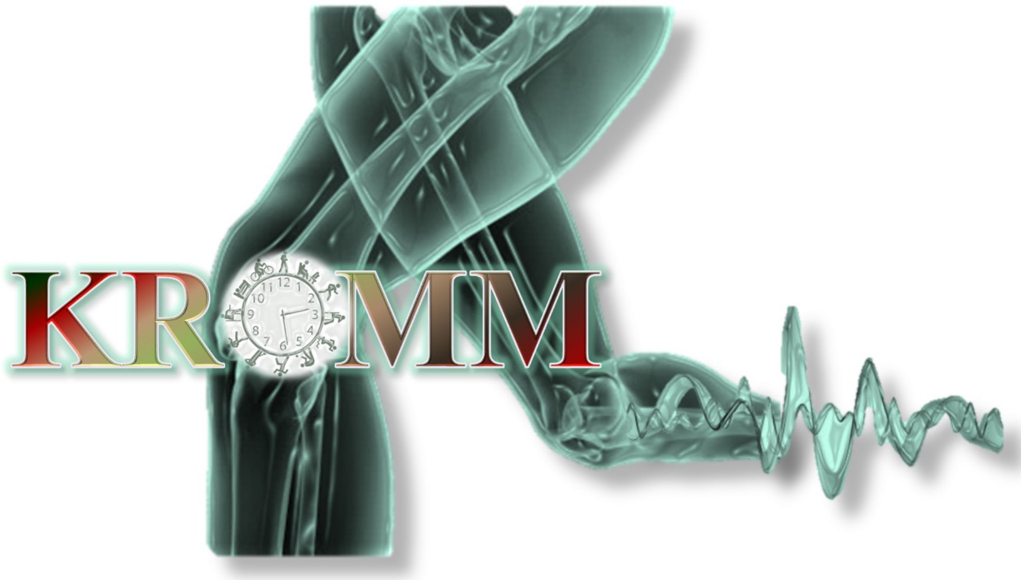


Knee Range of Motion Monitoring System for personalized rehabilitation in Osteoarthritis Patients

A machine learning exploratory study for more defined and flexible boundaries around walking and stair ambulation clusters in free-living conditions

Giulia Bottoni

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisor(s): Prof. Danilo Demarchi
Prof. Paolo Bonato
Dr. Stefano Sapienza

10 december 2018

Dedicated to biomedical engineering, to the passion, the curiosity, the creativity and the stubbornness that it requires

“Our vision is that, in the not too distant future, each patient will be surrounded by a ‘virtual cloud’ of billions of data points that will uniquely define their past medical history and current health status. Furthermore, it will be possible to mine the billions of data points from hundreds of millions of individuals to generate algorithms to help predict the future clinical needs for each patient.” (Muin J. Khoury, 2013)

Acknowledgments

Thanks to Danilo Demarchi and to Paolo Bonato for giving me the possibility of living my dream and at the same time making me understand how the real word of the research is.
To my supervisor Stefano Sapienza for always challenging me with wisdom and patience.
To all the team of the Motion Analysis Lab for supporting me personally and professionally.
To my professor Monica Visintin for showing me how a real teacher has to be, tough but fair, passionate about her job, confident in her students.
To my professor Guido Pagana that taught me all the creativity that engineering has with his brilliant mind and his enthusiasm.
To my professor Franca D'Agostini for making me do my first lesson in front of a whole class and giving me my first 30 cum laude at the Politecnico of Turin.
To my teacher Carla Mariucci for being the first one making me love math.
To my professor Ombretta Guerri for being the first one believing that I would have become an engineer, the first one that makes me love science.
To my teacher Lucia Tingoli for being the first one believing that I would have become a writer.
To my professor Paolo Amico for making me fall in love with physics.
To my professor Elena Pelliccia for pushing me over my limits and tears.
To my professor Francesco Gagliardi for opening my mind and my criticism during the hours of philosophy.
To my professor Maria Letizia Giontella for making me explore every corner of my emotions and my fantasies and encouraging me to write and take a chance in national competitions.
To my dance teacher Tiziana Basciani for teaching me discipline and commitment and for giving the unique smell of wood and the velvet of the theatre, the sensation of being free and alive while the music is playing.
To my skating teacher Ketty Salatin for teaching me that with hard work, shoulders back and head up it is possible to win every personal and professional challenge.

Thanks to the small town of Ponte Pattoli because I was born next to my best friend Roberta. Thanks to her because of her loyalty and support in all the most important moment of my life, but especially in the little "normal" everyday life, drinking a cappuccino, doing infinite talks walking in the countryside. Thanks to Letizia for the horror movies, the incredible laughs, the real, honest hugs.
Thanks to Perugia where I met Livia. Thanks to her because of our convoluted conversations, our freedom in sharing every thought and fear, our pure love.
Thanks to the Politecnico of Turin because it literally overturns my life.
The first day it makes me meet the blue intense eyes of Alessia Cafagna. Thanks to her for the song of Giorgia in the car, for the tears the day we passed Analysis 1, for all the beautiful mess we've been trought.
The second year it makes me meet Alessia Botta. Thanks to her that run to me every time I need her, that went over the mask, straight to the heart, that respects me and loves me like a sister, always believing in my goals and blaming me if I want to give up.

The third year it put me on the way of Giulia Luciano and everything changed again, completely. Thanks to her for the sushi during the session of the exam, for the concert of Laura Pausini, for the infinite talks in the bed about the crazy things happened during the day, thanks for the wild adventures and the unique relationship.

Thanks to Giulia I met better Carla and Barbara. Thanks to Carla for the comfortable advices, the delicious meal.

Thanks to Barbara for the walks at the same speed, for the magic days and nights, feeling good and alive.

The fourth year I met Veronica, thanks to her for bringing me always on the wrigth path, for knowing how to love.

In all these years my family of Turin stays with me.

First of all, my sister Ginevra, thanks to her for the unconditional love, for six years full of wonderful surprises and first times. Thanks to her for her soul and her generosity, for always being there for me, no matter the time and the circumstances, for taking care of my dreams, for holding my hand and my heart trought all these years without leaving it even for a second.

Thanks to my foster mum Alessia, for sharing with me all her thoughts from the first moment, for dressing, cooking and fighting for me.

Thanks to my foster dad Andrea for showing me how an hard worder and amazing husband has to be taking care of the house and loving his wife and children.

Thanks to my brother Lorenzo for his smart ideas, to my brother Edoardo for his vibrancy.

Thanks to my little sister Beatrice for her empathy and sweetness.

Thanks to Tommaso, for the Chamomile reflecting about our lives, and the hard words for making me growing up.

Thank to Simone for treating me like a princess and making me laugh beyond my tears.

Thanks to my foster grandmother Mirella that gave me important advices about my life because she really cares about me.

Thanks to my foster grandfather Franco that always says to Mirella to not give me advices because it trusts me.

To Argenide, for her determination that makes all these things happen.

Thanks to Boston because I made another family there too.

Thanks to my foster mum Ellie, the most beautiful and the most eclectic mum ever, for making me feel important and loved with small and big gestures.

Thanks to my foster dad Federico that always knows what to say to me and what I am thinking about, for giving me his time even when he hadn't.

Thank to Monserrat for being so much more than a roomate, for convincing me every day that I had something special to be successful and a great heart to deserve to be loved.

Thanks to Zoe for sharing with me all she had, for the honesty, the simpathy, the million laughs.

Thanks to Thomas that believed that I would have become a great biomedical engineer, that trusts me, that took care of me like a daughter and organized me the best birthday party ever. Thanks to Francesco, for his muffins and words to make the day sweeter.

Thanks to Isotta, for supporting and especially bearing me in all the situations.

Thanks to Flavio for being a loyal friend, for cooking and cheering me up when I was resigned. Thanks to Lucrezia for being a real friend and adventure mate, for making me feel safe and happy.

Thanks to my grandmother Annamaria and my granfather Bruno, for all the times I run into their couch to show them my skating's cup, because today I really want to do the same with my laurel wreath, to hear your voice say ”Brava”, granpa, to feel your caress, granma.
Thanks to my grandmother Diana that taught me how a strong and unstoppable woman has to be, running to Turin when I needed her help without making excuses, giving me unconditional love.
Thanks to my granfather Paolo for crying after I came back from Boston, for asking me every day how I feel, for always believing I am a smart woman.
Thanks to my aunt Elena and my uncle Nicola that are like grandparents to me because of they way they love me.
Thanks to Rossana and Francesco for making me feel like a daughter, for always cheering for me.
To my aunt Dina, that helps me clarify my heart and my mind a thousand times.
To my uncle Giacomo for always find a way to make me laugh and making me feel special, talented, loved.
To my aunt Rosita that taugh me how to look always at the bright side, to find a solution, to give the best hugs ever.
To Rebecca for her brightness and the way she loves me, to Edoardo for his infinite heart, to Ginevra for the brilliant answer she always gives me.
To my aunt Aurelia for caring about my future and my emotions since I was a baby, to my uncle Fabio for the fastes race cars for bringing me home.
To Gregorio for his sense of humor and to Laura for alway find a way to make me defuse the situation.
Thanks to my mum, the woman of the million ideas, that taught me how to write and how to listen to my heart, how to be passionate about the world around me, to never settle, thanks to the woman that is more like me in the whole world, the woman that knows me better, the one that dedicated her life to be a great mother.
Thanks to my dad, the man of the million resources, that sang me lullabies all the nights when I was a baby and find all the possible creative ways to make me stop crying now that I am a woman, for all his advices that now are my life's quotes, for being always my superhero, for teaching me that, no matter what, you have to pursuit your dreams, run until the end of the road.
Thanks to my sister Paola that taught me more things that I taught to her, for always knowing what to say to me, how to hug me, for being a more beautiful and smarter mini version of me.
Thanks to my sister Margherita, for always having my back, for her amazing voice that is my favorite soundtrack, for her deep eyes that knows all my story, for her heart that loves me over my faults.
Thanks to my weakness and my doubts
To my dreams and my reality
To my mistakes and my goals
Thanks to all the times that I had the patience to listen to someone advice and to all the times that I had the courage to follow my intuition.
Thanks to the ballet, the skating, the writing because they have been comfortable places for my body and my mind in all these years.
Thanks to the biomedical engineering that instead set me on fire, excited my brains, give me a crazy adrenaline.
Thanks to Vincenzo that is both these things at the same time, for had seeing the worst parts of me and still staying, loving me, being family, a friend, a teacher, a team-skating-dance-soul mate, for making me believe, even when I about to surrender, that I am gonna be a creative writer, a great mum, a successful engineer.

