratory environment.

Therefore, these PAC systems are unable to recognize unstructured and unplanned activities in real-life conditions, which emphasizes the need of developing in-field, validated, PAC systems, as we did when considering the out-of-lab scenario.

Our findings are in-line with the work by Ganea et al. [28], where performance deteriorated when the laboratory-trained system was tested in real life. Our analysis generalizes the fact of performance deterioration over several activities in real life conditions by analyzing sitting, standing, walking, and lying instead of only postural transitions, as analyzed by Ganea et al.

3.3.3 Computational Complexity/Burden in the Real-Life Settings

The computational complexity/burden of testing out-of-lab data (when trained on in-lab) is also analyzed by measuring the time required for the feature extraction and for classification (Table 3.8). The feature computation time is the time required to compute the features of all 16 subjects from out-of-lab data using the sensor set S_{OUT} (Table 3.4). The testing out-of-lab time is the total time to test all the out-of-lab data for 16 subjects. Mean and standard deviation of 10 runs (to account for computer performance variability) are reported in Table 8. The total window instances obtained (after the feature extraction of the out-of-lab data) for all systems are 36,139 except the system by Leutheuser et al. [10], for which the samples are 35,088 because of the software dependencies.

Table 3.8: Computational complexity in the in-lab training/out-lab testing scenario.

Author	Feature Computation Mean ± Std (s)	Testing Out-of-Lab Mean ± Std (s)
Bao et al.	337.07 ± 3.10	25.27 ± 0.95
Cleland et al.	458.79 ± 6.57	738.21 ± 1.09
Leutheuser et al.	772.41 ± 11.99	957.83 ± 18.38

The time consumption analysis of the features computation shows that the time required to compute the features has a direct relationship with the number of sensors. All three systems used multiple sensors and took longer time for feature computation.

Moreover, the number of features, and the nature of the features, also plays an important role in the computational complexity of the system. For instance, in the work by Leutheuser et al., activity-specific features and hierarchical structure increased the time consumption for the validation. The complexity of the classifier, along with the number of sensors increased the computational time in the systems by Leutheuser et al., and Cleland et al. On the other hand, the time taken by Bao et al. is much shorter since it utilizes a simpler classifier approach (decision tree classifier). The computational analysis suggests that to make the PAC system operational in real time, the optimum number of sensors, proper feature selection to eliminate redundant features, and the choice of simpler and more robust classifiers, is very critical. Most of the existing systems do not highlight these factors, especially the selection of features, and of a reduced set of sensors. These factors are crucial for the practical implementation of these systems out of the laboratory.

3.3.4 Performance Comparison of the PAC Systems in the In-Lab Setting Using a Reductionist Approach and Sensitivity Analysis to the Window Size

The overall performance of the PAC systems using a reductionist approach obtained from the in-lab sensitivity analysis to window size is depicted in Figure 3.3. In-lab sensitivity analysis using a single sensor at L5 location (Figure 3.3) follow a decay in performance with the increase in window size (similar to that presented in Section 3.1) for the systems by Bao et al. and Cleland et al. The deterioration in accuracy from w = 1 s to w = 10 s was 5.3% by Bao et al. and 4.8% by Cleland et al. However, an improvement of 1.7% in accuracy is observed in the work by Leutheuser et al. In this case, the use of activity specific classification systems instead of using the generalized systems for ADLs seem to be the probable cause.

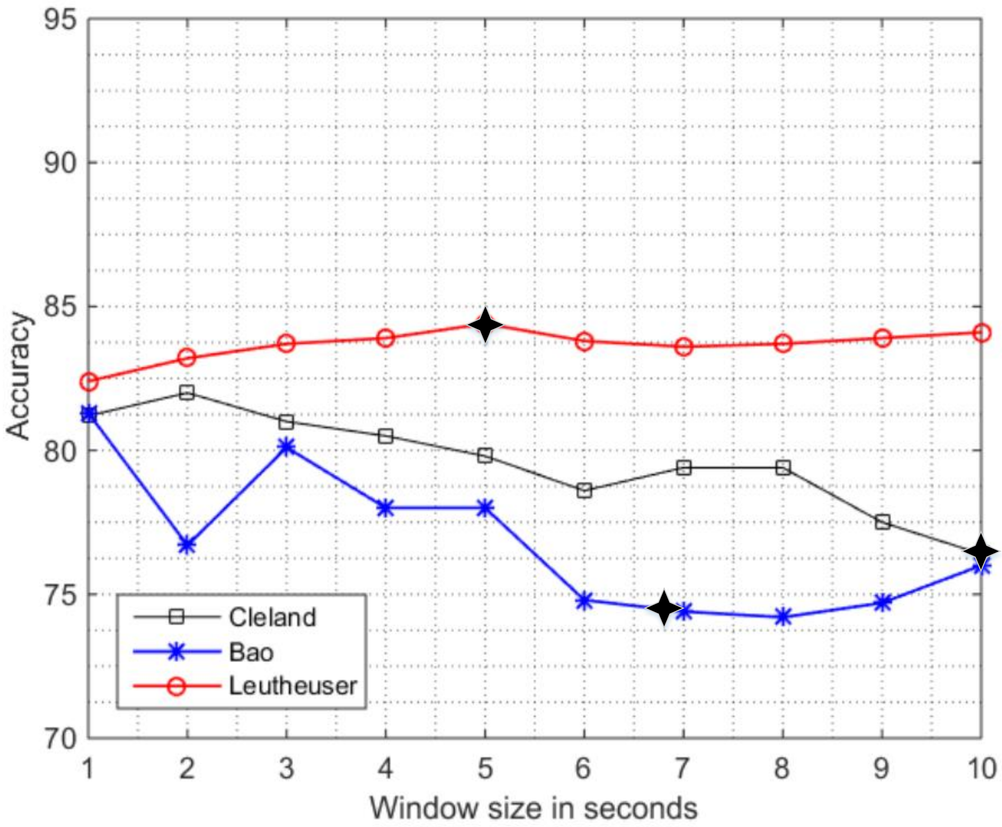


Figure 3.3: Sensitivity analysis of overall accuracy of in-lab data when window size is increased from w = 1 s to w = 10 s using reductionist approach. The symbol (✦) specifies the window size used in the original PAC system by the authors.

3.3.5 Performance of the PAC Systems in Real-Life Scenarios Using a Reductionist Approach

3.3.5.1 In-Lab vs. Out-of-Lab

The analysis using the reductionist approach (Figure 3.4) shows that accuracy of all systems is decreased except for Cleland et al. in the out-of-lab when compared to in-lab.

The decrease is: 2.7% in the work by Bao et al., 6.4% in the work by Leutheuser et al. The slight increase of 1% is observed in the work by Cleland et al. The best performance of 80.9% is achieved by the work of Cleland et al. (similar to Section

3.3.2.1) when trained and tested on the real-life data which show the potential of using a single sensor in real life conditions. This performance can be enhanced by developing a PAC system which incorporates more discriminative features (e.g., biomechanical features) and a robust classifier.

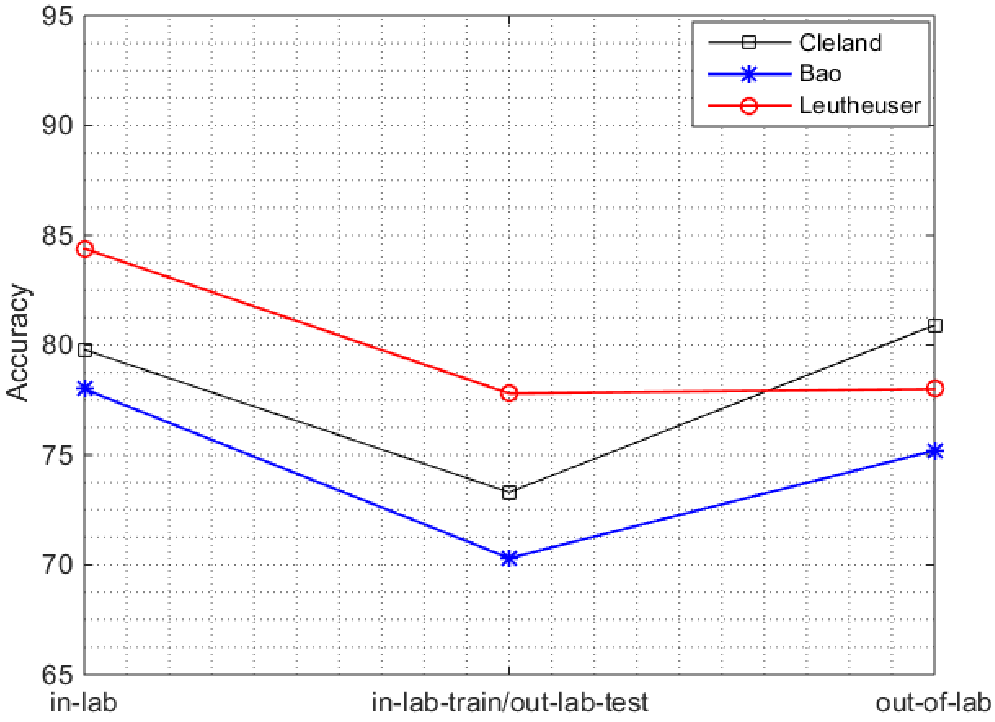


Figure 3.4: Performance analysis of in-lab, out-of-lab, and in-lab training/out-lab testing scenario for all PAC systems using a reductionist approach.

3.3.5.2 In-Lab Training/Out-Lab Testing

The in-lab training/out-lab testing analysis on the single sensing unit also followed the deterioration in overall accuracy and the differences are a bit larger (between 6–8%) than in the multi-sensor setting (Section 3.3.2.2) as described by Figure 3.4. The reduction in the accuracies are: 79.8% to 73.3% (6.5%) by Cleland et al., 84.4% to 77.8% (6.6%) by Leutheuser et al., and 78.0% to 70.3% (7.7%) by Bao et al.

The performance of all systems, both in the original framework and in the reductionist approach degrades for the in-lab testing/out-lab training scenario (when compared to in-lab analysis). Therefore, it is very important to develop a PAC system in the real-life data before releasing it for real life applications, as we did in the out-of-lab analysis. Most of existing system lack this perspective so their performance cannot be generalized for the real-life conditions.

3.4 Conclusions

A benchmark study is presented which investigates the performance of various SOA systems for PAC in the in-lab and out-of-lab environment. The sensitivity analysis to window size shows that the increase in window size generally degrades the performance. The in-lab training/out-lab testing analysis concludes that the systems developed in controlled settings are not capable of performing well in real-life conditions where the ADLs are performed more naturally. Therefore, the newly developed systems should be trained and tested on the dataset collected in the real-life conditions. The reductionist approach also obtained similar results for all analyses (in-lab sensitivity analysis to window size, out-of-lab analysis, in-lab training/out-lab testing) but the degradation is much larger than the multi-sensor setup. Furthermore, investigation of the computational complexity is conducted for the feature extraction stage and the classifier testing stage of out-of-lab data. The findings, as we expected, show that the systems with more complex classifier approaches and large numbers of sensors increases the computational complexity of the system.

The reductionist approach we developed, derived from existing systems, is an important first step to study the effect of reducing the number of sensors to find an optimal trade-off between usability and performance (the use of multiple sensors on various body locations can be impractical in real-life).

Our future aim is to develop a physical activity classification system in real life conditions with an optimal number of sensors (by exploring various sensor locations), improved feature set (using various feature selection approaches), and robust classification methods to perform comparably to, or better than, existing systems.

Chapter 4

Physical Activity Classification for Elderly Population in Free-Living Conditions

---

The findings of this chapter have been submitted as an article (currently under review) to the IEEE Journal of Biomedical and Health Informatics.

4.1 Introduction

Our previous findings (Chapter 3) 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 [47]. Therefore, the present work is in continuity with earlier works. It presents and validates an inertial sensors-based physical activity classification system developed in free-living conditions with older adults as the target population. The main objectives of this work are as follows.

- To develop an inertial sensors-based PAC system trained and tested in free-living conditions for older adults;
- To analyze the impact on its performance (accuracy and computational complexity) of various feature selection techniques;
- To analyze multi-sensor versus single-sensor solutions, to highlight the optimal number of sensors that can achieve an acceptable level of performance.

The flow diagram of data analysis performed to achieve the aforementioned objectives is presented in Figure 4.1. Before developing the PAC system for the elderly population in free-living conditions, we also developed a PAC system [60] on younger adults' data in the laboratory controlled environment. The results were compared with the state of the art methodology [14] using the same benchmark dataset. Our proposed PAC system outperformed the state of the art methodology [14] and this study findings are presented in Appendix C.

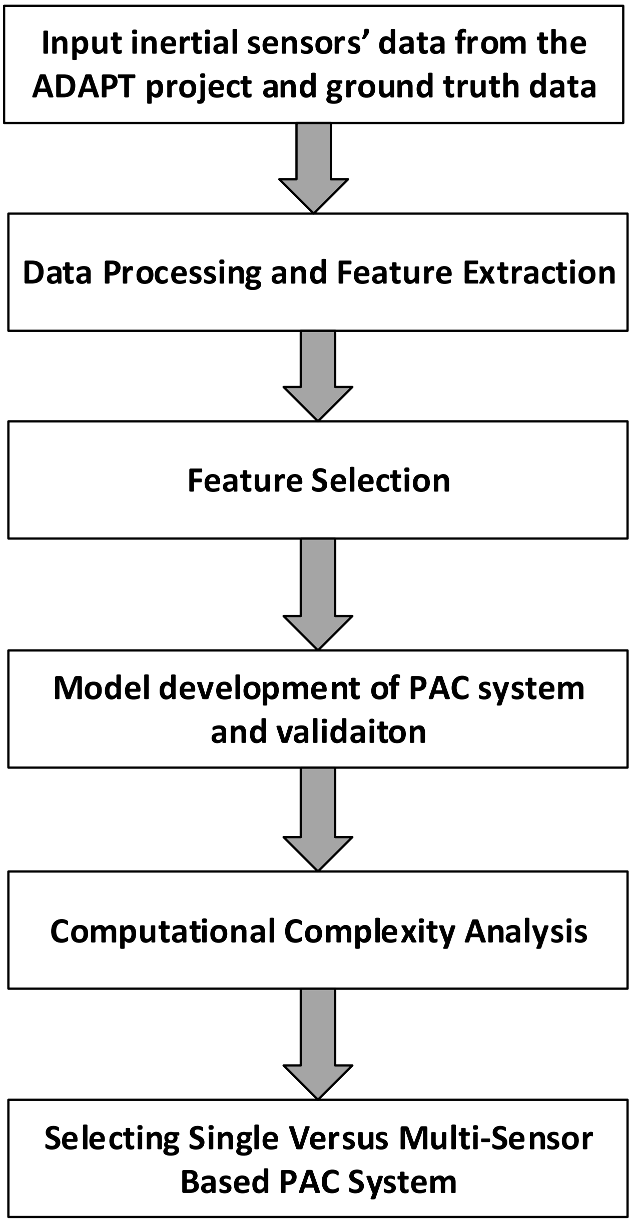


Figure 4.1: Flow diagram of the data analysis performed to develop PAC system

4.2 Materials and Methods

4.2.1 Data Collection in Free-Living Conditions.

This work utilized the free-living protocol dataset as described in Section 3.2.1 (Table 3.3) collected by the Norwegian University of Science and Technology (NTNU) under the ADAPT project. The 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. 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 4.1) without any instruction or supervision on how to perform them. Therefore, they could choose whether to perform these tasks (Table 4.1) or not and chose how and when they wished to perform the activities. The subjects performed the free-living protocol at their home environment resulting into more natural pattern and distributions of ADLs. Predetermined categories of ADLs used for the analysis were: sitting, standing, walking, transitions, shuffling, leaning, lying, ascending stairs, descending stairs, picking, leaning. The total length of recording were 28.7 hours for the 20 subjects [61]. Data from the wrist sensor was missing for four subjects due to technical issues during recordings, and these subjects were excluded from analysis. Consequently, the analyses have been performed on the remaining 16 subjects. The ADLs analyzed in this particular the study were: sitting, standing, walking and lying and the detailed summary of these ADLs is provided in Table 4.2. 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).

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 and a subset of these sensors were used in our analysis: chest (C), wrist (W), lower back (L5), and thigh (T) as shown in Figure 4.2. The synchronization between the sensors and the camera unit 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 accelerometer signal and by correlating this point in video, the synchronization between the camera and sensors was achieved.

Table 4.1: Free Living Unsupervised and Unstructured Task Based Protocol [61]

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
Climbing and 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 4.2: Characteristics of the total dataset of ADLs analyzed from the free-living conditions (N=16) as labelled from the video data

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

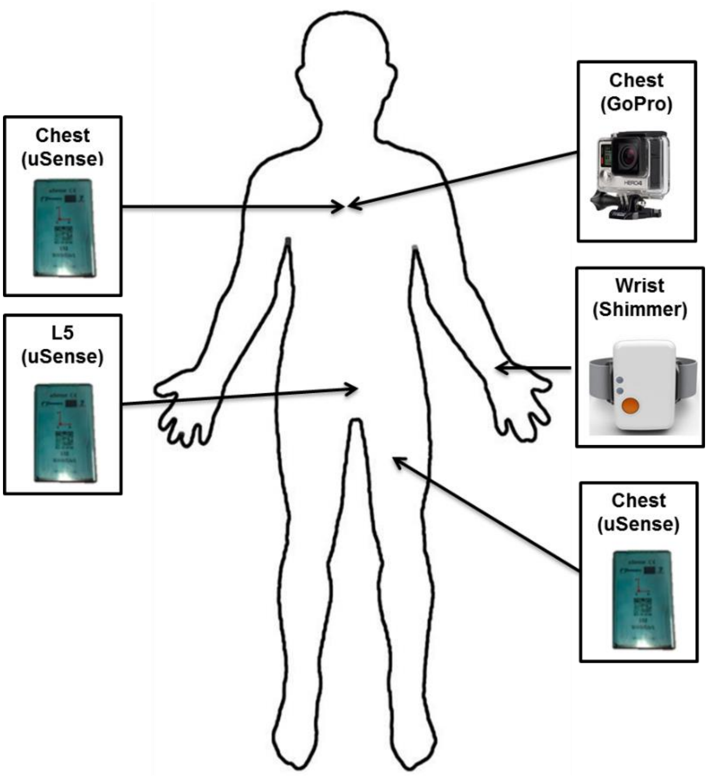


Figure 4.2: Sensors' placements chosen from the free-living protocol

Table 4.3: Description of the Sensors Used for Data Analysis [61]

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 /Recording time	72 h	11.75 days @, 10 Hz/4.6 days @ 1 kHz (450 mAh)
Sensor	3D accelerometer, gyroscope, magnetometer	3D accelerometer, gyroscope, 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 subjects were instrumented in the lab and they went home afterwards to perform the ADLs in free-living conditions. The detailed description of the sensors used for data analysis is presented in Table 4.3. 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.* [61].

4.2.2 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 100Hz 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 [29]. 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.

4.2.3 Features

Several features were extracted from acceleration and angular velocity (Table 4.4) which are described in detail in the following subsections. The computed features were of different categories: biomechanical features, statistical features, orientation free features and across sensor features. The aim was to collect maximum information from the sensors' data and then apply the feature selection process to select the robust features. Each of the features listed in Table 4.4 were computed across a time window of N samples (N=500, i.e. 5 seconds of data) with a 50% overlap. The letters x, y, and z in Table 4.4 represent the mediolateral, anteroposterior, and vertical axes, respectively. However, it is important to note that the sensor frame is moving in the world, so the axes of the sensors are approximately aligned with this body-centric axes.

Table 4.4: 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 gravitational acceleration (GA) component in vertical direction
15-17	Mean of GA components (x, y, z)
18	Mean of magnitude vector (MV) of bodily acceleration (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 show 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 69-77 were considered only if a sensor combination was analyzed.

4.2.3.1 Features Extracted from Acceleration

AThe mean, variance, and correlation between axes were computed from the raw acceleration (Table 4.4; 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 [30]. The mean of all three GA components [14] was used as a separate feature. The GA component was also used to compute the tilt angle [21, 30] from the expression below:

$$\mathrm{tilt}_{\mathrm{angle}} = \mathrm{acos}(z) \tag{4.1}$$

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) [30, 49], energy [22], and the magnitude vector (MV) [51] from the expressions below (Eqs. 2-4, Table 4.4: features # 10-20):

$$SMA = \frac{1}{N}\sum_{i=1}^{N}(|x(i)| + |y(i)| + |z(i)|) \tag{4.2}$$

$$Energy = \frac{1}{N}\sum_{i=1}^{N}(|X(i)|^2) \tag{4.3}$$

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

$$MV = \sqrt{x^2 + y^2 + z^2} \tag{4.4}$$

where x,y, z in Eq. 2-4 are from BA components. The mean, variance, and energy were then computed from the MV.

Note: The sqrt is monotonic and does 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 20Hz) 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 4.4; features # 21-33). Furthermore, the mean, variance, and energy were also computed (Table 4.4: features #: 34-36) from the MV (Eq. 4.4) of the jerk signal from acceleration.

4.2.3.2 Features Extracted from Angular Velocity

The mean, variance, correlation between axes, SMA, and energy (Table 4.4; features # 37-49) were extracted from angular velocity and jerk signal of angular velocity (Table 4.4; features # 53-65). The mean, variance, and energy of the MV from angular velocity (Table 4.4; features # 50-52) and MV from the jerk signal (Table 4.4, features #: 66-68) were also derived. The jerk signal was obtained by low-pass filtering (4^{th} order Butterworth low-pass filter with a cutoff frequency at 20Hz) the angular velocity and then taking its second derivative.

4.2.3.3 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 4^{th} order Butterworth low-pass filter with a cutoff frequency at 20Hz [62]. 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 [62]:

$$C_{ij} = (1 - \frac{RMS_j}{RMS_i}) * 100 \tag{4.5}$$

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 4.5) was available in the chosen sensor solution (e.g. 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 4.4; 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).

4.2.4 Class Distribution in the Dataset

The dataset [58] originally contained 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 [47]. The pie chart in Figure 4.3 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 seconds or 500 samples of data).

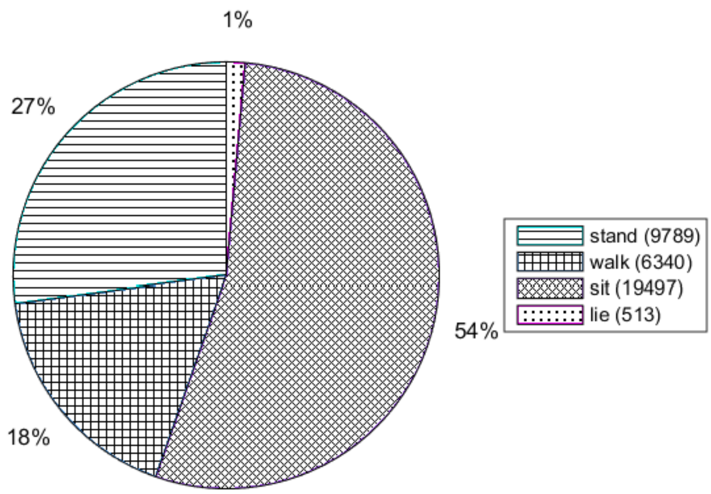


Figure 4.3: Percentage distribution of the four ADLs (sitting, standing, walking, and lying) for the 16 subjects in the dataset.

4.2.4 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, 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, the 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 correlation between features. A feature subset is determined by computing linear correlation [63].

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 [64].

The third method used for feature selection is ReliefF [65]. 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 [18].

4.2.5 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 the default settings [66]. 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. [67] 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.

Overall accuracy (A), F-measure, specificity and sensitivity by class (S_C) were computed as performance metrics using the expressions described in Appendix B. The accuracy measure is not the best metric to evaluate the performance in our dataset because of the of unbalanced class sizes. Thus, more balanced parameter F-measure was analyzed and is interchangeably with the term "performance" throughout the remainder of this paper.

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

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

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

4.2.6 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, the performance of 15 sensor combinations listed in Table 4.5 was analyzed and compared.

Table 4.5: 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	×	×	×	×

4.2.7 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 [47] 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.

Table 4.6: Performance Analysis of Multiple-Sensor Combinations

		F-measure (%)				Accuracy (%)			
No	**Sensors**	**ALL (SE)**	**CFS (SE)**	**FCBF (SE)**	**ReliefF (SE)**	**ALL (SE)**	**CFS (SE)**	**FCBF (SE)**	**ReliefF (SE)**
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)

All feature selection methods were implemented in MATLAB (Release 2014b, The Math Works, Inc., Natick, MA, USA) using the feature selection repository [68]. The SVM classifier was implemented using the LibSVM library [66] 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).

4.3 Results and Discussion

4.3.1 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 4.6 for the F-measure and for accuracy.

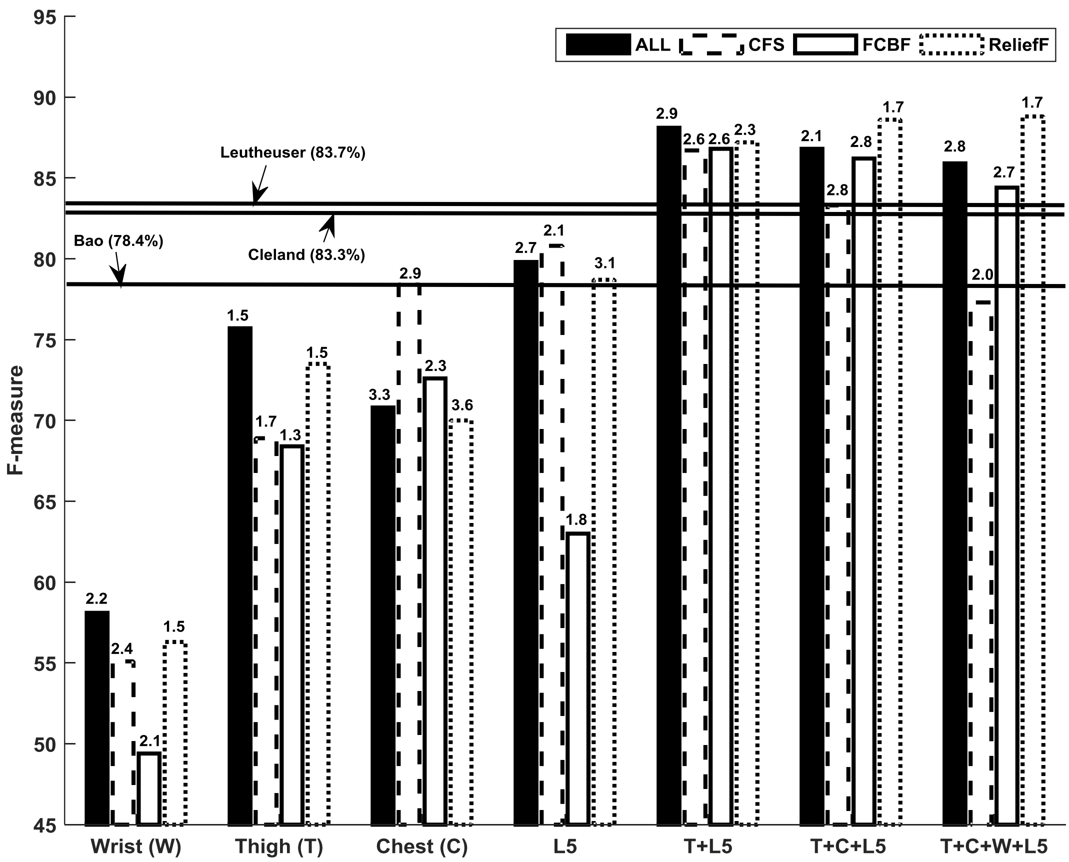


Figure 4.4: F-measure analysis using SVM Classifier with and without feature selection methods across various sensors combinations.

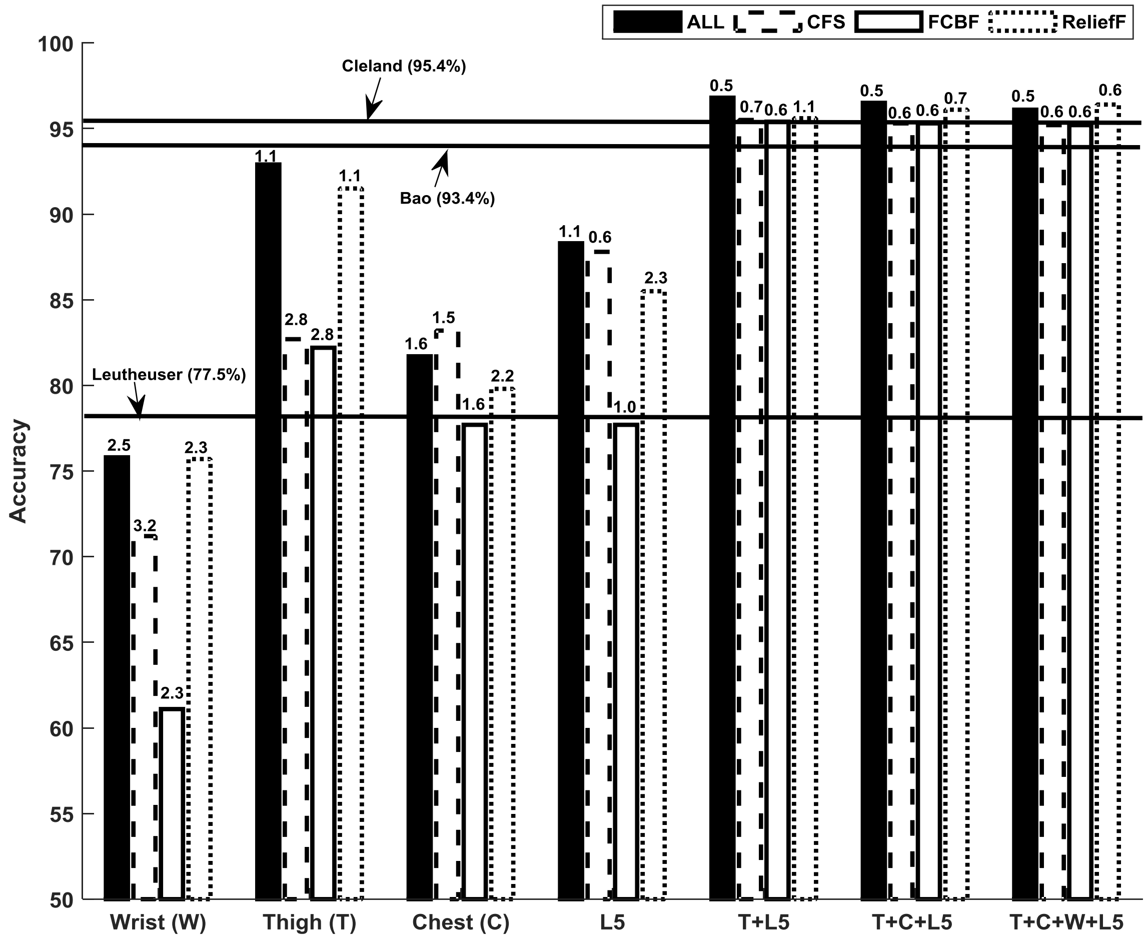


Figure 4.5: Accuracy analysis using SVM Classifier with and without feature selection methods across various sensors combinations

The F-measure for all single-sensor solutions and the best multi-sensor solutions (with 2, 3, and 4 sensors) are presented in Figure 4.4.