Abstract

The KROMM, “Knee Range of Motion Monitoring” is a knee wearable electrogoniometer composed of a knee angular sensor and a three-axis accelerometer placed on the thigh.
The clinical relevance of this device is to guarantee to the clinicians the possibility of an accurate assessment of the knee range of motion during walking and stair ambulation and to the patients with knee osteoarthritis a personalized rehabilitation also in the home setting.
The aim of this stage of the study is developing the next version of the KROMM system achieving three main goals:

- improving the activity detection capability of the algorithm incorporated into the sensors;
- improving the ergonomics of the design for avoid sudden breaks and uncomfortable usability;s
- improving the user-friendliness of the app on the smartphone for assist the patients in the home environment providing information about the time spent in each activity.

In particular, the focus of this thesis is the implementation of the activities detection mechanism of this platform that will construct more flexible and refined decision boundaries around the clusters of walking and stairs ambulation amongst dynamic, heterogeneous, and individualized free living activities.
In the first session of the study 17 healthy subject, aged 18-64 years, working in the Motion Analysis Laboratory at Spaulding Rehabilitation Hospital, were selected to perform the following instructed activities: sitting, standing, walking, ascending stairs, descending stairs, riding an exercise bicycle, using a rowing machine. Walking, rowing and bicycling were performed at different speeds.
These scripted data were segmented, filtered and used for the construction of the model. The features selection step was performed using the relief and the DBI index and the classification with the application of a hierarchic random forest classifier with 100 trees.
The first classifier of the cascade classified between “activities of interest” and “activities of no interest”; then between the “activities of interest” the second one discriminated between “walking “, “descending stairs” and “ascending stairs”.
The hypothesis is that the segmentation and the label of the instances will influence the features selection step and the classification and that the method of cross validation will affect the classification step. For verifying this hypothesis different attempts have been effectuated, changing one parameter at the time, starting from a model with the following parameters:

- Window of 5 s with no overlap
- Different labels for all the activities
- 10 cross fold validation method

This model was then validated on the unscripted data, data that were recorded for nine hours for four of these seventeen subjects, comparing the predictions of the classifier with the labels obtained from the images of a wearable camera that was wore facing the anterior direction of the human body.

During the second part of the study to the subjects was also given a mobile phone with an application for reminding the activities to do.
The aim of this thesis is to provide a flexible classifier for a personalized rehabilitation a systematic approach to detect instances of ground walking and stair ambulation in uncontrolled living conditions and to understand how the borders of the clusters of the activities move depending on the number, the correlation and the labels of these instances.

Keywords: machine learning, activity recognition, walking recognition, stair ambulation recognition, knee OA

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The KROMM, “Knee Range of Motion Monitoring” is a knee wearable device for monitoring the knee range motion during functional activities in patients with knee osteoarthritis.
The Motion Analysis Laboratory at Spaulding Rehabilitation Hospital in Boston, the Department of Physical Medicine and Rehabilitation at Harvard Medical School, the MGH Institute of Health Professions are the institution involved into this award-winning project that will have a strong clinical relevance in the field of personalize physical rehabilitation providing a more quantitative analysis of the patient condition in the home setting.

1.1 Objectives

Long term monitoring of daily activities in the home environment can provide reliable outcomes that will help the clinicians in tailoring the treatment for each patient and the patients in having a better understanding of the improvements of the therapeutic intervention.
Over the last decade, this necessity opened the path to the development of the wearable technologies system based on Inertial Measurement Units (IMUs: accelerometers, gyroscopes and/or magnetometers) for tracking joint kinematics. These systems present the same accuracy of motion capture system such as Vicon (within few degrees) in the estimation of joint angles but they overcome the problem of the confined space and the high cost of the data collection.
However, the existent devices still don’t solve the issue of the very short period of observation (1 -2hours) because of the large errors resulting from gyroscope drift and the positioning of the IMU and because of the battery limitations.
The aim of this stage of the study is developing the next version of the KROMM system achieving three main goals:

- improve the ergonomics of the design for avoid sudden breaks and uncomfortable usability.
- improve the activity detection capability of the algorithm incorporated into the sensors
- improve the user- friendliness of the app on the smartphone for help the patients in the home environment giving information about the time spent in each activity

In particular, the focus of this work is the activities detection mechanism of this platform based on a machine learning method that will construct more flexible and refined decision boundaries around the clusters of walking and stairs ambulation amongst free living activities.

1.2 Thesis Outline

First of all, it will be explained the concept and the aim of "personalized medicine" beyond the different definitions and its successful application in the pharmacological and oncologic field in terms of improvement of the efficacy, reduction of the toxicity and minimization of the cost thanks to predictable and reliable outcomes.
Secondly, it will be shown how this approach could be transposed in the field of physical rehabilitation for overcome the variability of clinical evaluation and therapeutic decision within different patients and for well define an individual growth curve that will guarantee reliable information for both the patients and the clinicians.
Especially it will be shown how to realize this aim choosing useful parameters and seeing their changes all over the time using wearable technologies.
After that, it will descript knee osteoarthritis (0A), the target of this study, in terms of definition, statistics, causes, symptoms, risks, severity, biomechanical consequences and possible treatments.
In particular, it will be shown why physical rehabilitation is normally preferable than surgery and how the changes in the knee flexion and extension patterns during walking and stairs ambulation are an effective clinical measure of knee OA.
Afterward, in the state of the art section, there will be described the wearable technologies for monitoring the knee comparing the aim of the study, the experimental setup, the population, the experimental procedure, the results, the Error, the Pro and the cons, the sensing technology and the machine learning method for the activity detection.
Then, after this literature review, it will be described the hardware platform of the KROMM and its peculiarities in respect to the previous technologies. First it will be explained the Kromm goal from clinicians', patients' and engineers' perspective, secondly it will be displayed the technology of the sensors, their system of orientation, the way to extract raw data, the calibration of the shimmer 3R and finally its challenges.
After that, it will be explained the study procedure with particular attention on the experimental setup, the choice of the activities and the protocol used for the data collection.
Finally, in the section of the data analysis, it will be explained the strategy for realize an accurate detection in highly dynamic, heterogeneous, and individualized conditions using a hierarchic classifier with a random forest algorithm.
First the data analysis on the scripted data (previously filtered) for build the model of the hierarchic classifier. In this part of the work it will be demonstrated the hypothesis that the segmentation and the label affect the feature extraction and the classification outcomes and that the method of crossvalidation affect the performances of the classifier.
Different attempts will be made for evaluate the relevance of this changes and for select the best parameters.
The choices on the scripted data will be made also loking forward at the analysis on the unscripted data that will finally validate the model. In this case, the evaluation will be realized comparing the output of the classifier to the

images from the camera worn during the data collection.
Finally, in the section of the results it will be chosen the best model comparing the performances of different model on both the scripted and unscripted data.
This comparison will involve the confusion Matrix rates, the distribution of false negative and false positive and the critical decision boundaries.
In the last section, it will be described the weakness and the strengths of the model selected and the future challenges for the kromm system.

Chapter 2

Clinical relevance

2.1 The path of personalized medicine: its aim and its methods behind the different definitions

In the 5th century the Greek physician Hippocrates was the pioneer in the discovery of familiar clusters between human disease and in the 20th century Archibald Garrod was the founder of human genetics: these were the two main steps that lead to the Human Genome Project in 1990, to the HapMap project and to the 1000 Genomes Project[1]

Thanks to these projects the knowledge about the relations between human disorders and genetic became wider but the biggest revolution that this new way of thinking the individual's unique genetic introduced was the model of the personalize medicine.

The term "Personalized medicine", also called "genomic medicine" or "precision medicine", has been described in a lot of different ways, accordingly to the choice of highlight every time a different aspect (understanding, diagnosing, preventing or treating the disease).

In particular, the three main definitions of "personalized medicine "are:

- "a medical model that proposes the customization of healthcare, with decisions and practices being tailored to the individual patient by use of genetic or other information" [3];

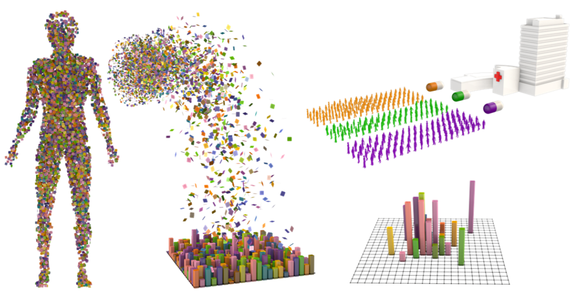


Figure 2.1: The path of personalization [2]: the picture shows how the acquisition of billion of data from a single the patient will help in prescribing the correct drug

Figure 2.2: The decomposition of the information hidden in the DNA carried out by the human genome project [7]

- "the tailoring of medical treatment to the specific characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient. Rather, it involves the ability to classify individuals into subpopulations that are uniquely or disproportionately susceptible to a particular disease or responsive to a specific treatment." [4];
- "a form of medicine that uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease." (National Cancer Institute, 2012)[5]

These three definitions describe from different points of view a new approach to the healthcare with the aim of assessing the risk of a particular disease, predicting its development and prognosis and tailoring it according to the patient personal information.
The background behind this conclusion is the assumption that every patient is unique and so its peculiar characteristics require different treatments.
This recognition of the differences among patients start in the 20th century looking at the different responses of different patients to the same drug therapy but this model has been widely diffused after the Human Genome Project.
The main factors playing an important role in this diversification are age, health condition, life style, nutrition and environment and they create an intricate pattern of responses to each theraphy.
Control this complex map of million outcomes is the aim of the big data analysis and this start a real revolution in the medicine world. In fact, the possibility of the individuation of a specific pattern of response for every patient will improve efficacy, reduced the toxicity and minimized the cost.[6]

2.2 Application of personalized medicine in the physical rehabilitation field

Nowadays, also the field of the physical rehabilitation is going in the direction of the personalization illustrated in the previous section.In fact, looking at the different abilities, improvements and needs of every patient, physiotherapists started to apply personalized rehabilitation programs, making changes to the focus of therapeutic intervention and to the training protocol.[8]
The main goal is to improve the efficacy of treatments tailoring the therapy and the tools for assesting the progress. Therefore, the personalization is on two main levels: a physical one due to the different stages of the disease and

a communicative one due to the personality of the patient. In fact, one of the most important issue in the long

Figure 2.3: An interactive game for rehabilitation [9]

difficult path of the rehabilitation is to make the patient proactive in the rehabilitation process and try to make him motivated because this can improve the outcomes, especially in an unsupervised environment.[10]
Differents techniques were adopted for this purpose, like for example fun interactive games or supportive tools to bring at home.These are attempts of recording more data from the patients without putting too much pressure on them and of continuing the rehabiliation also in the home setting without a sharp border between the house and the hospital.