For every sensing solution, each of the four columns in Figure 4.4 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 in parentheses above each column show the associated standard error, as computed in Eq. 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 (Figure 4.4). 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 Figure 4.4, 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 solutions, beyond which the performance cannot be improved further even by increasing the number of sensors.

Similar kind of behavior is observed (Figure 4.5) without any improvement in accuracy even if the number of sensors is increased over two. However, having an unbalanced problem (Table 4.3), directly suggests that using accuracy as a metric is not appropriate. This is because the accuracy metric will not decrease significantly even if the under-represented class (lying in our case) is completely misclassified. Therefore, to avoid this, we used a more balanced and accurate metric F-which takes into account the instances of each ADL including the minority classes, while computing the performance.

The use of weighting scheme is also helpful if the intentions are to improve the classification rate of under-represented class, having small number of samples as compare to other classes. For example, if the goal is to classify the instances of moderate to vigorous activity as it is shown to be beneficial for health, but this activity is performed quite rarely around 2% of the day with bouts less than 30 s. This can be achieved by tuning the weights of each ADL in such a way that the trained PAC system accurately classifies the rarely performed ADL. Moreover, the short bouts of 30 s or below can easily be classified using the 5 s windowing process used in our PAC system.

4.3.2 Comparison with State-of-the-Art Systems

The performance of three representative systems for PAC tested in our earlier work [69] is also presented in Figure 4.4 (solid lines) to provide a direct comparison with the newly proposed system. All the three systems by; Bao et al. [9], Cleland et al. [22] and Leutheuser et al. [14] were implemented using the same dataset, type of ADLs, windowing approach, and cross validation procedure. The performances (F-measure, Figure 4.4) obtained by these systems are: 83.7% by Leutheuser et al. which uses three sensors (chest, wrist, L5), 83.3% by Cleland et al. which uses five sensors (chest, L5, wrist, waist, thigh) and 78.4% by Bao et al. which uses three sensors (L5, wrist, thigh).

The performance of our single-sensor based solution at L5 is 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) is much better than the state of art systems and still uses less number of sensors than these systems (3 or more).

4.3.3 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 Figure 4.4. 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 4.7. These results are computed across 16 folds and the corresponding mean and standard deviation reported for each of the seven systems. The highlighted text in Table 4.7 corresponds to the best feature selection method. The type of features selected by the best feature selection method are listed in Table 4.7.

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

Table 4.7: Statistics of the Features Selected by the Three Feature Selection Approaches for the Sensor Combinations Presented in Figure 4.4

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.*

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 [70], 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 (Figure 4.4). 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 (Figure 4.4). 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.

4.3.5 Computational Complexity of the System

The computational complexity of the best single-sensor solution was analyzed for a subject (all window instances) and a single window instance (consisting of 5 seconds 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 a subject containing 1741 instances are shown in Figure 4.6 and 4.7 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 (Figure 4.5 (a)) was low in the selected feature subsets compared to the time taken to compute the whole feature set. 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 and compared to Relief.

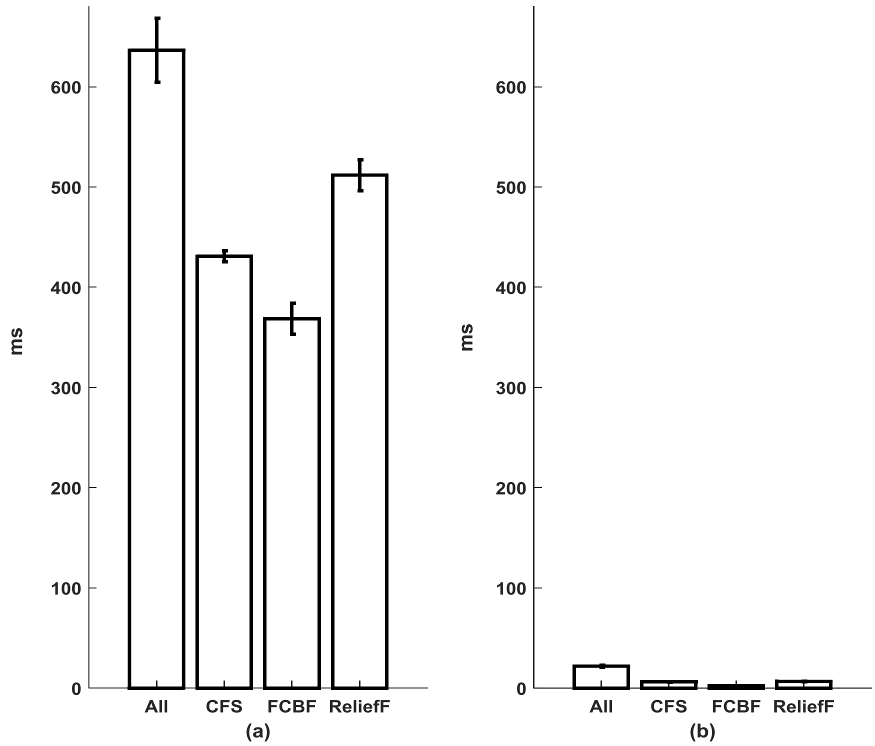


Figure 4.6: Computational complexity analysis of single window instance: a) feature extraction time, b) classifier testing.

Table 4.8: Computational Complexity Analysis of Single Window Instance

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

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 (Figure 4.5 (b), Table 4.8). 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 (Figure 4.6, Table 4.9) 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. The proposed PAC system was implemented on a personal computer and it would be interesting to see how computational complexity measures behave when implemented in mobile wearable platforms. Still, this is beyond the aims of this thesis and should be considered for future analyses.

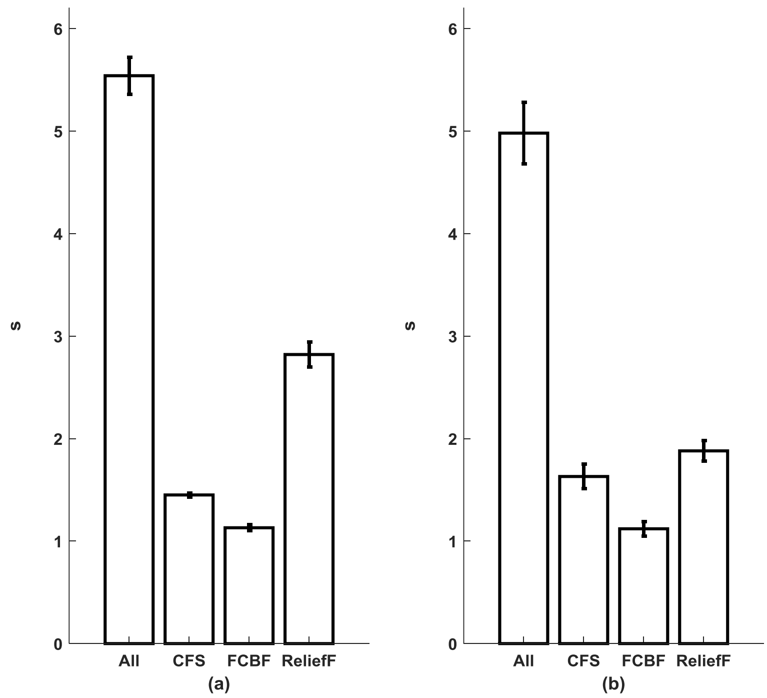


Figure 4.7: Computational complexity analysis of single subject's data: a) feature extraction time, b) classifier testing time.

Table 4.9: Computational Complexity Analysis for Single Subject

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

4.3.6 Single-Sensor vs Multi-Sensor Solution: What to Choose?

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 4.10 along-with F-measure for each case. The sensitivities and specificities are presented in Table 4.11. In the first solution, a sensor at the chest successfully classified walking and sitting but did not performed well in classifying standing and lying (Figure 4.8). 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 each class 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 (Figure 4.8). 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 of lying class are too small even with weighted SVM classifier.

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, a single sensor-based PAC system at the L5 is the best option to achieve the highest overall accuracy. 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.

Table 4.10: Confusion Matrix Using SVM Classifier (all features) for the Sensors at (a) Thigh (b) L5 (C) Chest (D) Thigh +Chest

F-measure **78.4%**	**(a) Chest Sensor (CFS subset)** Predicted Class				
Actual Class	**classified as →**	**walk**	**stand**	**sit**	**lie**
	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				
Actual Class	**classified as →**	**walk**	**stand**	**sit**	**lie**
	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				
Actual Class	**classified as →**	**walk**	**stand**	**sit**	**lie**
	walk	5688	498	154	0
	stand	421	9151	217	0
	sit	5	2	19470	20
	lie	0	0	256	257

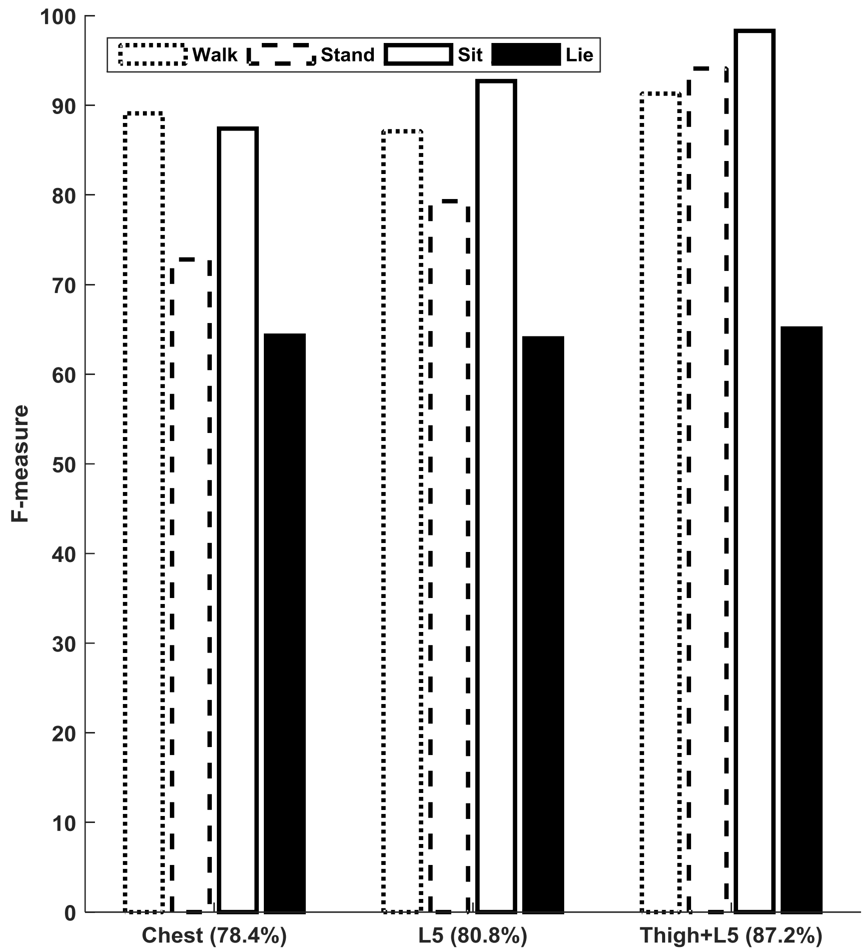


Figure 4.8: F-measure by class for three sensing solutions: a) Chest, b) L5, c) Chest, d) Thigh + L5. Value in parenthesis show the averaged F-measure.

Table 4.11: Sensitivity and Specificity by Class for three sensing Solutions

Sensor	Mean	Sensitivity by Class				Mean	Specificity by Class			
	Sens.	Walk	Stand	Sit	Lie	Spec.	Walk	Stand	Sit	Lie
Chest	84.5	91.4	74.5	84.8	87.3	93.4	97.1	88.8	89.1	98.8
L5	78.4	87.9	78.4	93.3	54.0	95.0	97.1	92.9	90.5	99.8
Thigh + L5	84.5	94.7	94.6	99.8	48.9	98.9	98.1	97.8	99.8	100

Sens.— Sensitivity, **Spec**.— Specificity.

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.

4.4 Conclusions

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 as well as the limited number of ADLs included. However, the dataset analyzed is among the largest of its kind, so far collected in free-living conditions for older adults and annotated manually with very high frequency of 25Hz (annotation every 0.04 s) [58].

Chapter 5

Predicting Unlabelled Activities of Daily Livings Using a Single Sensor Based Physical Activity Classification System for Elderly Populations in Free-Living Conditions

The findings of this chapter will be submitted as a journal publication in MDPI Sensors.

5.1 Introduction

Obtaining ground truth information in free-living conditions is not an easy task, especially when the aim is to perform the activity monitoring for a longer duration. This is because the commonly used labeling procedures such as video observation, marking by the experimenter/observer or subject itself [71, 72] have their own limitations and concerns [73]. The direct annotation methods such as video recordings or presence of observer is not always feasible because of ethical considerations and privacy issues. Although, these methods are reliable and accurate, the associated costs and resources make the labelling procedure time consuming and expensive. On the other hand, self-labelling by the subject is not as accurate and reliable as direct observation and it also interfere with the activities of the subject.

Conversely, the data collection of unlabeled data is much easier than the labeled data since it only requires a data collection device (smartphone/smartwatch/body-worn sensor) carried by the subject. Furthermore, another benefit is that the subjects can freely perform their daily life activities more naturally, without the need to self-label their ADLs or to be observed by somebody else.

Consequently, an effort has been made in this work to predict the unlabeled ADLs of daily living. This aim was accomplished by developing the single-sensor based PAC system on the labeled dataset (ADAPT dataset) and testing it on the unlabeled dataset of older adults (PreventIT) in free-living conditions. The sensor placed at the lower back (L5) was used from both datasets for data analysis. The PAC system developed in chapter 4 was adapted for being applicable to the specific dataset. A complete flow diagram of the data analysis performed is presented in Figure 5.1.

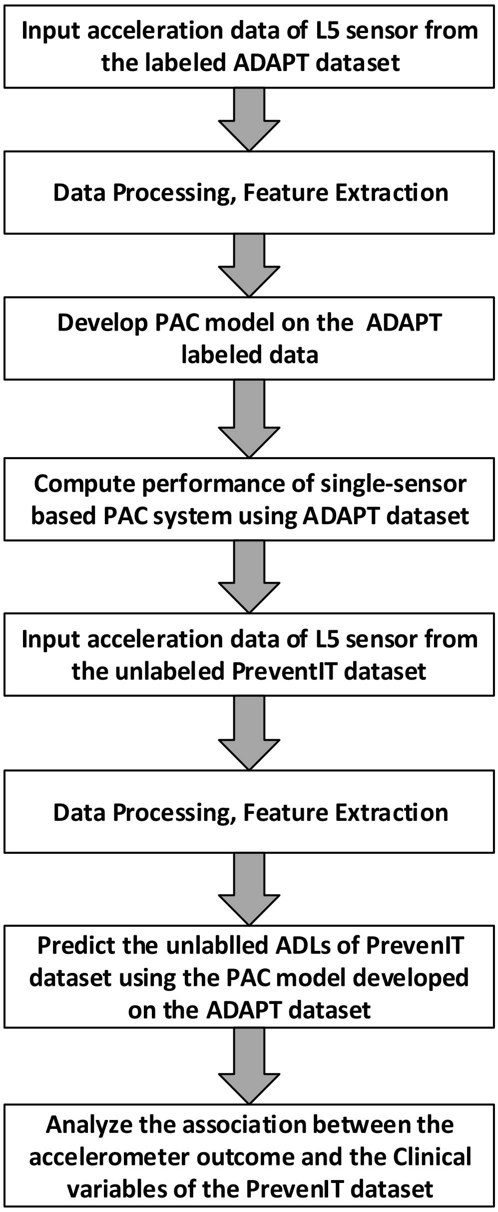


Figure 5.1: Flow diagram of the analysis performed to predict unlabeled ADLs

5.2 Materials and Methods

5.2.4 Data Processing and Model Development of PAC System Using the ADAPT Labeled Dataset

The ADAPT dataset was used to develop the PAC system model. The aforementioned dataset contains labeled ADLs, annotated by multiple raters using the video recordings and the brief description is provided in Chapter 3. The PAC system model is developed using only the sensor at lower back (L5). Several features were extracted from the acceleration signal of L5 sensor as described in Table 5.1.