The main issue indeed is that sometimes even if the patient follow the instruction of the clinicians during the

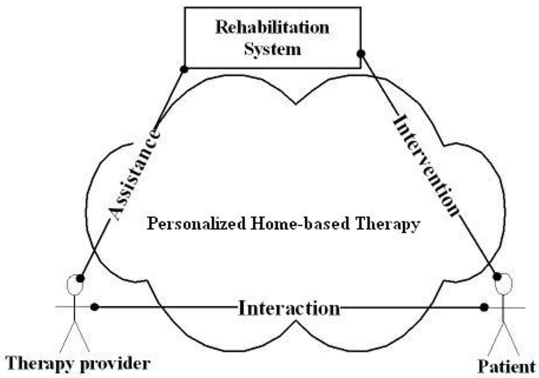


Figure 2.4: Rehabilitation system [11]

session of therapy, then when they get at home they don't have the constancy of doing physical therapy alone or sometimes they just feel pain in the impaired part of the body an so they have the tendencies to not use it.
For example, as it has been shown by Wolf, Taub and others[12], stroke survivors often tend to use the less affected arm in the real life even after the stroke rehabilitation, and so this demonstrated that the clinician can't be sure whether the patient has the same behaviour at home or not.
Therefore one of the aim of the personalized rehabiliation is to attend the patients all the day and everywhere all along the entire period of the rehabiliation, following his changes, his problems, recording his data while the patient is living his life and so obtaining as much information as possible.
All the data collected are like the genes in the Human Genome Project and the hypothesis is that they will define a personalized code that will help predicting the outcomes.

In particular, for realizing a personalized rehabiliation is important to focus the attention on a specific target of patients, then evaluate the reliable and useful outcomes for that particular disease and finally find a system for monitoring the patient continuously.

2.3 The target: knee osteoarthritis

The target chosen for the study is knee osteoarthritis, the most common form of arthritis and accordingly to the World Health Organization (1997) the fourth and eighth most common cause of disability in women and men, respectively.[13]
This degenerative disease is called a "wear-and-tear arthritis" because the cartilage between joints, essential for the shock-absorbing function, become thinner and consequentially there is a sliding friction between the two bones that can change their shape and eventually make them wear away.
The most common symptoms are pain, especially during the movement of the knee at the end of the day, stiffness,

Figure 2.5: Healthy vs not healthy joint: differences visualization [14] and differences description[15]

particularly after rest, a grinding sensation during the movement of the joints, hard swellings due to the osteophytes and soft swellings due to extra fluid in the joint.
Furthermore, there are other symptoms related to mechanical changes. First of all, muscle weakness and loss in the stability of the joint but also less degrees of movement for the knee that become bowed [16] The progress of the osteoarthritis of the knee is generally made by small changes but it is predictable with difficulty because the patients affected are different (variability intra-patients) and each of them is affected in different ways during the progression of the disease (variability intra -patient).
The factors that affect the first type of variability are the age, in fact the risk become higher after 50 years old because of the weakness of the muscle; the gender (in particular woman are generally more affected in terms of number and severity, the weight (a lower one is preferable), hereditary question, injuries or operation on the knee, daily routine in terms of type of job and amount of physical activity. Instead the second type of variability is due mainly to wheter parameters like humidity and pressure, to variation in the daily routine and to the fatigue at the end of the day

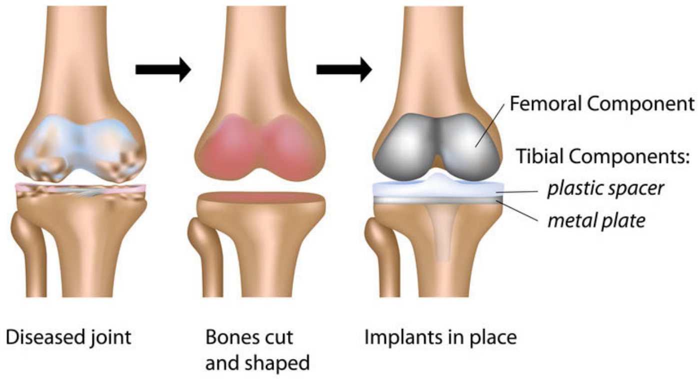


Figure 2.6: Knee replacement surgery [22]

2.3.1 Possible treatments

Knee OA compromise the possibility of doing essential activities of daily living like walking, going up and down stairs so an interventionis essential, preferably in the first stages of this degenerative disease.
The primary goals for restoring the previous quality of life are reducing the pain and regaining mobility and the possible treatments (that perhaps can be used combined) are:

- Weight loss: according to a study of the American college of Rheumatology each pound of weight loss will lead to a 4 fold reduction in the load applied on the knee per step during daily activities, so even a small loss of wieght, will have a great impact on the clinical improvements [17].
- Pain relievers and anti-inflammatory drugs:[18]
- Exercise:strengthening exercises alone have some effects on improving pain and functional outcomes in clients with OA however, in order to maximise the effectiveness of strengthening exercise it is necessary to combine them with a more complete exercise programme including ROM, stretching, functional balance and aerobic exercises[19].
- Braces:the main type of braces are ”unloader” braces, which take the weight away from the side of the knee affected by arthritis and ”support” braces, which provide support for the entire knee [20]
- Surgery:bone realignment surgery, osteotomy and for the most severe cases a total knee replacement or an arthroplasty [21] (2.6)

The literature about the different possible treatments for this disease is wide but all the papers have in common the classification of the treatments depending on the different stage of progression of the disease (from 0 which corresponds to an helathy joint to 4 which corresponds to the most severe stage)
In the stage 1 the patient is starting experiencing the growth of some bone spurs. The treatment can be an integration with glucosamine and chondroitin or a specifical exercise routine.
For the stage 2 even if there is a sufficient synovial liquid and if there is still space between the bones the patients will start feeling pain or other symphtoms, especially after physical activity. In fact, at this stage, the growth of the bones is consistent even if the cartlagine maintain the healthy size. At this stage a non pharmacological therapy

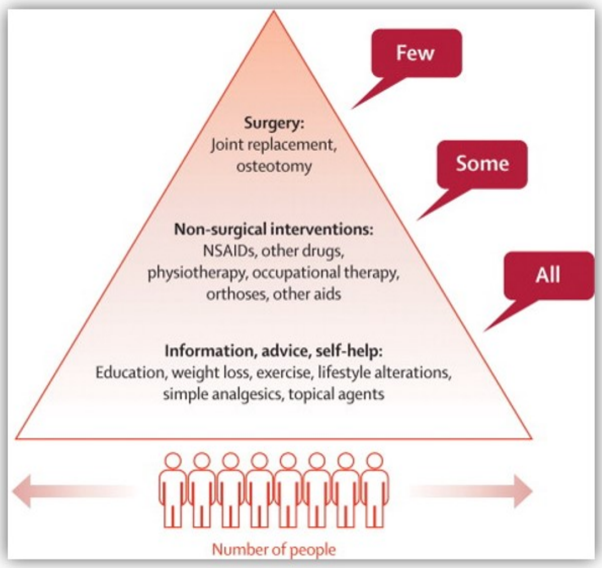


Figure 2.7: Pyramid of the possible treatments of knee OA. From the bottom to the top there is an increased severity of the diasease.

has been suggested because this solution will imply consequences worse than the discomfort due to this mild stage of the disease. It has been proposed instead focus daily exercises for losing weight and strengthening the joints. The stage 3 showed instead a damaged cartilagine frequent pain, joint swelling and stiffness. This is a moderate stage and it can be treated following a pharmaceutical way in combination with physical therapy and weigth loss.

Finally, the last stage is the severe one in which the patients experience great pain, discomfort.

Figure 2.8: Different stages of knee OA [23] from the early stage to the most severe one

Most of the time both the pharmaceutical solution and the physical rehabiliation don't work at this stage of the disease so surgery it is required. This last solution is sometimes the only possiblities for patient in the 4th stage but it isn't the preferable because of its consequences. In fact, even if it guarantees a shift in the weight of the body that will decrease pain or to a complete replacement of the damaged part, it will lead to possible infections, possible further surgeries and to a long recovery period.