Table 5.1: Feature Computed from Acceleration

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

The feature set is described in Table 5.1. Each of the features listed in Table 5.1 were computed across a time window of N samples (N=500, i.e. 5 seconds of data) with a 50% overlap.

5.2.2 Relabeling ADAPT Dataset Annotation into Hierarchical Way

The ADLs of the ADAPT dataset was divided into four main classes, i.e., active, sedentary, walking and lying as listed in Table 5.2. This was done to analyze the general profile of the subject in terms of active periods and sedentary periods throughout the days as these can provide better insight of the daily life activity patterns.

Table 5.2: Reassigning ADAPT classes into Hierarchical Distribution

Reassigned Class	Label	ADLs from ADAPT Dataset
Lying	1	Lying
Sedentary	2	Sitting
		Standing
Active	3	Shuffling
		Transitions
Walking	4	Walking
		Stairs Up
		Stairs Down

5.2.3 Performances Evaluation of Single Sensor Based PAC System Using the ADAPT Labeled Dataset

Before predicting the unlabeled ADLs, we evaluated the single sensor based PAC system to observe the performance on an annotated dataset of ADAPT collected in the free-living protocol (Chapter 4, Figure 4.3). For this purpose, we implemented SVM classifier and random forest (RF) classifier for performance computation with leave-one-subject-out cross-validation procedure. The four ADLs classes to be classified were: lying, sedentary, active and walking.

5.2.4 Data Collection of Unlabeled PreventT Dataset in Free-Living Conditions.

The unlabeled dataset was collected in the framework of the European project PrevenIT (http://www.preventit.eu/) from which the University of Bologna is one of the partner Institution. This dataset was collected in three different locations, i.e. Norwegian University of Science and Technology (NTNU), Trondheim, Norway; Vrije Universiteit (VU) Amsterdam, Netherlands, and Robert Bosch Hospital (RBK), Stuttgart, Germany. The dataset was collected in two sessions, pre- and post-intervention, in a group of young elderly subjects, 60 to 70 years old. Each session comprised of 7 to 8 days of continuous recordings of accelerometer data (sampling at 100Hz) by placing only one sensor at Lower back (L5) position. The L5 sensor did not embed a gyroscope due to practical considerations in terms of battery life. The intervention program consisted of a list of exercise that subjects performed at home. Not all the subjects were part of the pre- and post-intervention, since some subjects only participated in the pre- intervention session while others joined later and participated only in the post-intervention session. Furthermore, for some of the subjects, accelerometer recordings lasted for less than 7 days (2 days, 3 days etc.). Therefore, we shortlisted the subjects so that each subject had the accelerometer recordings in both sessions (pre- and post- intervention) for at least 6 days long. From the aforementioned criteria, a total of 16 subjects were shortlisted for the analysis of predicting unlabeled ADLs (Table 5.2) in free living conditions. The subjects performed their daily living tasks naturally in free-living conditions without any scripted guidelines on the sequence and act of performing ADLs. The accelerometer data was collected throughout the day, i.e., both for the daytime as well as for the nighttime when the subjects were sleeping.

5.2.5 Data processing and Predicting Unlabeled ADLs of PreventIt Dataset

Similar set of features was extracted from acceleration signal of L5 sensor of PreventIT dataset as described in Table 5.1. Each feature was computed across a time window of N samples (N=500, i.e. 5 seconds of data) with a 50% overlap.

The prediction of unlabeled ADLs was accomplished by building the PAC system model (Section 5.2.3) on the annotated dataset (ADAPT) and testing it on the unlabeled dataset (PreventIT) of older adults collected in free-living conditions. Then, activity intensity [74] (i.e., metabolic equivalents- METS) was computed for each class apart from calculating the activity duration/bout for each class and the number of steps/bout for walking class.

5.2.5.1 Acceleration Based Outcomes Computed from the Predicted Labels Obtained Through the Pre- and Post- Intervention Sessions of PreventIT Dataset

Pre- and post-intervention analysis was conducted on the predicted labels to observe if there exist any differences in the activity behaviors. For this purpose, several features were computed such as: the proportion of the measurement time [75] total activity time of each class/day [76], steps counts per day [77] etc. A list of the features analyzed for the pre- and post- analysis is presented in Table 5.3.

Table 5.3: Features Computed for the Predicted Labels of Pre- and Post-Intervention Sessions

Feature #	**Feature description**
*1-4**	the proportion of total activity time (for each class)
5-8	total activity time/day by taking the mean across 6 days (for each class)
9-12	total activity time/day by taking the median across 6 days (for each class)
*13-17**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of lying class duration/bout
*18-22**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of sedentary class duration/bout
*23-27**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of active class duration/bout
*28-32**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of walking class duration/bout
*33-37**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of active class intensity/bout
*38-42**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of walking class intensity/bout
*43-47**	5^{th}, 10^{th}, 50^{th}, 90^{th} ,95^{th} percentiles of walking class number of steps/bout
48	number of steps per day by taking the mean across 6 days
*49**	number of steps per day by taking the median across 6 days
*50**	total number of steps

* features were computed across 6 days, i.e., the total duration of the data analyzed

The appropriate statistical test was then applied to check the significance of each feature across the pre-and post-intervention. The data normality was checked using the Kolmogorov-Smirnov test [78]. Then, we selected non-parametric test, i.e., paired Wilcoxon signed rank test [79] since the data was not normally distributed.

5.2.5.2 Clinical Variables Obtained from PreventIT Dataset

We analyzed the associations between clinical variables and the accelerometer measures described in Table 5.3. The clinical variables have been divided into four categories: functional capacity, cognition, strength, and balance. Categories and variables are described in Table 5.4. Correlation coefficient and the statistical significance were computed between the acceleration based measures (Table 5.3) and the clinical measures (Table 5.4) using the spearman correlation [80]. We also analyzed the association with weight, height, and body mass index (BMI).

Table 5.4: Clinical tests analyzed for the correlation analysis

Category	Test
Functional Capacity	1- Time to complete gait 7 meters 2- Time to complete gait 7 meters (repeat) 3- Time to complete gait 400 meters
Cognition	Montreal Cognitive Assessment (MoCA) [81]
Strength	Chair stands (number of repetitions in 30s)
Balance	Fullerton Advanced Balance (FAB) scale [82]

5.3 Results and Discussion

5.3.1 Performance Analysis of the Single Sensors Based PAC system on Labeled ADAPT Dataset

The results obtained from the performance analysis of L5 sensor of ADAPT dataset are presented in Table 5.5. The overall accuracy of both classifiers was quite good (around

90%), considering the single sensor solution and only the features from acceleration signal. The random forest classifier performed slightly better than SVM, but the difference in the performance was quite negligible (1.1% improvement). To get a better insight of the classifiers' performance against each class, we computed the accuracies by class and sensitivities by class as shown in Table 5.6. The sensitivities by class provide better knowledge of the classifiers' effectiveness in terms of identifying true positive labels.

For better understanding, sensitivities by class for both classifiers are depicted in Figure 5.2 against each class, i.e., sedentary, walking, active and lying. The sensitivities by class of both classifiers were quite close to each other in classifying sedentary and lying classes. However, random forest classifier performed better than SVM in classifying walking and active class with a difference of 3.8% and 8.2% respectively. Therefore, random forest classifier was selected to perform the prediction of unlabeled ADL.

Table 5.5: Performance analysis using ADAPT dataset (a) SVM Classifier(b) Random Forest Classifier

Accuracy **88.67%**	**(a) SVM Classifier**				
	Predicted Class				
Actual Class	**Classified as →**	**Sedentary**	**Walk**	**Active**	**Lie**
	Sedentary	28241	466	419	160
	Walking	758	5875	382	15
	Active	1497	574	1127	14
	Lying	254	0	3	256
Accuracy **89.68%**	**(b) Random Forest (RF) Classifier**				
	Predicted Class				
Actual Class	**Classified as →**	**Sedentary**	**Walk**	**Active**	**Lie**
	Sedentary	28113	428	694	51
	Walking	541	6154	335	0
	Active	1235	583	1392	2
	Lying	255	0	6	252

Table 5.6: Accuracy by Class and Sensitivity by Class for SVM and RF classifiers using single sensor at L5

Classes of ADLs	Accuracy by Class		Sensitivity by Class	
	SVM	RF	SVM	RF
Sedentary	91.1	92.0	96.4	96.0
Walking	94.5	95.3	83.7	87.5
Active	92.8	92.9	35.1	43.3
Lying	98.9	99.2	49.9	49.1

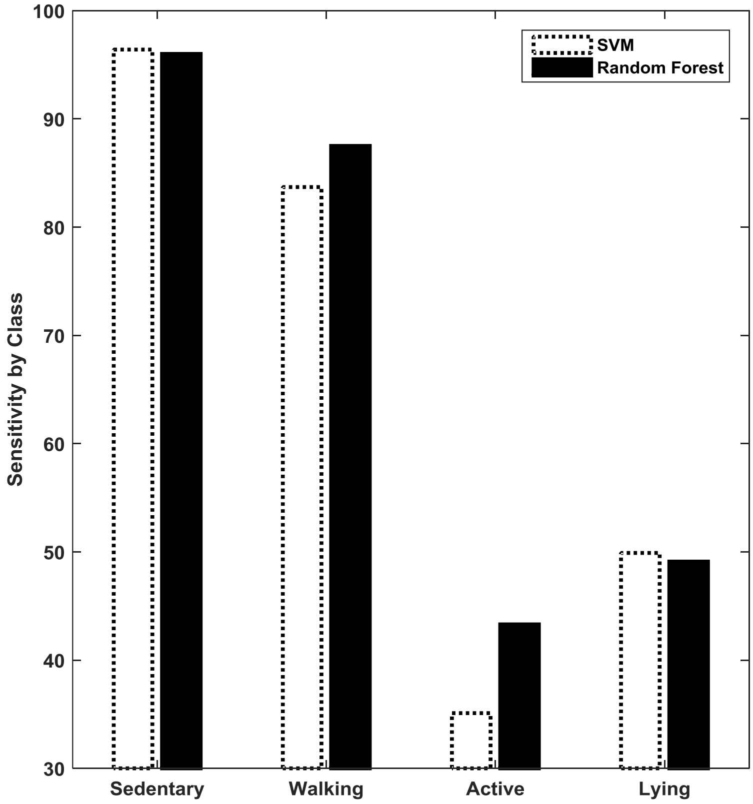


Figure 5.2: Sensitivity by Class Using SVM and Random Forest Classifier

It is important to highlight that both classifiers were not able to perform well in classifying lying and active class (Figure 5.2). The reason for misclassifying lying is possibly the low data samples of this class as stated earlier (chapter 4, Figure 4.3). Although there are resampling techniques [83, 84] which can be adapted to overcome the class unbalancing. However, the ADAPT dataset closely reflects the proportion of activities that can happen in real life conditions.

The second class is active, whose performance was the worst among all classes. The possible reason could be the diverse nature of this class since it is composed of transitions and shuffling. This class was mainly confused with the sedentary class and the walking class (Table 5.5). The reasons for this behavior are twofold. Firstly, the type of transitions included inside active label was composed of more than 50 types of transitions [58], e.g., stand to lean, stand to lie, stand to pick, sit to stand, etc. Thus, accurate classification of this diverse nature transitions using a single feature set with machine learning approach is not an easy task. Secondly, the inclusion of shuffling activity inside active class, since shuffling also comprised stepping in place and feet movements on the spot, which could have been confused with walking.

5.3.2 Classifying Unlabeled ADLs of the PreventIT Dataset Using the PAC System Model Developed on the ADAPT dataset

The model of the PAC system trained on the ADAPT dataset was applied to the PreventIT dataset to predict the unlabeled activities into one of the four classes described in Table 5.2.

The raw dataset and class predictions obtained from one of the subjects analyzed from PreventIT dataset are shown in Figure 1 (a) and (b) respectively. The time axis is in hours and the analyzed data length is about one-week time. We also implemented conditioning on the nighttime predictions. This is because the ADAPT dataset contained very few instances of lying as compare to sitting, standing and walking instances and none of the lying periods were recorded at nighttime while per subject was sleeping. Nighttime was defined as the time window between 11:00 PM and 6:00 AM, except for walking bouts activities are labelled as lying at nighttime. Apart from this condition, there were no other conditions defined for the rest of the analysis.

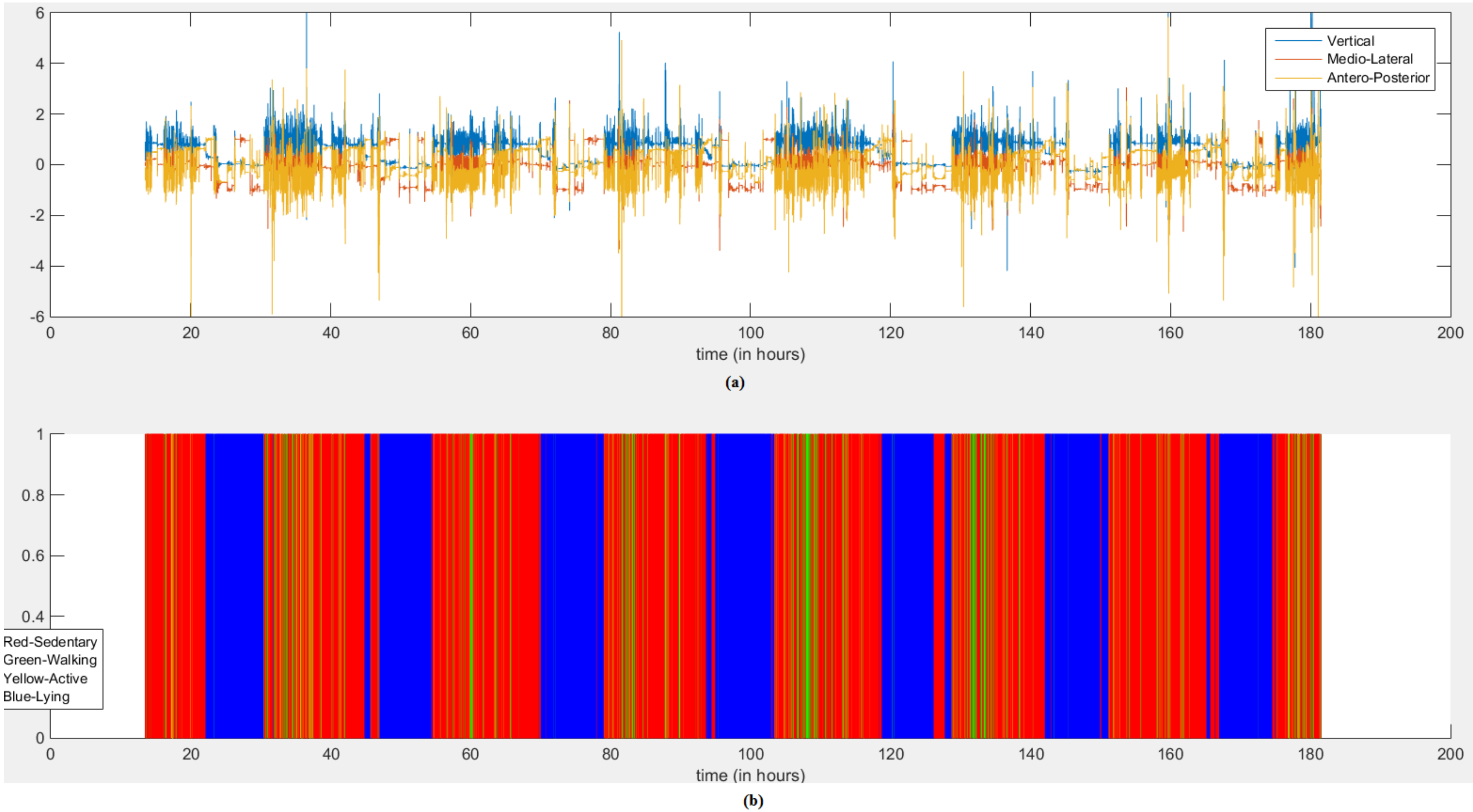


Figure 5.3: (a) Raw accelerometer data of L5 sensor collected across one week of recordings (b) ADLs predictions using color coding, i.e., Blue- Lying; Red- Sedentary; Active- Yellow; Walking- Green

5.3.2.1 Analysis on Predicted Labels Obtained Through the Pre- and Post-Intervention Sessions of PreventIT Dataset

We also went one step further to extract the patterns form the predicted labels (Figure 5.3) to observe if there exists any difference in the activity proportions of each class in the pre- and post- intervention sessions.

The proportion of the predicted classes for a single subject are depicted in Figure 5.4 (a) and (b) for the pre- and post-intervention sessions (measurement time of each session is 6 days), respectively. The proportions can provide a general knowledge of the subject's activity profile before and after the intervention. However, it cannot provide a clear indication of significant increases or decrease in the activity distribution of each class across pre- and post- sessions. Thus, paired Wilcoxon signed rank test was applied on the feature-set (Table 5.3) obtained from the pre- and post- interventions session. The statistical analysis showed that only one feature was significant ($p<0.05$) between the pre- and post- sessions, i.e., total sedentary time per/day computed by taking the median across 6 days. These results are listed in Table 5.7, which shows the mean and standard deviations of the total sedentary time/day (minutes) of 16 subjects in the pre- and post- sessions. The analysis highlighted that there was a significant decrease in the total sedentary time/day of about one hour (Table 5.7) between the pre- and post- intervention sessions.

Table 5.7: Pre- and Post- Intervention Sessions Statistical Findings

Feature	**Pre- Intervention (mean ± std)**	**Post- Intervention (mean ± std)**	***p*-value**
Total sedentary time per day (median across 6 days)	678.84 ± 100.94 min	607.68 ± 159.71 min	0.03

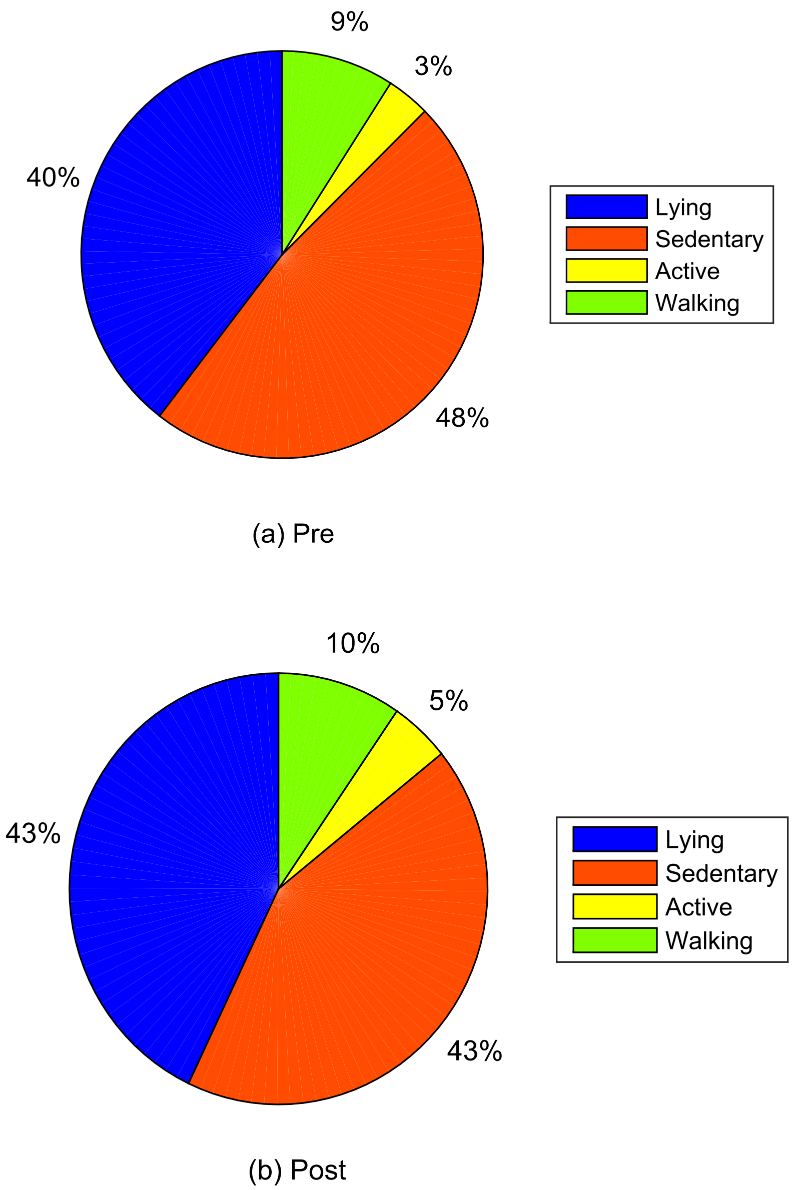


Figure 5.4: Proportion of the predicted classes of a single subject in (a) Pre-intervention session (b) Post-intervention session using the PreventIT dataset

5.3.2.2 Correlation Analysis Between the Acceleration Based Measures and the Clinical Variables of PreventIT Dataset

The findings of the correlation analysis performed between the clinical variables and the acceleration measures are presented in Table 5.8. Results are presented only for the significant pairs along-with their *p-values* and correlation coefficients.

Table 5.8: Analysis between the clinical measures and the accelerations based measures of PreventIT dataset

Class	Feature (*p*, *r*)	BMI (*p*, *r*) *	Weight (kg) (*p*, *r*) *	Gait 400m (*p*, *r*) *	Chair Stand (*p*, *r*) *	FAB Scale (*p*, *r*) *
Walking duration across 6 days	**Percentage distribution**	0.045, -0.564	✕	✕	✕	✕
	Duration per day (median across 6 days)	0.042, -0.569	✕	✕	✕	✕
	Duration per day (mean across 6 days)	0.045, -0.564	✕	✕	✕	✕
Intensity per bout† of active class (across 6 days)	**10th percentile**	✕	✕	✕	✕	0.042, 0.570
	50th percentile	✕	✕	0.043, -0.577	✕	0.040,0.575
	90th percentile	✕	✕	0.045, -0.571	✕	✕
	95th percentile	✕	✕	0.043, -0.577	✕	✕
Walking class bout duration (across 6 days)	**95th percentile**	✕	✕	✕	0.016, 0.653	✕
number of steps/bout (across 6 days)	**95th percentile**	✕	✕	✕	0.021, 0.630	✕
Total steps/day	**Median across 6 days**	0.024, -0.619	0.027, -0.608	✕	✕	✕

* *p* is the significance value, *r* is the correlation coefficient, † bout: uninterrupted period of any specific activity being considered

The correlation between the BMI was significant with the total walking duration and the total number of steps. Both these associations exhibited negative correlations, emphasizing on the fact that an increase in the body mass index has a negative influence

on the walking activity of the subjects with a significant decrease in the total walking duration and the number of steps per day. Similar findings were obtained between the body weight and the number of steps, showing that high body weight has a negative impact on the number of steps. It is likely that the association between the weight/BMI and walking time and number of steps is within a vicious circle since and increased weight can cause a reduction of walking time/number of steps and at the same time a reduction of walking time/number of steps can lead to an increased wright/BMI,

The 400m walk test used as a measure of functional capacity showed statistical significant association with the 50th, 90th and 95th percentiles of the active class' intensity per bout. All these acceleration measures were negatively correlated with the duration of 400 m test. These findings suggested that the more it takes to complete the 400m test, the lower functional capacity is; a lower function capacity would also have an impact on the intensity of the active class. The cognitive measure MOCA was not significantly correlated with any of the acceleration based measures.

The chair stands test used as a measure of strength was significant and positively correlated with the 95th percentiles of the walking duration/bout and the number of steps/bout. These results are also coherent, as the higher is the number of chair stand repetitions the higher is the time the subject can walk continuously. Similarly, for the number of steps in a waking bout.

The association of the FAB scale was statistically significant with the 10th and 50th percentiles of the intensity/bout of active class. A high FAB value corresponds to a better balance. The positive correlation suggests that a better balance allows the subject to perform the activities with a generally higher intensity.

The labelling procedure used for some of the hierarchical classes is in contrast with the literature. A terminology consensus project conducted by the sedentary behavior research network (SBRN) [85] provides a standardized definitions of various terminologies: physical inactivity, sitting, lying, reclining, sedentary behavior, stationary behavior, screen and non-screen time based sedentary time etc. The labelling procedure presented by SBRN should be adapted in the future studies to maintain a common taxonomy.

This study comes with some limitations. One of the major limitations is the sample size. The samples size used in this study is 16 which is insufficient for the generalization of the findings. Thus, use of a relatively large sample size in the future can provide a better overview of the proposed method. Secondly, the PAC system trained in the annotated dataset contains very diverse types of transitions under one category (active class) which can be avoided in the future either by creating more robust feature-set to classify transitions or by reducing the transitions' types to only those which are most commonly performed in real life conditions. This will improve the prediction capabilities of the activity classification system when tested in free-living conditions.

5.4 Conclusions

This chapter presents a single sensor based PAC system to classify the unlabeled ADLs of young older adults in free living conditions. Initially, the performance of PAC system is verified on the ADAPT annotated dataset using only the L5 sensors and only the accelerometer, and an overall accuracy of about 90% is achieved using the random forest classifier. Then, the same training model built on the ADAPT dataset is tested on the PreventIT dataset to predict the unlabeled ADLs. The predictions are performed both for the pre- and post- intervention sessions.

We analyzed several statistical features from the predicted labels to observe if there exists any difference in the distribution of activity classes during the pre- and post-interventions sessions. The statistical analysis found a significant decrease in the total duration of sedentary class/day, suggesting that the PreventIT intervention program had a positive effect in terms of reducing sedentariness. Furthermore, the correlation analysis between the clinical variables and the acceleration measures was also performed. The finding showed that high BMI values are negatively correlated with the total walking duration as well as the number of steps/day. Furthermore, the associations found between the clinical variables and acceleration measures were also coherent i.e.; 1) increase in the time require to complete gait 400m is associated with the decrease in the active class intensity/bout, 2) high number of chair stands performed in 30s are positively correlated with the increase in the walking bout duration and the number of steps, 3) high values of FAB scale are associated positively with the increase in the intensity/bout of active class. These findings suggest that the accelerometer based measures can be helpful in determining the health profile of the elderly population in unsupervised settings.

Chapter 6

A Pilot Study to Develop Automated Video-Based Labelling Procedure for Activities of Daily Living

Some of the material presented in this chapter is taken from our earlier published work [86].

6.1 Introduction

The most commonly used and precise method to obtain ground truth information of the ADLs is by capturing the video information [8, 11, 12, 29]. However, this approach consumes resources and is very costly. It also adds bias in the video annotation process as video marking is done by different raters. Therefore, inter-rater reliability must be carefully investigated. Although the direct video recordings raise privacy concerns, these can be avoided or reduced to a minimal level by designing the video capturing system in such a way that it only captures the relevant information and ignore the surrounding objects in the frame, e.g., third person view.

There are various systems developed in the literature using the image processing and computer vision techniques to classify human activities. However, they focused mainly to detect either the actions performed by the third person view in the scene or the gestures/activities performed during the hand movement of the first-person view [87-92]. These systems ignored the commonly performed ADLs in free-living conditions, i.e., sitting, standing, walking, ascending stairs, descending stairs, lying, etc. An effort has been made by Kim et al. [93] to automatically detect the walking activity of the subject using a first-person view camera pointing downwards toward the feet of the subject. However, they only focused on the detection of walking activity and other commonly performed ADLs were not detected.

This study aims to automatically label/classify these ADLs in free-living conditions without the presence of an observer in the field or the help of a rater to perform the offline video marking/annotation. The successful development of this the proposed system will provide a state of the art solution for video annotation, which make it possible to simultaneously validate the performance of wearable sensor based activity classification system.

6.2 Materials and Methods

6.2.1 Data Collection and Experimental Protocol

The data collection was performed at the Department of Electrical, Electronic and information engineering (DEI), University of Bologna, Italy. Three subjects aged between 25 to 30 years participated in the pilot study. A relatively small sample size is chosen for data collection to check the feasibility of the proposed methodology before implementing on a larger population. Various types of equipment were used in the data collection procedure: IMU sensors, GoPro cameras (GoPro, Inc., San Mateo, CA, USA, 1920x1080 pixels), and smartphone Samsung Galaxy S3® mini (Samsung Electronics Co., Suwon, Republic of Korea).

Table 6.1: Description of the IMU Sensors Used for Data Collection

Device	**uSense**
Location	Thigh, L5, Chest
Size	67 × 42 × 10 (mm)
Weight	36 g
Sampling frequency	100 Hz
Battery Life /Recording time	72 h
Sensor	3D accelerometer, gyroscope, magnetometer
Measurement range	±2 g, ±250_/s, ±1200 _T
Company/ Institution	University of Bologna, Italy

Three IMU sensors were placed at three different location; lower back (L5), right and left feet and the details are presented in Table 6.1. The IMU sensing device is comprised of accelerometer, gyroscope, and magnetometer. The sampling frequency was 100 Hz for IMU sensor and 60fps for the camera units. The Smartphone acted as a sink device to wirelessly collect the data for all IMU sensors. The experimental set-up is shown in Figure 6.1.

Two camera units were used; one was placed at the chest of the subject for the first person's view pointing downward towards the subject’s feet, and the other unit was carried by the observer for the third person view. The video recordings obtained from the first-person camera unit were processed only to develop automated video labeling method. Three different shapes (circular, square and triangular) of markers were placed on both legs, with a different color on each leg. The markers’ location was; feet, shank, and thigh.