2.3.2 Changes of the biomechanics in patient with knee OA

The pain due to the progression of the disease causes strategies of compensation in patients with knee OA. In particular, these modifications to the normal gait are made in the attempt to reduce the knee adduction moment. This moment is due to the fact that during the walking gate there is a disproportion of forces between the medial and the lateral compartment.[24]

In particular, during the gait this causes an adduction that makes the load on the medial compartment increase. This strategy is adopted with the purpos of reducing the pain and don't overload the osteoarthritic knee but in the

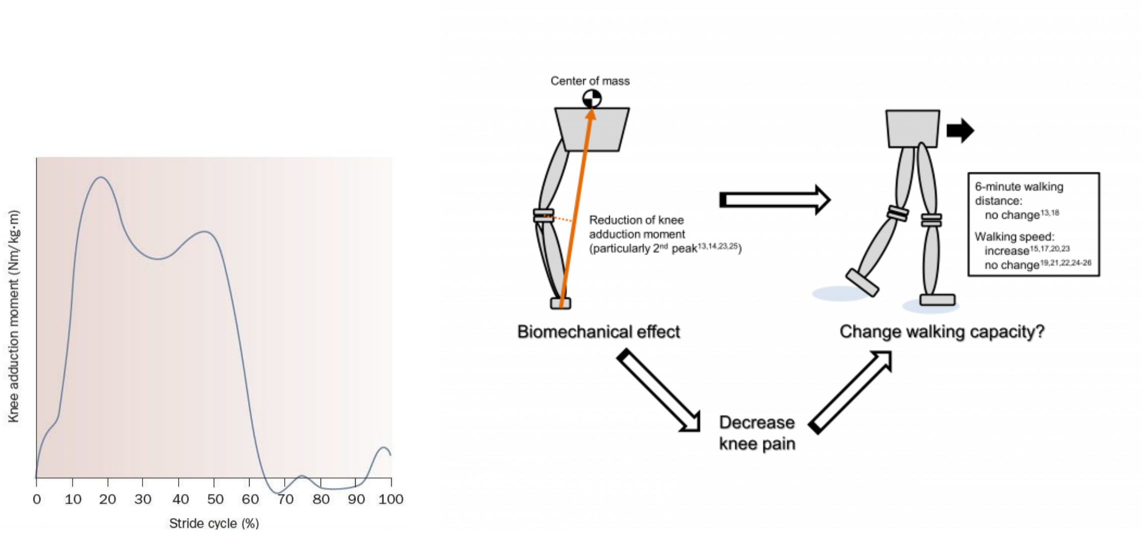


Figure 2.9: Knee adduction moment in function of the stride cycle [24] and compensatory walking strategy[25]

long distance they will lead to an insufficient loading of the joint because of the incorrect gait and the result will be a degeneration of the cartilage like in the case of excessive loading.

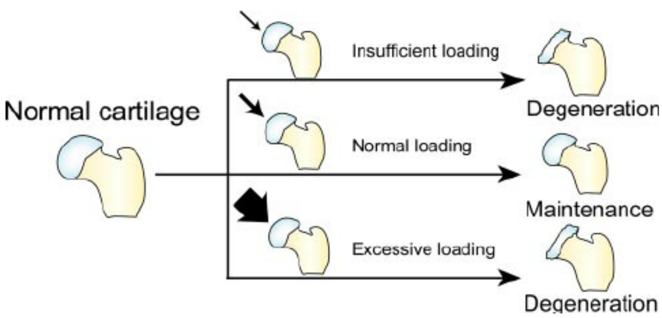


Figure 2.10: Cartilage degeneration [26]

Knee medialization

One effect of the external adduction moment that acts around the knee joint during the entire stance phase of walking is the rotation of medially with respect to the femur in the frontal plane.This phenomenon is caused by a medially acting ground reaction force both on ground and stairs ambulation to compensate the damaged leg.
The main factors that influence the magnitude of the knee adduction moment are:

- the magnitude of the ground reaction force
- the moment arm of the ground reaction force about the knee joint center (defined as the perpendicular distance between the action line of this force and the knee's center of rotation)
- the mass and acceleration of lower limb segments.

Studies in patients with OA have calculated the coefficient of correlation as 0.61 between varus knee alignment and the peak knee adduction moment during walking[27].

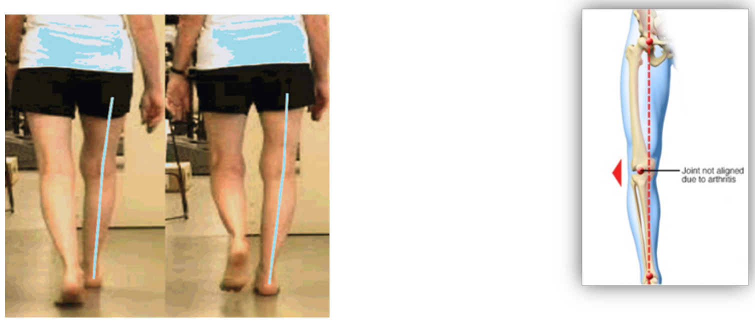


Figure 2.11: Walking gate in which is adopted the knee medialization [28] and misalignment of the joint in the knee medialization [29]

Lateral trunk lean

In this compensatory effect for reducing the knee adduction moment, for unloading the osteoarthritis knee the patient leans the trunk towards the side of the weight-bearing limb and consequently the centre of mass moves laterally, closer to the centre of pressure of the weight-bearing foot and it changes the moment of the ground reaction force that shift towards the knee joint centre[30]. The progression of the disease goes with a progression of the lean that become more visible.
However, even if this compensatory strategy is useful for the reduction of the knee adduction moment, it increases the risk of fall because of the upper body sway. Therefore in this case an incorrect rehabilitation could make the patient loose the possibility of restoration of the healthy gait.

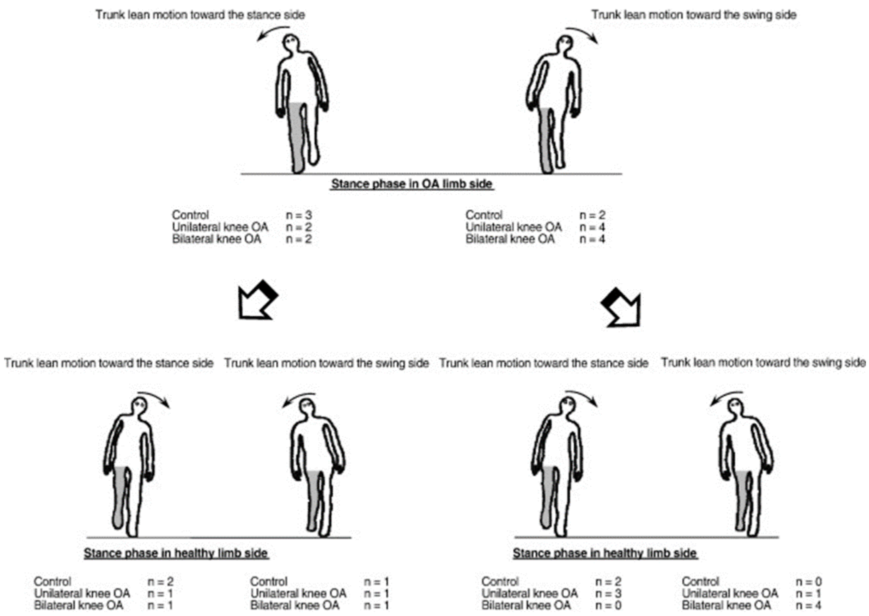


Figure 2.12: Lateral trunk lean [31]

Toe out

Another strategy of compensation adopted for reducing the knee adduction moment is the “toe-out gait” with the “toe-out angle”, a certain degree of rotation of the foot with respect to the direction of the physiological progression.

In this strategy the foot rotates externally and as a consequence there is a shift in the knee joint axis and the conversion of a part of the knee adduction moment in an external knee flexion moment.[32] This modification implies two different consequences on the two peaks of the knee adduction moment:

- a direct correlation on the first peak (that increase) due to a less medial ground reaction force;
- an inverse correlation on the second peak (that decrease) due to the movement of the centre of pressure towards the toes, that make the distance between the ground reaction force and the knee joint centre decrease;

Looking at the follow up of the patient an increasing toe-out angle is a consistent parameter for evaluate the progression of the disease. A lot of different solutions like for example cane in the hand contralateral to the symptomatic knee might shift the body’s centre of mass towards the affected limb, thereby reducing the medially directed ground reaction force, in a similar way as that achieved with the lateral trunk lean strategy described above.

All these strategies of compensation have in common the aim of temporary reducing the pain and thery are attempts that the patient makes on his own.Sometimes the consequences of these behaviours are irrecoverable and so the patient lose the possibility of restoring the healthy gate for good.

The crucial role of the interfere with this process regaining the healthy gait with specific programs like for example the training of the hip abductor that stabilize the frontal plane of motion of the pelvis and the trunk compensating the external knee adduction moment[35]

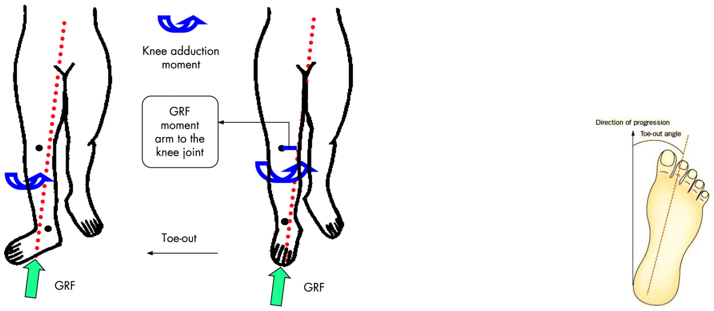


Figure 2.13: Toe-out gait [33] and toe-out angle [34]

2.3.3 Monitoring with wearable sensors

Even if it's already proven that exercises and weight loss have an impact on the pain and on the progression of this illness, the main problem is still the lack of a significant amount of data to prescript a detailed therapy that takes into account the stage of degeneration, the severity of the disease, the most efficient activities to do and the activities to refrain from doing.

Consequently, often this exercise programs lead to an aggravation of joint symptoms for an incorrect load of them especially in the home environment were the patient isn't helped by the clinicians.Therefore, having a constant monitor of the actual state of the patient will reduce this risk. In this contest the wearables offer the possibility of a fine description of the rehabilitation curve of the patient for personalize the treatment. The development of this category of devices started different challenges for the improvements of the characteristics of these devices [36].