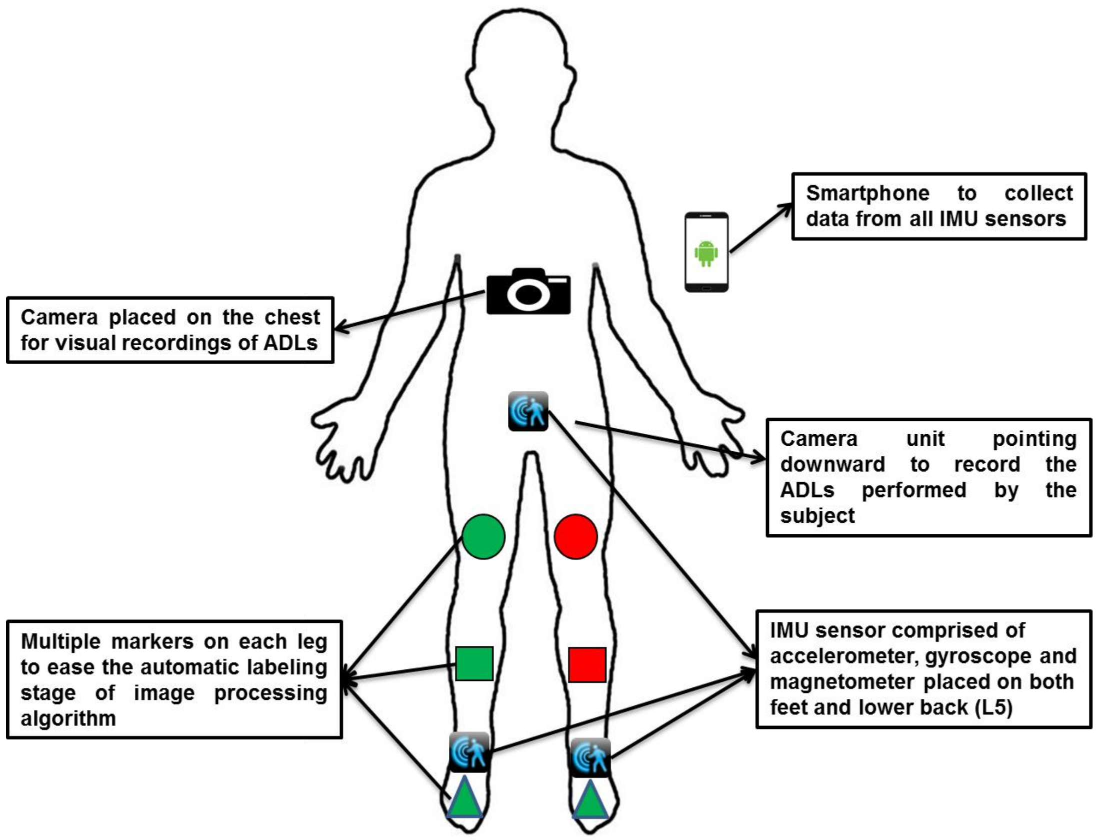


Figure 6.1: Placement of IMU sensors, wearable camera unit for data acquisition and multiple markers

Figure 6.2: Sequence of ADLs followed by the subjects during data collection

These markers aimed to facilitate the image processing algorithm in recognizing the type and duration of each ADL activity performed.

The experimental protocol consisted of a sequence of ADLs performed by the subjects as shown in Figure 6.2. The ADLs performed by the subjects were: sitting, standing, walking, ascending stairs, descending stairs and various transitions. The estimated length of the experiment was 7 minutes. Synchronization was performed twice between the IMU sensors and the camera units, once at the start of the experiment and then at the end of the experiment. This was accomplished by putting all three sensor units on top of one another and hitting them from an object (marker pen) thrice in the view of camera units. After the synchronization, the sensors and camera units were attached to the subject. The movements were evident in the acceleration signals as well as in the video recordings.

6.2.2 Data Analysis and Description of Methodologies

The proposed methodology for the data analysis is shown in Figure 6.3. The inertial sensor based PAC system is the same as we proposed in Chapter 4. The image processing based PAC system is the one that replaces the traditional offline video labeling procedure (e.g., the video annotation process used in the ADAPT project with the help of raters, Chapter 3).

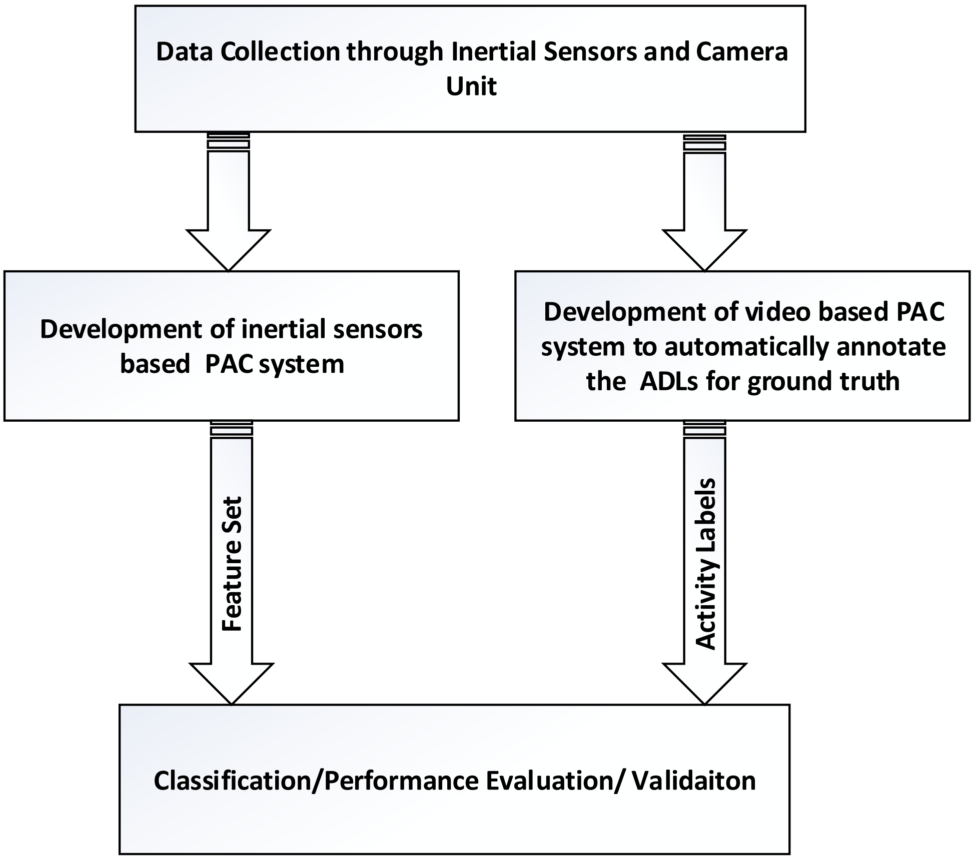


Figure 6.3: Flow diagram of the proposed methodology for the development of PAC system

6.3 Pilot Results and Discussion

The development of the inertial-sensor based PAC system is already achieved in Chapter 4, where the findings are quite encouraging and suggests to use a single/multi sensor based activity classification system to accurately classify the ADLs of elderly population in free living conditions. The same PAC system will be used to replace the left column of Figure 6.3.

The development of image processing-based PAC system includes preprocessing of the images as well as implementation of object detection algorithms. The raw images obtained through the first-person view camera are shown in Figure 6.4 for each of the ADL performed by the subject.

The image processing step was initiated by first converting the raw RGB (red, green blue) into the HSV (hue, saturation, value) image. The color space of HSV is less variant to noise as compared to RGB image and is commonly used when the aim is to separate the color components of the image from the intensity components. This is helpful in light varying conditions and to remove the objects' shadows. The raw RGB image and HSV images are presented in Figure 6.5 (a) and (b) for the sitting posture.

Then, the selected thresholding values each of the hue, saturation and value spaces was applied to separate the colored objects from the environment (Figure 6.5 (c)). Border smoothing of the detected objects and masking was applied. Finally, the color objects detected were mapped again to the RGB image as shown in Figure 6.5 (d).

In most of the ADLs, the image processing algorithm was able to detect the markers quite reasonably as shown in Figure 6.5 (d)-(h). However, there were also other objects detected in the image that were not of interest. This is because of the fixed thresholding of HSV parameters for all activities which means that fixed thresholding cannot be a permanent solution to detect all ADLs. Furthermore, there were no conditioning applied to the images to get rid of the connected components (objects) that are smaller than certain pixels. Therefore, further investigation is required to apply the appropriate algorithm for marker detection and to implement a connected component approach to get rid of the irrelevant objects in the scene.

a) Walking (green marker)

b) Walking (red marker)

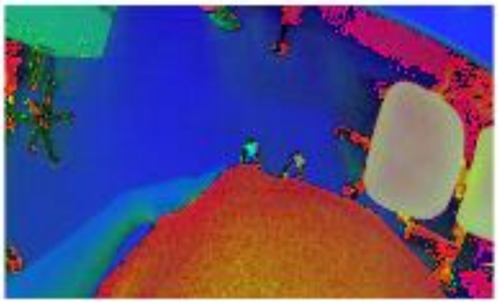
c) Sitting

d) Standing

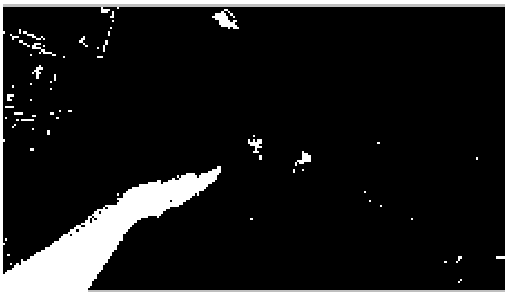
e) Ascending stairs

f) Descending stairs

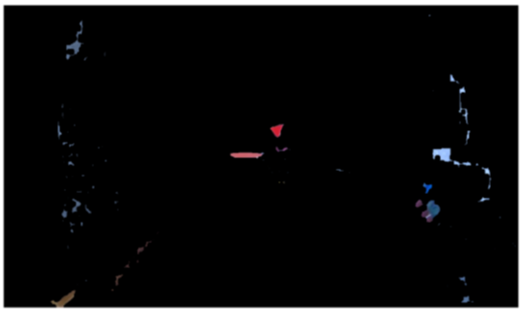
Figure 6.4: First person camera view of the various ADLs performed by the subject



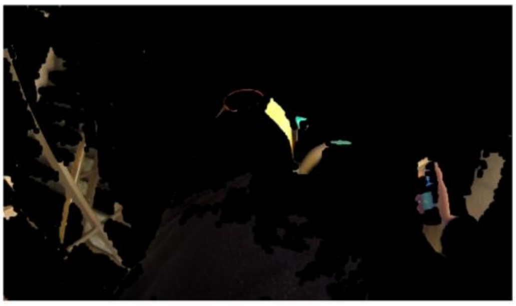
(a) Original Image (Standing) **(b) HSV Image (Standing)**



(c) After Thresholding **(d) Mapping and Markers Detection**



(e) Sitting **(f) Walking**



(g) Ascending Stairs **(h) Descending Stairs**

Figure 6.5: Various stages of the image processing algorithm and along with the marker detection for each ADL

There were also other important issues which we did not expect during the data collection, but we faced at the image processing stage. During the walking episode or the ascending/descending stairs activity, the image patterns were heavily blurred (2 out of 3) due to the to the relatively fast motion of a feet when the person is moving compared to the acquisition time of GoPro camera. This issue is particularly related to the exposure time of the camera and can be fixed by lowering the exposure time as much as possible. However, lowering exposure time will make the image much darker which can be controlled by increasing the camera gain. The ideal adjustment of these two parameters could be performed in such a way that the obtained images are slightly or not blurred for moving objects, and at the same time are neither too dark nor too bright in different lighting conditions and less noisy.

The second issue we encountered was the lighting variations. During the data recording of the subjects, it happened quite often that the lighting conditions were quite fluctuating (bright, dark, moderate) in the hallways. Due to these extreme variations, the performance of the color detector based image processing algorithm was highly affected. In the future trials, lighting variations should be limited by performing the data collection in the areas where the lighting conditions are not very bright as well as not very dark. Then, further trials can be conducted in light varying conditions, after the successful development of the image processing algorithm in controlled lighting environment.

6.4 Potential Applications

The automated video-based labeling method can be utilized in various domains other than the area of physical activity recognition. This is mainly possible because of the reliability of this method in terms of providing accurate ground truth, non-fixed nature of the first-person view camera and the placement of the camera unit. Firstly, the reliable nature of ground truth has a significant impact, since the ground truth keeping by the subject’s self-observation or by the help of external observer affects the natural flow of the performed activities in real life conditions. Secondly, the non-fixed nature is important, because the fixed camera-based systems (e.g., wall mounted) limits the usability only to the indoor controlled environment. Furthermore, these systems then require multiple camera units to be placed in each portion of the facility if the intended application is in the home environment. Lastly, the placement of first-person view camera towards the subject's feet reduces the privacy concerns and ethical consideration because of not capturing the irrelevant information and objects in the scene.

The successful development of the automated video-based labeling procedure has many implications for health care. For instance, the image processing based event detector can be utilized for the surveillance of epilepsy patients as well as the validation of epileptic seizures. The performance of wearable sensors based epileptic seizure detector can be validated through a first-person camera based event detector. Similarly, the video-based event detector can also be used for the detection and validation of freezing events in Parkinson Diseases (PD) patients and the detection of fall events in the elderly population. For these applications, the image processing based approach can be used alone or in an integrated manner by providing the ground truth information for the validation of the wearable sensor-based event detector.

Chapter 7

Conclusions/Final Remarks and Future Directions

7.1 Conclusions/Final Remarks

This dissertation has mainly focused on the development of wearable solutions for the activity classification of elderly populations. The state of the art methodologies for inertial sensors based physical activity classification were deeply reviewed to get an adequate insight of the strengths and weaknesses of the existing systems. Then, we proposed a benchmark approach to analyze the performance of the state of the art methodologies in an unbiased and fair way. The findings suggested that the performance of the existing systems is highly deteriorated when a laboratory-trained system is tested in free-living conditions. This analysis also highlighted that the newly developed systems should be trained and tested on the dataset collected in real-life, where the activities are performed in a more natural and unstructured way.

The gaps and limitations inferred by the benchmark study were addressed in the development of a novel physical activity classification system for older adults in free-living conditions. The performance of single-sensor and multi-sensors solutions were analyzed by implementing the filter-based feature selection approaches on the feature-set obtained from the acceleration signals and the angular velocity signals. The findings showed the potential of different solutions (single-sensor or multi-sensor) to accurately classify the ADLs of older people in free-living conditions.

Then, a single sensor based physical activity classification was developed to predict the unlabeled activities of daily living. This aim was due to the fact that collecting a large amount of unlabeled data in free-living conditions is quite straightforward because of the availability of IMU sensors in almost every smartphone and smartwatch. This study was implemented on a relatively small sample size and will be validated in future studies in a broader database.

Lastly, a preliminary study was proposed to automatically label the activities of daily living using the first person based camera system. Obtaining ground truth information using supervised methods is costly and time-consuming. Therefore, successful implementation of image processing based activity labeling method will help to save time and money, and to more easily test and/or (re)design systems for automated physical activity classification. The feasibility study faced several challenges regarding data collection and data processing which will be addressed in the future study to achieve the intended goals.

The wearable sensors-based activity classification system has implications in clinical practice as well as in home environments. Our findings regarding the single sensor-based system could enable real-time decision making, by implementing the activity classification algorithm in smartphones which will gather data from built-in IMU sensors. The activity patterns obtained from such activity classification systems can guide the healthcare practitioners to make informed decisions about the physical conditions and the onset of several diseases in elderly. These patterns can also guide the general population to adopt active and healthier lifestyle by observing the active and sedentary periods of long-term recordings.

7.2 Possible Future Directions

1. The set of ADLs analyzed in the developed PAC system were limited to four (sitting, standing, walking, lying) since these are the commonly performed activities by the elderly population. However, it would be interesting to see how the developed system behaves when the ADLs to be classified are scaled up with more complex activities e.g. transitions (sit-

stand, stand-sit, stand-walk, walk-stand, sit-lie, lie-stand etc.), shuffling, leaning, stairs-up, stairs-down etc.