First of all they are something that the patients needs to wear and so they have to be comfortable in term of wear-

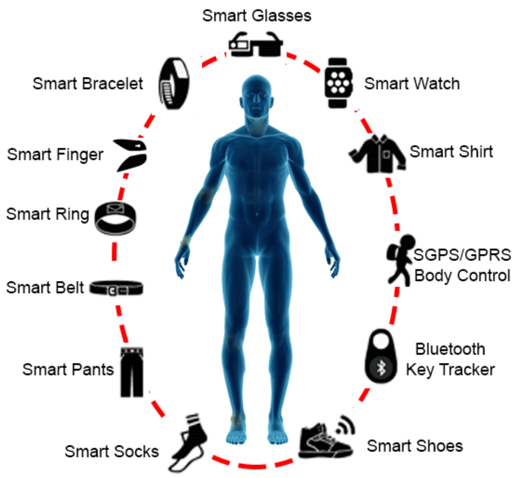


Figure 2.14: Wearable sensors on the market [37]

ability. Secondly they have to monitor a specific parameter of that part of the body.Thirdly, they have to observe the variation of this parameter in a specific contest. For the patient with knee OA the wearability is determined by a sleeve, one of the most significant parameters to monitor is the range of motion and the fundamental activities for observing these changes are walking and stair ambulation.

In a free living environment, when the patient isn't under the supervision of the clinicians he detection of walking among free living activities is required for observing all the small changes in the pattern of walking during the day.This tool will help optimizing the curve of rehabilitation because the continuous monitoring could provide a higher temporal resolution rather than the follow up in the lab twice a week.
In fact, there are some little changes that it is impossible to catch without a continuous monitoring because they can happen when the patient is at home.
When the patient come to the hospital it is possible to catch only a particular moment of the day and of the week, but for an overall, complete understanding it is necessary a deeper knowledge.On the contrary, with the data recorded, the clinicians will be able to analyze information previous missing and so they will better take care of the patients' conditions.
The pro for the patients' perspective will be a better treatment and also the possibility of being assisted also in the home setting with a device that, like the clinician, will suggest the activities to do. In this way it will not only feel better about the improvements of the theraphy but it wil also have have a bettere understanding of his treatment and it will feel essential part of the entire process.
So the main goals are:

- Find the pattern of walking and stair ambulation among all the possible free living activities;
- Observe these little changes in the pattern and making a clinical evaluation that will specifically tailor the theraphy;

From the engineers point of view the goal will be discriminating these patterns among the free living activities using a machine learning techinque in order to define strong and flexible decisions boundaries around these clusters of these activities of interest.

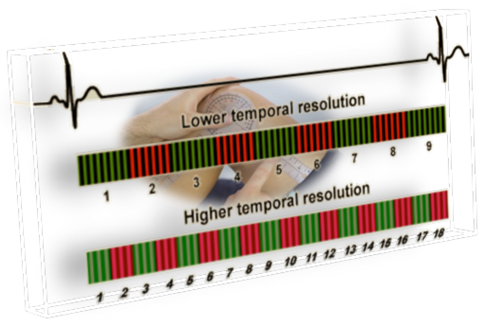


Figure 2.15: fig:Wearable sensors resolution

Chapter 3

State of the art

Over the last decade a lot of long term monitoring devices of human health has been developed for providing valuable information about the stage of the disease and his recovery.
The power of the wearable sensing technologies for health monitoring has been demonstrated but using body-worn sensor data is a nowadays an opened challenge.
Even if a lot of advances has been reached, improvements are still required in the electronics, in the sensor technology, telecommunication and in the machine learning techniques.
In fact, between the existing devices there are several issues like the number, the location and the type of sensors for the accuracy of the algorithm and the wearability of the device for the comfort of the patient.
It is possible to divide the main problems to solve in the following principal challenges.
The first challenge is regarding the activities to detect.In fact, a lot of different activities performed in daily life can be the focus of the reasearch, depending on the final aim of the study.
Therefore, some studies are focus on the detection of a greater number of activities like detection of as lying, standing, walking, running,football, swinging, croquet, playing ball, and using the toilet [38] but with low accuracy, instead other are focus on the recognition of a smaller subset of activities but with a higher accuracy[39].Another complexity factor is that the static activities are simpler to detect than periodic activity like running or riding a bicycle.However, the most difficult issue is to discriminate inside very similar activities like ground walking and stair ambulation because of their similar patterns than result in a overlap in the features space. Moreover, there is the issue that inside the feature space some activities are very similar from one point of view and very different from others point of view.
In this contest the data collected for the training set are very important. Normally the first step is the data collection inside the lab environment using a specific protocol. So the activities are instructed form the member of the lab and they are performed at the same speed and for the same time by all the subjects.
Even if in the real life the activities will be unconstrained this step is essential for make the classifier learn the differences and the similarities between activities. The ideal situation is that the algorithm is trained on a specific data set and then when applied on different subject it would have great performances.
Previous work underline however that independent activity recognition is difficult to achieve due to the high variability between people, especially patients[38]. It was proved instead that algorithms perform better when trained

with more subject-specific training data and so a lot of data are required.On the contrary, there is the necessity to keep as low as possible the data collected for avoid long computational time.
After that, another challenge of the previous device was the choice of the number, the types of sensors and their location on the body. Considering the number of sensors used in the previous studies it is possible to conclude that the power of systems with less sensor is to have low computational time but on the contrary with less sensor is more difficult to improve the recognition accuracy because less information is available.
Regarding the location instead for long term monitoring is essential that the sensors are attached on a comfort position because the patients has to wear them along all the day. So the possibilities offered by studies conducted only in the lab with a complex configuration of the sensor has to be excluded because it is feasible only for the short term acquisition.
Many past works showed that the size and the position of the sensors has to allowed the possibility of doing all the activities avoiding encumbrance that will lead to an incorrect execution of the activities[40].
It has also been demonstrated that there is a within-class variance due to the changes in orientation, magnitude, and frequency depending on the position of the accelerometers.
Looking at the literature it is showed how some of the earliest work in accelerometer-based activity recognition used multiple accelerometers placed on several parts of the user's body like for example Bao and Intille[41] that used five biaxial accelerometers worn on the user's right hip, dominant wrist, non-dominant upper arm, dominant ankle, and non-dominant thigh in order to collect data from 20 users or Parkka et. al [38] that used 20 different types of different sensor,Tapia et al.[42] used five accelerometers placed on different part of the body for real-time recognition of thirty gymnasium activities. Another significant work that used more sensors it is the one of Mannini and Sabitini [43] twith a five triaxial accelerometers system attached to the hip, wrist, arm, ankle, and thigh or Subramayana et. al.[44] that used the combination of a tri-axial accelerometer, two micro-phones, phototransistors, temperature and barometric pressure sensors, and GPS for the discrimination of a stationary state, walking, jogging, driving a vehicle, and ascending and descending stairs.
Even if these systems reach a pretty good accuracy in identifying various activities they demonsytrated to be not feasible for the home setting because of the complexity of the number of sensors that will change position during the daily activities.
A lot of creative solutions have been proposed for overcoming this solution like the one from Choudhury et al.[45] an activity recognition based on the manipulation of object by the users.
All these multisensory approach reach pretty good accuracy but the potential of this thesis it is demonstrate that just with one triaxial accelerometer the detection of walking and stair ambulation in the free living environment is possible.
Other studies has been previously conducted with just one triaxial accelerometer like the one of Long, Yin, and Aarts [46] on 24 subjects with a triaxial accelerometer worn on the waist.
Good results were reached also by Brezmes et. al. [47] that using the Nokia N95 phone develpoed a real-time system for recognizing six user activities. But in this case the activity recognition model require to be trained for each user, the purpose of the model of this work is instead to build and user indipendent algorithm.
Another problem is the choice of the algorithm. Decision trees, k-Nearest Neighbor, Naïve Bayes, and Bayes Net classifiers with five-fold cross validation were used for learning.

For example, Susanna Pirttikangas et al.[48] applicated a KNN classifier and a set of basic feature on data froma tri-axial acceletrometer for the recognition of 17 activities reaching as best accuracy 90.61%.
The same classification algorithm was adopted by Moncada-Torres1 et al.[49] that used it on data from an inertial and a barometric sensor reaching up to 95% accuracy for the classification of 16 activities.
Another different solution was adopted by Fabien Massé1 et al.[50] with a hierarchical fuzzy classifier for the recognition of sitting, standing, walking and lying using data from bothe an inertial sensor and barometric sensor reaching the recognition of 90.4%.
Enrique Garcia-Ceja et al.[51] instead implemented an Hidden Markov Models for the detection fo daily activities like running using acceleration data from a wrist wearable sensor, reaching as the highest accuracy 77%. Also T. L. M. van Kasteren et al. [52] used the sam techinque reaching up to 98.2% accuracy in the detection of 10 activities from real life.
Other works used variations of this model, like the one of La The Vinh et al.[53] use the semi-Markov Conditional Random Fields reaching an avarage precision of 88.47%.
Tao Gu et al. [54] the epSICAR algorithm to discriminate real life activities reaching an accuracy of 90.96%. Tolstikov et. al [55] with the implementation of a Bayesian Networks reached up to 100% accuracy in the detection of activity like showering.
Tˆam Huy‘nh et al. [56] tried different approaches using K-means, Support Vector Machine and Hidden Markov Models for the recognition of 16 daily activities reaching the highest precision (97.6%) in the detection o sleeping.
Other authors like Dean M. Karantonis et al.[57] try the recogntion of 12 tasks reaching an accuracy of 95.6% using a multilayer approach baseD on the calculation of the signal magnitude area.
A hierarchical, binary tree was implemented instead by Mathie et al.[58] data collected from a triaxial accelerometer worn on the waist recognized the activity with an average specificity of 98.7% for the 9 categories.
The same technique was adopted by bonomi et al.[59] that reached the accuracy of 93% for the classification of walking, running and cycling at different speeds using data from a triaxial accelerometer worn on the lower back.
F. Chamroukhia et al. [60] recognized 12 different activities like walking, sitting, standing and stair ambulation testing different machine learnirng technique like the static classifier and the hidden Markov model and different positions (chest, thigh, ankle) reaching an accuracy up to 90.3%.
Lyons et al.[61] reached an overall accuracy for ”moving activities” of 97% with a model with a fixed threshold.
Finally Lei Gao et al.[62] tested three different method: KNN, ANN an Decision Trees reaching more than 96% of accuracy. However, even if there are a lot of works in this field with different solutions in terms of sensor setting and technology, study procedures, features selected and machine learning algorithm, there is still a lack in the investigation of the activity recognition of walking and stair ambulation in the home setting using a single accelerometer that will guarantee low computational time and easy attachment on the body of the subjects.
Another issue is the deficiency in the exploration of the features space which result in a difficulty of discriminate between activities with high similarity and high difference.
Some works explored the frequency and time-domain because even if they require a large number of components for the discrimination and so a high computational power. Time domain features were used as well because the simplicity of the extraction can be easily even if they didn’t lead to a great accuracy. These performances get worse on the home setting.