2. The ADAPT dataset analyzed in this dissertation is unique of its kind since it is the largest dataset so far collected in free living conditions for older adults and ground truth is maintained with very high frequency (annotations every 0.04 s). The subjects were not supervised to perform a certain activity more frequently than others due to the nature of the free-living protocol. As a consequence, the data samples of lying were less than those in the other three classes, i.e. sitting, standing, walking. In the future studies, it is important to collect the dataset for longer duration (couple of days or more) to capture sufficient samples of each activity class, for better generalization of the developed PAC system.

3. The use of deep-learning based approaches in health informatics has grown rapidly in recent years due to the advancements in computational power and data storage devices. These methods provide an automatic feature set, derived directly from the raw data to extract complex behaviors instead of using hand-crafted features with human intervention [94]. Therefore, it would be interesting to see how deep-learning based PAC system will classify the activities of older adults in free living conditions.

APPENDIX A

Computation of Performance Metrics: F-measure, Accuracy, and Specificity and Sensitivity

This section provides the details about the computation of the performance metrics used in this study. The expressions to calculate overall accuracy, accuracy by class, and sensitivity by class are described below:

$$Fc - measure = \frac{2 * TPc}{2 * TPc + FPc + FNc} \times 100 \tag{A1}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100 \tag{A2}$$

$$Ac = \frac{TPc + TNc}{TPc + FNc + FPc + TNc} \times 100 \tag{A3}$$

$$Sc = \frac{TPc}{TPc + FNc} \times 100 \tag{A4}$$

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

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

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

$$Sensitivity_C = \frac{7673}{7673 + 2116} \times 100 = 78.4\%$$

$$Specificity_C = \frac{7673}{7673 + 24471} \times 100 = 92.09\%$$

Table A.1: Confusion matrix of the PAC system by Bao et al. in in-lab training/out-lab testing scenario.

classified as →	walk	stand	sit	lie
walk	5573	754	13	0
stand	776	7673	1337	3
sit	103	1124	18197	73
lie	0	1	235	277

APPENDIX B

Detailed Description of the Training and Classification Process Used

This section provides the details about the classifiers used and the training process adapted. The details about the classification procedure and cross-validation procedure are described in Table B.1.

The cross-validation process is leave-one-subject-out for the in-lab windowing analysis (trained and tested on in-lab data) and for the out-of-lab analysis (trained and tested on out-of-lab data). The training and testing procedure was different in the in-lab-training/out-lab-testing analysis. In this case, the model was trained using the in-lab data of all subjects, but one, which is being tested on the out-of-lab data.

Table B.1: Classification procedure used for each PAC system

Authors	Classifier Used	Cross-Validation Procedure
Cleland et al.	SVM Classifier (with universal Pearson VII function based kernel and complexity value of 100 using WEKA libraries)	Leave-one-subject-out-cross-validation
Bao et al.	Decision Tree Classifier (J48 with default parameters using WEKA libraries)	Leave-one-subject-out-cross-validation
Leutheuseur et al.	Hierarchical Classification (KNN and SVM using WEKA libraries)	Leave-one-subject-out-cross-validation

APPENDIX C

Physical Activity Classification Using Body-Worn Inertial Sensors in a Multi-Sensor Set-Up

This study presents a novel approach to classify ADLs using a multi-sensor configuration: multiple inertial sensors, each mounted to a different body location, are used to capture a variety of movements from both the upper- and lower-body segments. Section II describes the methods and dataset used to develop the PAC algorithm. Section III reports the PAC algorithm classification performance and interprets the classification results. Section IV discusses the findings of the current study and presents a comparison with existing work. Section IV also discusses the limitations of this study and proposes ways to overcome said limitations to make the PAC algorithm more effective in real-life conditions.

C.1 Materials and Methods

The study uses a benchmark DaLiAc dataset [14] acquired from the University of Erlangen in Germany. Nineteen healthy young subjects participated in the data collection protocol by performing a series of prescribed ADLs on the university campus. The activities were: sitting (SI), lying (LY), standing (SD), washing dishes (WD), vacuuming (VC), sweeping (SW), walking (WK), ascending stairs (AS), descending stairs (DS), treadmill running (TR), bicycling (50W) (B50), bicycling (100W) (B100), and rope jumping (RJ). Wearable inertial sensors (Shimmer Research, Dublin, Ireland) [95] were positioned and mounted on the left ankle, right wrist, chest and right hip (Figure C.1), for a total of four inertial sensors (sampling at 204.80 Hz) used to collect the raw 3D accelerometer and 3D gyroscope data. Additional, more-detailed information about the DaLiAc dataset is available online [14].

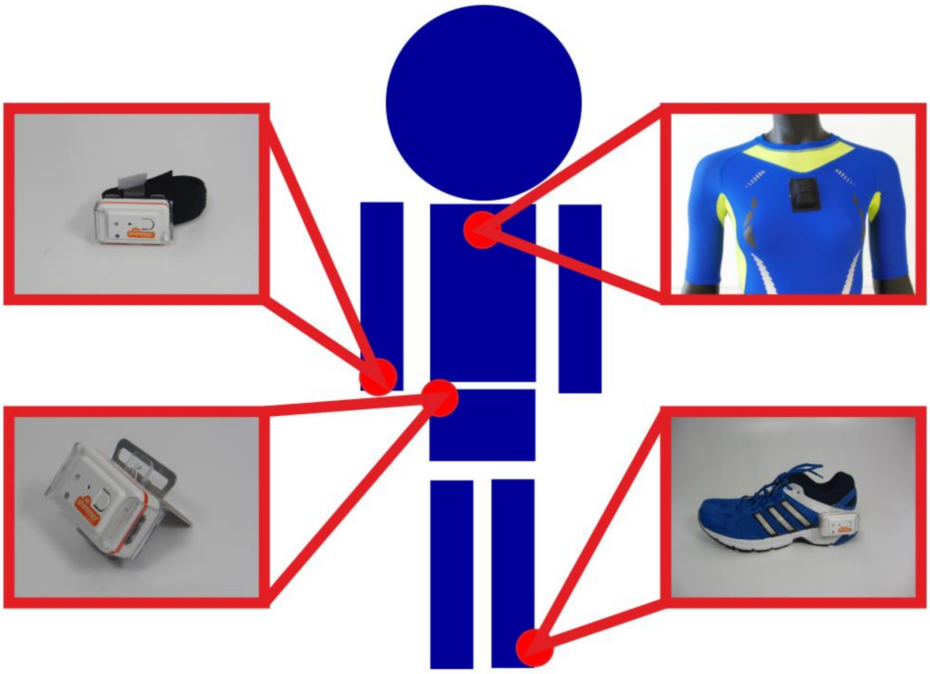


Figure C.1: Sensor placement, courtesy of Leutheuser et al [8].

C.2 Feature Extraction

Feature extraction is necessary before PA classification because pattern recognition algorithms cannot analyze raw signals. The below features were extracted from the raw 3D accelerometer and 3D gyroscope signals and were computed across a window length of 5 sec with 50% overlap.

- Mean
- Standard Deviation
- Variance

- Median
- Range
- RMS [49]
- Skewness [50]
- Kurtosis [50]
- Energy

$$E = \frac{1}{N}\sum_{i=1}^{N}|X(i)|^2$$

The energy of the signal was computed by the sum of the squared FFT components X(i), divided by the total length of the window N for normalization [10].

- Signal Magnitude Area (SMA)

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

SMA was derived by summing up the absolute values of all the axes across window length of N samples [21].

The feature set obtained after the processing of raw data contained a total of 224 features, from which each feature was computed for every single axis of accelerometer and gyroscope of the four sensing units except the feature SMA, which combined the three axes of accelerometer and gyroscope separately in order to get a single value for each, against a single inertial sensor.

C.3 Physical Activity Classification

The KNN clasifier was used as a pattern recognition algorithm. It was evaluated with 10-fold cross-validation and K was set to 1 as a default setting to classify the 13 ADLs detailed above in Section II, A. Weka data mining software (University of Waikato, Version 3.6.12 [59]) was used to build the classifier on the computed feature set. The metrics computed after performing the classification algorithm on the feature set are shown below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

Where, TP-True Positive, TN-True Negative, FP-False Positive, and FN-False Negative derived from the confusion matrix in Table I.

C.4 Results

The classification algorithm was implemented on the entire features set of 224 features and an overall classification rate was computed. The classifier output was compared with the original ADL labels across a window of 5 sec and is presented in the form of confusion matrix in Table C.1. From the confusion matrix, accuracy, specificity, sensitivity, and precision were computed for each ADL and are reported in Table C.2. The classification rate of the proposed PAC algorithm was compared with the original PAC algorithm by Leutheuser et al. [14], presented in Table C.3. The performance metrics computed from the confusion matrix showed remarkably high detection rates of above 90% to differentiate each activity. The comparison of the proposed algorithm with the original work by Leutheuser et al. [14] also shows quite promising results in classifying the ADLs. The overall mean classification rate of the proposed PAC algorithm (97.38%) exceeded that of the original Leutheuser PAC algorithm (89.6%).

Table C.1: Confusion Matrix of the PAC Algorithm Where Every Value Corresponds to Window of 5 sec Activity

SI	**LY**	**ST**	**WD**	**VC**	**SW**	**WK**	**AS**	**DS**	**TR**	**B50**	**B100**	**RJ**	**← Classified as**
446	0	1	1	1	0	0	0	1	0	0	0	0	**SI**
1	455	0	0	0	0	0	0	0	0	0	0	0	**LY**
1	0	441	8	1	0	0	1	0	0	1	0	0	**ST**
0	0	0	927	7	1	1	0	0	0	0	1	0	**WD**
0	0	0	3	443	6	1	0	0	0	0	1	0	**VC**
0	0	1	3	37	683	7	6	2	0	1	3	0	**SW**
0	0	0	1	1	2	2009	10	10	5	2	1	0	**WK**
0	0	0	0	0	0	12	305	3	0	0	0	0	**AS**
0	0	0	0	0	0	6	3	264	0	0	0	0	**DS**
0	0	0	0	0	0	3	0	0	905	3	0	0	**TR**
0	0	1	1	0	0	3	0	1	4	876	37	0	**B50**
0	0	0	1	0	0	0	0	0	0	37	882	2	**B100**
0	0	0	0	0	0	0	0	0	0	0	0	243	**RJ**

Table C.2: Classification Rates Comparison of the Proposed System with the Previous Work by Leutheuser [8]

No.	Activity	Leutheuser et al.[8]	Proposed
1	SI	88.90	**99.33**
2	LY	100.00	99.78
3	ST	89.80	**97.35**
4	WK	99.00	98.53
5	AS	95.50	**95.63**
6	DS	95.20	**96.70**
7	WD	98.1	**98.72**
8	VC	85.4	**97.8**
9	SW	89.9	**91.79**
10	TR	100.00	99.45
11	B50	69.10	**95.12**
12	B100	53.50	**95.99**
13	RJ	100.00	100.00
Mean		89.6	97.38

The bold values show that the proposed PAC algorithm performed better than the original.

Table C.3: Performance Metrics Derived from the Confusion Matrix

Activity	Accuracy	Specificity	Sensitivity	Precision
SI	99.96	99.99	99.33	99.78
LY	99.99	100.00	99.78	100.00
ST	99.82	99.95	97.35	99.10
WK	99.34	99.58	98.53	98.53
AS	99.63	99.77	95.63	93.87
DS	99.75	99.84	96.70	94.96
WD	99.65	99.76	98.72	97.88
VC	99.36	99.45	97.8	90.24

SW	99.23	99.89	91.79	98.7
TR	99.86	99.90	99.45	99.12
B50	99.08	99.52	95.12	95.75
B100	99.11	99.46	95.99	95.26
RJ	99.98	99.98	100.00	99.18

C.5 Discussion and Conclusion

The mean classification rate of the proposed PAC algorithm is high and exceeds that of the original PAC algorithm described by Leutheuser et al. [14], emphasizing the significant impact of this work. The proposed PAC algorithm outperformed the existing PAC algorithm when classifying nine out of the 13 total ADLs; the two algorithms performed the same for one ADL, and the original algorithm slightly outperformed the proposed algorithm for the remaining 3 ADLs (Table II). The proposed PAC algorithm outperformed the original PAC algorithm when classifying two out of the three sedentary ADLs (sitting and standing). One possible reason is that the energy feature (along with other features) was computed for every single axis in our case while in the original algorithm described by Leutheuser et al., the energy feature was derived by combining all the axis to get a single value for each sensor. This possibly minimized the information about variation in a single axis (e.g. vertical axis). The proposed algorithm performed slightly better in the ascending stairs and descending stairs activities and slightly worse for the walking activity. This could be because in this study a generalized classification method was used for all activities instead of using hierarchical classification approach used in [14]. In the ADLs: lying, walking, and treadmill running, our proposed PAC algorithm performed slightly worse than the original PAC algorithm by Leutheuser et al., but the difference was less than 0.6%. Furthermore, in the ADLs: ascending stairs and washing dishes, the improvement of proposed PAC algorithm over the original was less than 0.75%. Most importantly, the improvement in the classification rate of the proposed PAC algorithm was significantly higher for the activities: sitting, standing, sweeping, vacuuming bicycling (50W) and bicycling (100W), making the proposed PAC algorithm superior to the original by Leutheuser et al. The reason behind the overall higher classification rates in detection of the ADLs is related to the utilization of features which show significant variations in behavior when compared between static and dynamic activities (i.e. energy and SMA). The SMA values were lower when there were sedentary ADLs such as sitting, standing, and lying, while in dynamic ADLs (i.e., walking, stairs up, stairs down and running) the SMA values were comparatively high. The SMA feature enabled us to discriminate between static and dynamic ADLs, which in turn resulted in higher overall classification rates for the proposed algorithm.

Additionally, the proposed algorithm is preferable to the existing algorithm due to its simple, efficient design. Instead of using a different classification approach for each subset of ADLs, we used a single procedure to classify all 13 ADLs. The algorithm described by Bao and Intille [9] was also implemented in [14] using the same benchmark dataset and achieved an overall mean classification rate of 80%, which is significantly less than the classification rate achieved in this study. Specifically, the proposed algorithm outperformed the algorithm described by Bao and Intille when classifying activities 1-6 in Table III, the six basic ADLs for activity classification algorithm development and testing [9, 10, 13, 14, 21, 22, 29]. This is likely because Bao and Intille used only accelerometers signals for feature extraction and algorithm development. In contrast, both accelerometer and gyroscope signals were used in this study.

In sum, the proposed algorithm has shown encouraging results in classifying both sedentary and mobile ADLs. It is important to note though that most of the analyzed features in this study are based on statistical computations and are quite sensitive to change. Slight variations in sensor placement can significantly influence the feature values, which will, in turn, affect the algorithm’s classification performance rate. There is a need to explore the use of biomechanical features to enhance algorithm classification performance under various sensors orientations. Integrating biomechanical features with the current features used in this study will help compensate for variations in sensor placement and in turn will make the algorithm more robust and versatile. It is also worth mentioning that in most of the previous studies, not many efforts were spent on the selection of feature set prior to classification if any at all. This stage is often neglected. Mindful feature selection is very important and, if performed properly, can easily rid redundancies within the feature set that could affect the detection capabilities of the PAC algorithm.

The benchmark dataset used in this study was based on a structured protocol in a restricted environment where subjects were instructed to perform prescribed ADLs. Both the duration and the way of performing these ADLs were predefined. These conditions are quite different to those of real life where ADLs are not structured (e.g., walking in a laboratory environment can be totally different to walking in the home environment). A possible solution to this issue would be to measure ADLs in daily life conditions using the same data acquisition methods detailed in this study to capture more realistic behaviors. In addition, testing the existing algorithms for PAC on this new dataset would give more insight to the challenges and gaps that must be addressed in the new PAC algorithms in order to make them more robust and practical in daily life settings.