But even if most of these studies reached very good performances with instructed activities inside the lab there is still a lack in the prediction in a real life environment because the algorithm isn't flexible enough on a new, various, uncontrolled and unbalanced data set.
Additionally, it is still not very clear how to define the borders between activities that are very similar from some point of view and different from others.
So, the literature review underlined these needs:

- A comfortable and easy to use device for the patients
- A reliable activity detection for the clinicians
- A wide and complete exploration of the features space
- A segmentation that can detect both short term and long term activity

The debate about the best sensor platform, subste of activities to recognize, way of segmentation, machine learning technique is still open.
In this work these choiches will be adopted for suit at the best the needs of this paticular applicatio.

Chapter 4

Material and methods

4.1 Hardware platform

The platform of the Kromm (4.1)contains a biomechanically designed electrogoniometer composed of:

- Rotatory potentiometer-based angular sensor that measures the knee angles;
- E-textile sensor that monitors the compliance;
- Three-axis accelerometer placed on the thigh that monitors the activities;
- Bluetooth transceiver that delivers the acquired data to a mobile device;

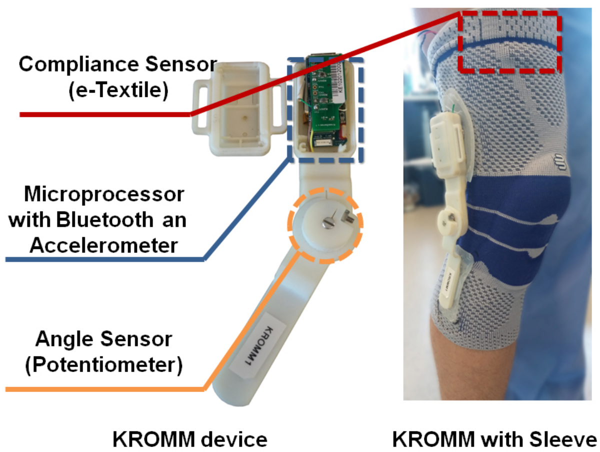


Figure 4.1: The Kromm platform [63]

The Shimmer2R board was the one selected for this application, enclosed in a rapid prototyped case designed by the team of the motion Analysis Lab and in addition to the Shimmer board, it accommodates a 450mAh Li-Ion rechargeable battery. To improve robustness for a long term monitoring, simplify the design and the starting of the sensor, the electrical connection with the external sensors is achieved via a snap buttons on the bottom.

Moreover, two other external tools were used for the second part of the experiment:

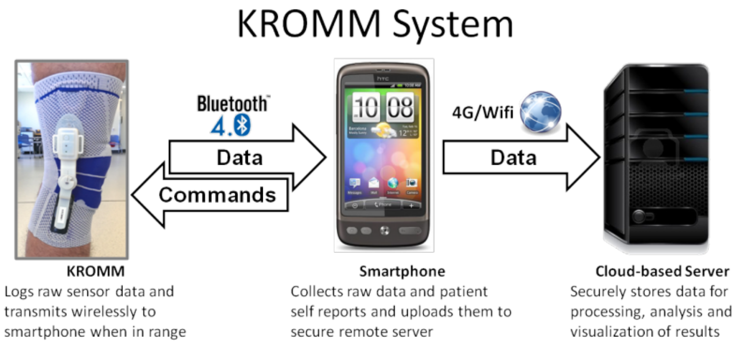


Figure 4.2: The Kromm system [64]

- a wearable camera that captures images every second;
- a smartphone that receives via Bluetooth raw data from the Kromm and for security reasons upload them to a cloud- based served from whom the data are processed and analysed for building the model.

4.1.1 Knee Brace

The DonJoy Spiral Elastic Knee Support (EKS) is characterized from several peculiarities:

- it is made with breathable material for improve the comfort for long term wearing;
- it is made with stretch cotton material that make it simple to wear;
- it has an open patella design that ensure proper alignment of the knee attachment with respect to the knee joint;

The Kromm device was attached to this brace with the extremities on thigh and lower leg with the aim of contrast the valgus knee during the day. In this way the knee abduction moment (valgus moment) and the overload on medial compartment is reduced during walking because the brace counteracts with a moment opposite to the external knee adduction moment.

Therefore, for realizing this aim, is important the fixation and the alignment of the device especially in obese patients that have the highest joint loads and the most difficult transmission of forces because of the great amount of soft tissue.

Another essential function of the knee brace is decreasing the pain, one of the most critical symptoms of knee joint OA.

For all the reasons the patients should wear the knee brace continually. Therefore, it is important that it respect specific ergonomic characteristc for avoiding to be bulky, potentially uncomfortable, but to be practical to wear in daily life and userfriendly.

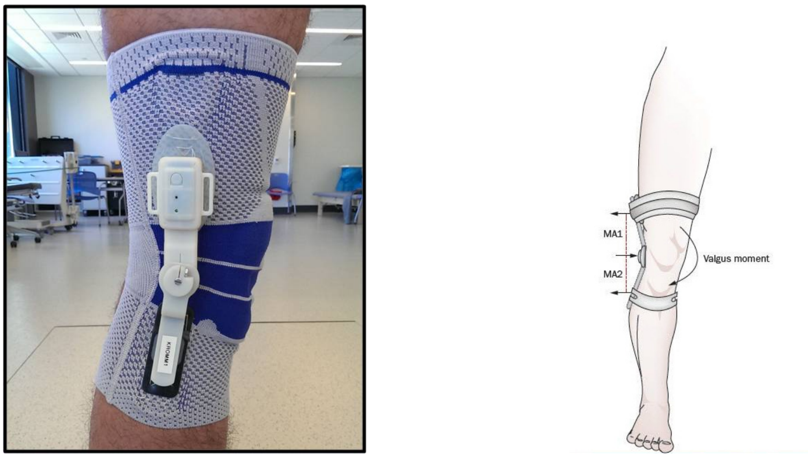


Figure 4.3: Lateral view of the knee brace [28] and exemplification of the structure of the brace [65]

4.1.2 Compliance sensor

The e-textile sensor (marked by the red box in 4.1) is integrated inside the brace with a conductive thread and it has the aim of compliance monitoring with a simple and robust output changing the electrical resistance when it is stretched. There is the possibility of activating or not this sensor thanks to an apposite switch.

4.1.3 Tri-axial accelerometer

The accelerometer is a sensor which measures acceleration due to all forces acting on the device. in these forces are included both the gravitational force due to the mass of the earth and any inertial force which may be applied to the device. Therefore, the two primary components of acceleration are the inertial and gravitational acceleration. The inertial acceleration is due to the application of a force other than gravity to a body. In fact, the inertial acceleration is equal to zero only if the body is completely motionless or if it is moving with a constant velocity. So this type of acceleration is defined as the rate of change of velocity of the body in motion and it is measured in units of m/s2.
The other type of acceleration is the gravitational one due to the natural phenomenon of gravity by which physical bodies attract each other with a force proportional to their masses. The measured vector of acceleration due to this type of acceleration points vertically upwards from the Earth's surface.
So, in a tri-axial accelerometer (an accelerometer formed of three mutually orthogonal uni-axial accelerometers), each axis measures a certain proportion of both the gravitational acceleration and the inertial acceleration. In particular this proportion measured by a given axis depend on the angles between that axis and the directions of the acceleration components.

4.1.4 Extraction of the raw data

The Shimmer2R configuration is contained inside the 256 bytes of the header of the raw data files.
After these 256 bytes, there are the data from the sensor written continuously in blocks of up to 512 bytes until data are logged to a given file. After one hour the file is closed and a new one is opened, with the same format.
In this case, with the accelerometer and the gyroscope enabled, there are three 2-byte channels for the accelerometer

Figure 4.4: The system of acceleration of the Shimmer [66]

Figure 4.5: The orientation of the axes [67]

and three 2-byte channels for the gyroscope in each sample and so the blocks are of 510 bytes.
Moreover, the order, the exact number and the interpretation of the data depends on which sensors are enabled and it follows a specific protocol.

4.1.5 Calibration

The main factors of noise due for the IMU's sensor are for example the range for measure phenomena with an infinite range or the quantisation due to the digital output affect.
Moreover, even if ideally the three axes of the IMU should be mutually orthogonal in reality this orientation isn't very precise because there are misalignment errors due to a slightly incorrect placement of sensors and to an imperfect alignment of the sensor inside the casing.
In particular in the table 4.1 the average noise for different signal bandwidths are showed. These issues can be

Table 4.1: Noise for Shimmer2R with different signal bandwidths

Signals Bandwidths (Hz)	**Accelerometer RMS noise (m/s2)**
50	24.2×10^{-3}
100	36.8×10^{-3}
250	60.9×10^{-3}

Figure 4.6: How the data are stored inside the Shimmer2R [67]

	Scripted data		Unscripted data	
	Bytes	Size	Bytes	Size
Angle	1229600	153700x1 double	25097600	3137200x1 double
Acceleration	3688800	153700x3 double	75292800	3137200x3 double

Figure 4.7: Data contained inside the Shimmer2R

solve defining a rotation matrix R that can allow to the user to calculate the real sensitivity axes from the assumed ones [67].

So for a tri-axial sensor, if the value of the sensed phenomenon vector is $\underline{v}$, then the sensor output, $\underline{Y}$, can be calculated by:

$$\underline{Y}=KR\underline{v}+ \underline{b} +\underline{n}$$

In this formula K is the diagonal matrix of sensor axis scale factors, R is the rotation matrix, $\underline{b}$ is the offset bias vector and $\underline{n}$ is the noise vector.

4.2 Experimental setup and study procedures

The subjects selected for this study were 17 healthy subject, aged 18-64 years, working in the Motion Analysis Laboratory at Spaulding Rehabilitation Hospital.
The requirements were that they would have been able to walk at comfortable walking speeds for 2 minutes without interruptions on a treadmill and that they didn't have any known orthopedic, musculoskeletal, neurological, cardiovascular, pulmonary, or gait disorder that results in an abnormal gait pattern.
In particular, the study involved 2 testing sessions.

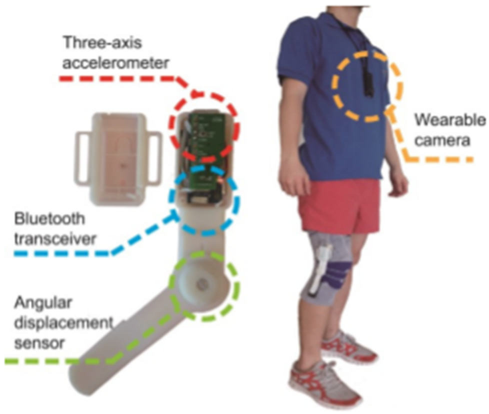


Figure 4.8: Experimental Setup [63]

The first part of the study consisted in a data collection inside the Motion Analysis Lab and it last up to 3 hours. To these subject was asked to perform, while wearing the electrogoniometer, the following instructed activities:

- Sitting;
- Standing;
- Walking;
- Descending stairs;
- Ascending stairs;
- Riding an exercise bicycle;
- Using a rowing machine.

The data collected from this controlled and instructed activities were defined as the "scripted data". The activity of walking, rowing and bicycling were performed at slow, medium and fast speeds.
After that, to the subjects were given also other two tools for the second part of the study:

- A wearable camera wore facing the anterior direction of the human body 4.8, which captured images at 1 Hz for have a reference for the machine learning algorithm;

- A mobile phone with an application for remind to the subject to perform the activities;

The second part of the study involved 4 of these 17 subjects. To them was asked to perform everyday activities at Spaulding Rehabilitation Hospital for 9 hours whithout the clinicians following the execution of the activities and without limits in the activities they were allowed to do.

However, considering that the subjects were wearing the device in a work environment, it was asked to them with the app on the smartphone twice a day to go into the gym and perform the activities of the scripted data. This strategy was adopted for preventing to have a static output.In fact, otherwise they would have been sitting most of the day because of the work environment in which they were during the data collection. The data collected in the uncontrolled part of the experiment were addressed as"unscripted data".

In this work, the scripted data were used for the training of the model and the unscripted data for its validation.

4.2.1 Choice of the activities

The choice of the activities to record was oriented by the analysis of the knee biomechanics in the activities target of the study: walking and stair ambulation.

These activities were selected as" activities of interest" for two main reasons. First of all, because normally these are the dynamic activities that the patients perform the most during the day and so the most important activities to restore in this group of patients for a good quality of life.

Secondly, because it has been proven that monitoring the performances of a specific patient affected by knee OA in these two daily activities and consequently tailoring specific walking program has a clinical impact [68].

For example, in these activities, it is possible to monitor clinical finding like changes in the balance, in the muscle strength, flexibility and proprioception[69]. Furthermore, these are the activities in which is simpler to observe the

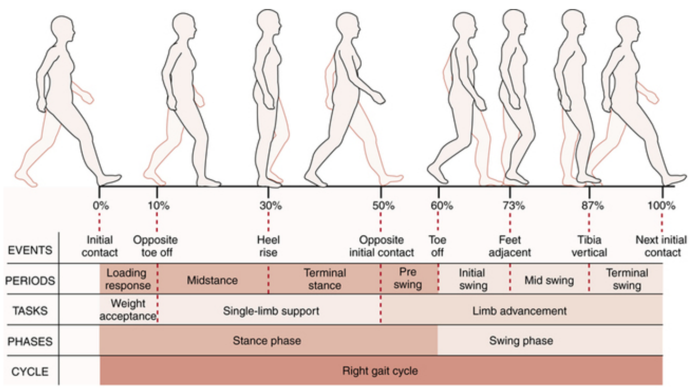


Figure 4.9: Walking phases [70]

altered patterns of movement, correlated mainly to the loss of functionality of the quadriceps femoris that can lead to an overload on the knee joint with an external flexion moment for compensate the balance and consequently to pain and disease progression.

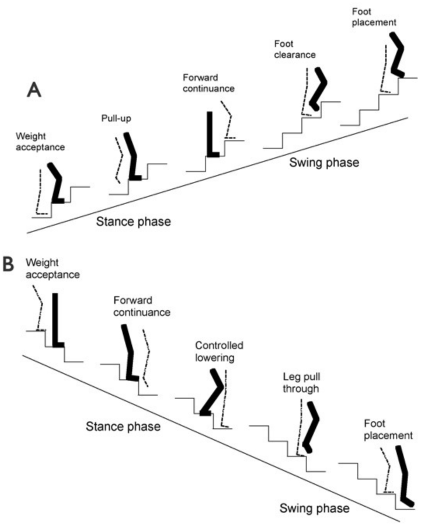


Figure 4.10: Stair ambulation gate [71]

The main ephiphanies of this phenomenon are:

- a smaller peak external knee flexion angle than healthy control subject during ascending stairs [71];
- the increased coactivation of the muscles in both grade and ground walking[72];
- knee flexion from heelstrike to the angle of peak knee flexion before midstance during both level walking and stair descent;
- Knee angle at initial contact;

Therefore, looking at these changes in the gait,it is clear the clinical relevance of monitoring the range of motion(ROM) [73] for assesting the grade of disability and the importance of monitoring this parameter during walking and stair ambulation.

If the "activities of interest" were selected for a clinical purpose, the activities of no interest were selected instead for machine learning reasons.In fact, the classifier has to be feed with a solid training set of data from different daily activities.This step is essential in order to make the classifier learning the differences and the similarities between activities and so to make it being able to discriminate a new data set based on this previous input information.

The start point for this choice is the analysis of the charachteristics of the activities of interest and the space of the features around them. Considering that walking and stair ambulation are dynamic activities the classifier has to learn the difference between static, moderate and very power consuming activities.Moreover, for let the classifier distinguish between different activities with different orientation sitting, half-sitting and half standing activities have to be part of the data collection.

In particular, bicycling, and rowing activities (dynamic activities) were selected to provide similar knee movements as walking and stair ambulation looking at the energy On the contrary, sitting and standing were selected as activities with more limited knee movements (static activities) and so almost a null value of acceleration and of range of motion (despite small fluctuation due to the noise and negligible movement of the body). Moreover, inside

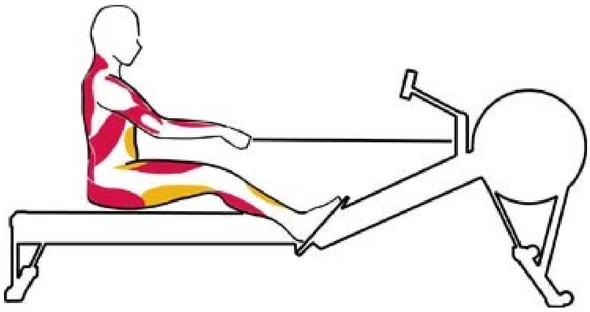


Figure 4.11: Muscle activation during rowing [74]

Figure 4.12: Muscle activation during bicycling [75]

the cluster of the static activities, sitting and standing have different orientation of the body. In particular, standing has an orientation similar to that activities of interest and so, especially when these activities are perfomed at low speed there will be an slight overlap.In this way, the cluster of walking and stair ambulation will be surrounded by the clusters of these activities and the borders around them will be defined. For giving information also about the different velocities and so different values of the accelerations, walking, bicycling and rowing were performed at slow, comfortable, and fast speeds.

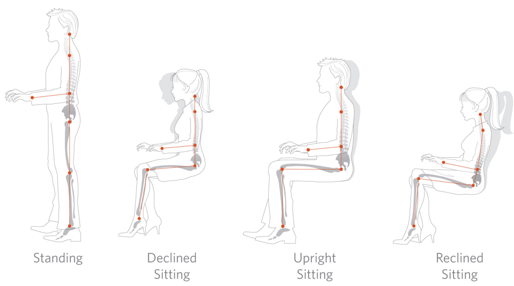
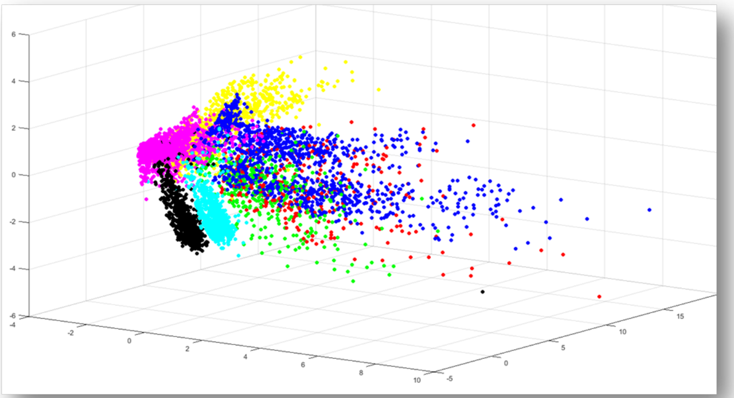


Figure 4.13: Sitting and standing



	Dynamic activities	Static activities
Sitting activities	Rowing	Sitting
Half sitting activities	Bicycling	
Standing activities	Walking Descending stairs Ascending stairs	Standing

Figure 4.14: Overall view of all the activities

4.2.2 Anatomical landmark calibration

Before the subject arrives the Vicon system was calibrated in the space next to the force plates inside the motion analysis lab where is treadmill.
After that, it was done the subject setup and the clusters of markers were positioned following a specific anatomical Landmark calibration.
Finally the following trials were executed:

1. Perform static trial with clusters only;
2. Perform functional calibration for acetabulum:
   - RACant: ask subject to swing right leg antero-posterior and collect for 5 seconds;
   - RAClat: ask subject to swing right leg medio-lateral and collect for 5 seconds;
   - LACant: ask subject to swing left leg antero-posterior and collect for 5; seconds
   - LAClat: ask subject to swing left leg medio-lateral and collect for 5 seconds;
3. Perform pointer calibration for pelvis, medial epicondyle and medial malleolus:
   - RMEstatic: point to right medial epicondyle and hold for 3 seconds;
   - LMEstatic: point to left medial epicondyle and hold for 3 seconds;
   - RMMstatic: point to right medial malleolus and hold for 3 seconds;
   - LMMstatic: point to left medial malleolus and hold for 3 seconds;
4. Perform anatomical landmarks calibration static:

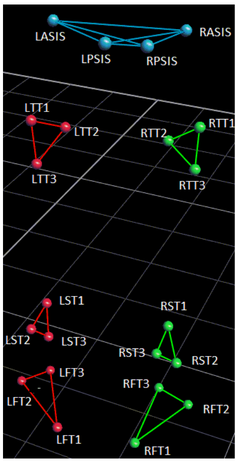


Figure 4.15: Anatomical landmarks

- ALstatic: position extra markers on both sides on:
  - GT (grater trocanter) and LE (lateral epicondyle);
  - HF (head of fibula), TT (tibial tuberosity) and LM (lateral malleolus);
  - SM (second metatarsal head).