The benchmark dataset used in this study was based on a structured protocol in a restricted environment where subjects were instructed to perform prescribed ADLs. Both the duration and the way of performing these ADLs were predefined. These conditions are quite different to those of real life where ADLs are not structured (e.g., walking in a laboratory environment can be totally different to walking in the home environment). A possible solution to this issue would be to measure ADLs in daily life conditions using the same data acquisition methods detailed in this study to capture more realistic behaviors. In addition, testing the existing algorithms for PAC on this new dataset would give more insight to the challenges and gaps that must be addressed in the new PAC algorithms to make them more robust and practical in daily life settings.

Proper selection of sensor type, count, and placement are all important considerations when developing and validating PAC algorithms for real-life applications. This study used four sensors, which is a relatively large number when considering a PAC system in real life environment (as sensor count increases, issues such as sensor wearability, battery life and data storage limitations arise). Therefore, sensor selection and placement are a critical task that must be solved efficiently so that the classification performance of the PAC algorithm is minimally affected by decreasing the number of sensing units. Smartphones and smartwatches may serve as suitable alternatives in terms of wearability and user-friendliness as compared to body-worn inertial sensors alone since most smartphones and smartwatches on the market today come embedded with the inertial sensors. These devices also have their limitations such as low data storage, shorter battery life and limited computational capabilities, which is another topic that must be addressed by the research community.

Bibliography

[1] M. Awais, S. Mellone, and L. Chiari, "Physical activity classification meets daily life: Review on existing methodologies and open challenges," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 5050-5053.

[2] *The 2012 Ageing Report*. Available: http://ec.europa.eu/economy_finance/publications/european_economy/2012/pdf/ee-2012-2_en.pdf

[3] W. H. O. R. O. f. Europe. (2012). *Policies and priority interventions for healthy ageing*. Available: http://www.euro.who.int/__data/assets/pdf_file/0006/161637/WHD-Policies-and-Priority-Interventions-for-Healthy-Ageing.pdf?ua=1

[4] W. J. Chodzko-Zajko, D. N. Proctor, M. A. Fiatarone Singh, C. T. Minson, C. R. Nigg, G. J. Salem, *et al.*, "American College of Sports Medicine position stand. Exercise and physical activity for older adults," *Med Sci Sports Exerc,* vol. 41, pp. 1510-30, Jul 2009.

[5] W. H. Organization. (2010, July 19, 2016). *Global Recommendations on Physical Activity for Health*. Available: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf

[6] H. Zheng, N. D. Black, and N. D. Harris, "Position-sensing technologies for movement analysis in stroke rehabilitation," *Medical and biological engineering and computing,* vol. 43, pp. 413-420, 2005.

[7] M. Uiterwaal, E. Glerum, H. Busser, and R. Van Lummel, "Ambulatory monitoring of physical activity in working situations, a validation study," *Journal of medical engineering & technology,* vol. 22, pp. 168-172, 1998.

[8] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, "Physical activity monitoring based on accelerometry: validation and comparison with video observation," *Medical & biological engineering & computing,* vol. 37, pp. 304-308, 1999.

[9] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive computing*, ed: Springer, 2004, pp. 1-17.

[10] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *AAAI*, 2005, pp. 1541-1546.

[11] P. M. Grant, C. G. Ryan, W. W. Tigbe, and M. H. Granat, "The validation of a novel activity monitor in the measurement of posture and motion during everyday activities," *British Journal of Sports Medicine,* vol. 40, pp. 992-997, 2006.

[12] 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," *Biomedical Engineering, IEEE Transactions on,* vol. 56, pp. 871-879, 2009.

[13] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition,* vol. 43, pp. 3605-3620, 2010.

[14] H. Leutheuser, D. Schuldhaus, and 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, p. e75196, 2013.

[15] D. Trabelsi, S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "An unsupervised approach for automatic activity recognition based on hidden Markov model regression," *Automation Science and Engineering, IEEE Transactions on,* vol. 10, pp. 829-835, 2013.

[16] L. Gao, A. Bourke, and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Medical engineering & physics,* vol. 36, pp. 779-785, 2014.

[17] J. J. Guiry, P. van de Ven, J. Nelson, L. Warmerdam, and H. Riper, "Activity recognition with smartphone support," *Medical engineering & physics,* vol. 36, pp. 670-675, 2014.

[18] A. Moncada-Torres, K. Leuenberger, R. Gonzenbach, A. Luft, and R. Gassert, "Activity classification based on inertial and barometric pressure sensors at different anatomical locations," *Physiological measurement,* vol. 35, p. 1245, 2014.

[19] K. R. Westerterp, "Assessment of physical activity: a critical appraisal," *European journal of applied physiology,* vol. 105, pp. 823-828, 2009.

[20] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," in *Pervasive Computing,* ed: Springer, 2006, pp. 1-16.

[21] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *Information Technology in Biomedicine, IEEE Transactions on,* vol. 14, pp. 1166-1172, 2010.

[22] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, *et al.*, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors,* vol. 13, pp. 9183-9200, 2013.

[23] 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," *Medical engineering & physics,* vol. 37, pp. 705-711, 2015.

[24] 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.

[25] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Büla, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly," *Biomedical Engineering, IEEE Transactions on,* vol. 50, pp. 711-723, 2003.

[26] G. Lyons, K. Culhane, D. Hilton, P. Grace, and D. Lyons, "A description of an accelerometer-based mobility monitoring technique," *Medical engineering & physics,* vol. 27, pp. 497-504, 2005.

[27] A. Godfrey, A. Bourke, G. Olaighin, P. Van De Ven, and J. Nelson, "Activity classification using a single chest mounted tri-axial accelerometer," *Medical engineering & physics,* vol. 33, pp. 1127-1135, 2011.

[28] R. Ganea, A. Paraschiv-lonescu, and K. Aminian, "Detection and classification of postural transitions in real-world conditions," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 20, pp. 688-696, 2012.

[29] M. B. Del Rosario, K. Wang, J. Wang, Y. Liu, M. Brodie, K. Delbaere, *et al.*, "A comparison of activity classification in younger and older cohorts using a smartphone," *Physiological measurement,* vol. 35, p. 2269, 2014.

[30] 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 transactions on information technology in biomedicine,* vol. 10, pp. 156-167, 2006.

[31] A. Hickey, S. Del Din, L. Rochester, and A. Godfrey, "Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis," *Physiol Meas,* vol. 38, pp. N1-N15, Jan 2017.

[32] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors (Basel),* vol. 10, pp. 1154-75, 2010.

[33] K. Lee and M.-P. Kwan, "Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results," *Computers, Environment and Urban Systems,* vol. 67, pp. 124-131, 2018.

[34] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *IEEE journal of Biomedical and Health Informatics,* vol. 17, pp. 553-560, 2013.

[35] J. J. Dominguez Veiga, M. O'Reilly, D. Whelan, B. Caulfield, and T. E. Ward, "Feature-Free Activity Classification of Inertial Sensor Data With Machine Vision Techniques: Method, Development, and Evaluation," *JMIR Mhealth Uhealth,* vol. 5, p. e115, Aug 04 2017.

[36] F. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors,* vol. 16, p. 115, 2016.

[37] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster*, et al.*, "Collecting complex activity datasets in highly rich networked sensor environments," in *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*, 2010, pp. 233-240.

[38] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880,* 2016.

[39] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *Wearable Computers (ISWC), 2012 16th International Symposium on*, 2012, pp. 108-109.

[40] M. Altini, J. Penders, R. Vullers, and O. Amft, "Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning," *IEEE journal of biomedical and health informatics,* vol. 19, pp. 219-226, 2015.

[41] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," *Pervasive Computing,* pp. 1-16, 2006.

[42] A. Chowdhury, D. Tjondronegoro, V. Chandran, and S. Trost, "Physical Activity Recognition using Posterior-adapted Class-based Fusion of Multi-Accelerometers data," *IEEE J Biomed Health Inform,* May 17 2017.

[43] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," *IEEE Transactions on Biomedical Engineering,* vol. 61, pp. 1780-1786, 2014.

[44] A. M. Khan, Y. K. Lee, S. Lee, and T. S. Kim, "Accelerometer's position independent physical activity recognition system for long-term activity monitoring in the elderly," *Med Biol Eng Comput,* vol. 48, pp. 1271-9, Dec 2010.

[45] M. Kamada, E. J. Shiroma, T. B. Harris, and I. M. Lee, "Comparison of physical activity assessed using hip- and wrist-worn accelerometers," *Gait Posture,* vol. 44, pp. 23-8, Feb 2016.

[46] F. S. Ayachi, H. P. Nguyen, C. Lavigne-Pelletier, E. Goubault, P. Boissy, and C. Duval, "Wavelet-based algorithm for auto-detection of daily living activities of older adults captured by multiple inertial measurement units (IMUs)," *Physiol Meas,* vol. 37, pp. 442-61, Mar 2016.

[47] 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 (Basel),* vol. 16, Dec 11 2016.

[48] E. K. Antonsson and R. W. Mann, "The frequency content of gait," *Journal of biomechanics,* vol. 18, pp. 39-47, 1985.

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

[50] S. Liu, R. Gao, and P. Freedson, "Computational methods for estimating energy expenditure in human physical activities," *Medicine and science in sports and exercise,* vol. 44, p. 2138, 2012.

[51] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell, "Activity recognition using a single accelerometer placed at the wrist or ankle," *Medicine and science in sports and exercise,* vol. 45, p. 2193, 2013.

[52] K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall, "A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers," *Physiological measurement,* vol. 35, p. 2191, 2014.

[53] X. Long, B. Yin, and R. M. Aarts, "Single-accelerometer-based daily physical activity classification," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE,* 2009, pp. 6107-6110.

[54] M. Arif, M. Bilal, A. Kattan, and S. I. Ahamed, "Better physical activity classification using smartphone acceleration sensor," *Journal of medical systems,* vol. 38, p. 95, 2014.

[55] Y. Zheng, "Human activity recognition based on the hierarchical feature selection and classification framework," *Journal of Electrical and Computer Engineering,* vol. 2015, p. 34, 2015.

[56] U. Lindemann, W. Zijlstra, K. Aminian, S. F. Chastin, E. D. De Bruin, J. L. Helbostad*, et al.*, "Recommendations for standardizing validation procedures assessing physical activity of older persons by monitoring body postures and movements," *Sensors,* vol. 14, pp. 1267-1277, 2014.

[57] M. Arif, M. Bilal, A. Kattan, and S. I. Ahamed, "Better physical activity classification using smartphone acceleration sensor," *Journal of medical systems,* vol. 38, pp. 1-10, 2014.

[58] A. K. Bourke, E. A. 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 (Basel),* vol. 17, Mar 10 2017.

[59] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, pp. 10-18, 2009.

[60] M. Awais, Palmerini, L., Chiari, L., (In Press), "Physical Activity Classification Using Body-Worn Inertial Sensors in a Multi-Sensor Setup," presented at the 2nd International Forum on Research and Technologies for Society and Industry (RTSI), Bologna, Italy, 2016.

[61] 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, p. 559, 2017.

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

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

[64] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *ICML*, 2003, pp. 856-863.

[65] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, 1992, pp. 249-256.

[66] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST),* vol. 2, p. 27, 2011.

[67] Y.-M. Huang and S.-X. Du, "Weighted support vector machine for classification with uneven training class sizes," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, 2005, pp. 4365-4369.

[68] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, "Advancing feature selection research," *ASU feature selection repository,* pp. 1-28, 2010.

[69] 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, p. 2105, 2016.

[70] 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, p. e0124414, 2015.

[71] T. Huynh, M. Fritz, and B. Schiele, "Discovery of activity patterns using topic models," in *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, pp. 10-19.

[72] M. Stikic, D. Larlus, S. Ebert, and B. Schiele, "Weakly supervised recognition of daily life activities with wearable sensors," *IEEE transactions on pattern analysis and machine intelligence,* vol. 33, pp. 2521-2537, 2011.

[73] S. Bhattacharya, P. Nurmi, N. Hammerla, and T. Plötz, "Using unlabeled data in a sparse-coding framework for human activity recognition," *Pervasive and Mobile Computing,* vol. 15, pp. 242-262, 2014.

[74] J. E. Sasaki, D. John, and P. S. Freedson, "Validation and comparison of ActiGraph activity monitors," *J Sci Med Sport,* vol. 14, pp. 411-6, Sep 2011.

[75] P. Husu, H. Vähä-Ypyä, and T. Vasankari, "Objectively measured sedentary behavior and physical activity of Finnish 7-to 14-year-old children–associations with perceived health status: a cross-sectional study," *BMC public health,* vol. 16, p. 338, 2016.

[76] K. Taraldsen, M. H. Granat, and J. L. Helbostad, "Quantification of outdoor mobility by use of accelerometer-measured physical behaviour," *BioMed research international,* vol. 2015, 2015.

[77] L. Verlaan, S. Bolink, S. Van Laarhoven, M. Lipperts, I. Heyligers, B. Grimm, *et al.*, "Accelerometer-based physical activity monitoring in patients with knee osteoarthritis: objective and ambulatory assessment of actual physical activity during daily life circumstances," *The open biomedical engineering journal,* vol. 9, p. 157, 2015.

[78] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association,* vol. 46, pp. 68-78, 1951.

[79] J. D. Gibbons and S. Chakraborti, "Nonparametric statistical inference," in *International encyclopedia of statistical science*, ed: Springer, 2011, pp. 977-979.

[80] D. Best and D. Roberts, "Algorithm AS 89: the upper tail probabilities of Spearman's rho," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 24, pp. 377-379, 1975.

[81] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, *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.

[82] D. J. Rose, N. Lucchese, and L. D. Wiersma, "Development of a multidimensional balance scale for use with functionally independent older adults," *Archives of physical medicine and rehabilitation,* vol. 87, pp. 1478-1485, 2006.

[83] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.

[84] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering,* vol. 21, pp. 1263-1284, 2009.

[85] M. S. Tremblay, S. Aubert, J. D. Barnes, T. J. Saunders, V. Carson, A. E. Latimer-Cheung, *et al.*, "Sedentary behavior research network (SBRN)–terminology consensus project process and outcome," *International Journal of Behavioral Nutrition and Physical Activity,* vol. 14, p. 75, 2017.

[86] M. Awais, "Physical activity classification meeting daily life conditions," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 393-398.

[87] Y. Ming, Q. Ruan, and A. G. Hauptmann, "Activity recognition from RGB-D camera with 3D local spatio-temporal features," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, 2012, pp. 344-349.

[88] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2847-2854.

[89] M. S. Ryoo and L. Matthies, "First-person activity recognition: What are they doing to me?," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2730-2737.

[90] L. Baraldi, F. Paci, G. Serra, L. Benini, and R. Cucchiara, "Gesture recognition in ego-centric videos using dense trajectories and hand segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 688-693.

[91] M. Ryoo, T. J. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies, "Robot-centric activity prediction from first-person videos: What will they do to me?," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 295-302.

[92] S. Singh, C. Arora, and C. Jawahar, "First person action recognition using deep learned descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2620-2628.

[93] A. Kim, J. Kim, S. Rietdyk, and B. Ziaie, "A wearable smartphone-enabled camera-based system for gait assessment," *Gait Posture,* vol. 42, pp. 138-44, Jul 2015.

[94] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo*, et al.*, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics,* vol. 21, pp. 4-21, 2017.

[95] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer*, et al.*, "SHIMMER™–A wireless sensor platform for noninvasive biomedical research," *Sensors Journal, IEEE,* vol. 10, pp. 1527-1534, 2010.