This standard biomechanical method allows to extract information about various aspects of gait: it takes data captured by Vicon cameras and convert it into joint angle trajectories, peak flexion/extension angles and range of motion.
Using matlab software the same set of parameters were extracted from the knee sensor for comparison looking at the root mean squared error (RMSE).

4.2.3 Data collection protocol

A different protocol of execution was selected for the first and the second session due to the different time of acquisition and the different freedom given to the subjects in the two part of the study.
However, the two protocols were thought one in function of the other.
Both the sessions were executed inside the Spaulding Rehabilitation Hospital: the first one inside the Motion Analysis Lab, the second one in the all the possible space of the structure, mostly the offices of the Motion Analysis Lab.
This study didn't involve any risk for the subjects except the possibility of fall on the treadmill, fatigue or an allergic reaction to the adhesive tape used to secure the reflective markers to the skin.
For security reason to each patient was assigned a study number, which was used for all documentation except for a master lists (electronic and paper) matching subjects' names and study numbers, and intake interview forms that were kept in a secure location in locked offices. All subjects were informed of their privacy rights and sign a

HIPAA-compliant authorization from previously approved by the Spaulding IRB.

Session 1: scripted data

Before the start of the data collection small reflective markers were placed on the subject's lower body for gait analysis by Vicon motion analysis system and they have to wear the knee sensor.
Data from Vicon Motion Capture System and knee sensor were recorded simultaneously during the testing sessions which will were also videotaped by the Vicon system.
After that, the KROMM device was set: the device was connected to the kneepad, then switched, after that the calibration procedure was completed on the smartphone app and the marker cable was connected.
Before the walking trial 3 swings of the right leg had to be executed for the calibration of the swing.
In the table4.2 the duration of each activity of the data collection. Each acquisition was repeated 3 times; the

Table 4.2: Protocol of the data collection for the instructed activities

Activities	**Duration**
Sitting	2min
Standing	2min
Walking slow (2.2 mph)	2min
Walking medium (3.2 mph)	2min
Walking fast (4.2 mph)	2min
Stair Descent	7th floor – 1st floor
Stair Ascent	1st floor – 7th floor
Bicycle slow	2min
Bicycle medium	2min
Bicycle fast	2min
Rowing slow	2min
Rowing medium	2min
Rowing fast	2min
Sitting	2min
Standing	2min

subjects are allowed to rest for a period of 15-20 minutes after the test.
After the trials the device was switched off and the data from the KROMM device downloaded for check the quality of the acquisition for eventually repeat them.
For example, the movements of the markers accidentally during procedures require the repetition of the Anatomical Landmark calibration.
At the end of the session, the knee sensor sleeve to wear during the daily life and a smart phone to use in conjunction with use of the knee sensor sleeve will be given to 4 of the 17 subjects.

Session 2: unscripted data

During the second session to the subjects was asked to wear the knee sensor in the morning at approximately 9am and keep it on for most of the day.

They were also wearing a wearable camera recording images every second with the option of being turn off and on during the day.

Also the start and the stop of the knee sensor using the smartphone was possible whenever the subjects wanted for discomfort or other necessities like going the bathroom. Despite for that the subjects were doing their typical life, without wearing anything special.

Considering during the data collection the subject were working the app asked during the day to do the activities of the data collection inside the lab for avoid the problem of a static data set.

Finally, at the end of the period of observation to the subjects was asked to provide subjective feedback about the usability of the device, especially in terms of wearability.

4.3 Data Analysis

4.3.1 Challenge

The challenge of the detection mechanism is to design a classifier that will construct more flexible and defined decision boundaries around the walking and stair ambulation clusters. In fact, the optimal model is the one that will perform better on a future data set being adaptive to the changes of the network and being also identity independent.The issue to deal with during the construction of the model is the bias –variance-complexity trade-off (fig.4.16) that can lead to two opposite problems, both conducting to bad performances of the classifier.

In fact, from one point of view a high complexity of the model will lead to the memorization of the noise or of

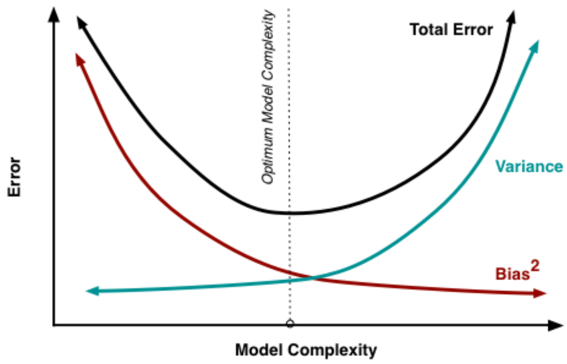


Figure 4.16: Exemplification of the method for obtaining the best model [76]

the outliers that don't represent the underlying structure and the consequence is an "overfitting" (high variance and low bias): the model will have a low generalization and even if it will have great performances on the training set it will not perform well on a new data set.

On the other hand, if the generalization is too high, the complexity of the model is too low and there will be "underfitting" (low variance, high bias) and independently from the amount of data that will be add to th training set it will be impossible to detect the underlying structure.

Furthermore, it has to be taken into account the Okham's razor: "from all models describing the process with the same level of accuracy, the simplest is the best".

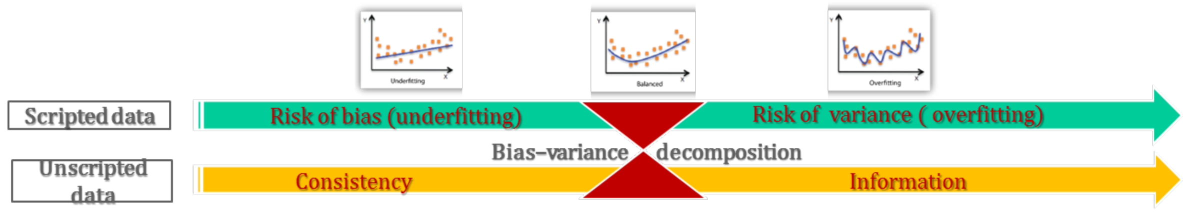


Figure 4.17: Bias-variance tradeoff

4.3.2 Model building strategy

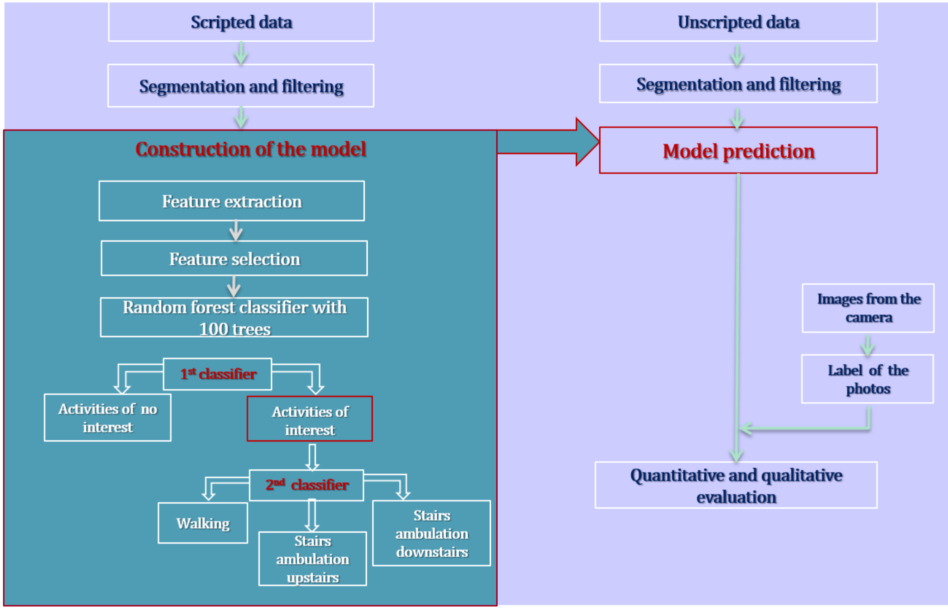


Figure 4.18: Exemplification of the model building strategy

The scripted data will be segmented, filtered and used for the construction of the model.
The first step will be the features extraction, followed by the features selection using the relief and the DBI index and finally the application of a hierarchic random forest classifier with 100 trees.
The first classifier of the cascade will classify between "activities of interest" and "activities of no interest"; then between the "activities of interest" the second classifier will discriminate between "walking", "descending stairs" and "ascending stairs" This model will be used for the prediction on the unscripted data previously segmented and filtered.
Then this prediction will be compared with the label obtained from the images of the camera for realize a quantitative and qualitative evaluation of the model.
The hypothesis is that the segmentation and the label of the instances will influence the features selection and the classification step and the method of cross validation will affect the classification step.
For verify this hypothesis different attempts will be effectuated, changing one parameter at the time starting from a model with the following parameters:

- Window of 5 s with no overlap;
- Different labels for all the activities ;
- 10 cross fold validation method;

First of all, the segmentation will be changed in terms of window width (2.5s, 5s, 10s) and window overlap (0%,25%,50 %). Then, with each one of this different segmentation the features selection and the classifier step will be evaluated. In particular, in the step of the features selection the number and the types of features will be

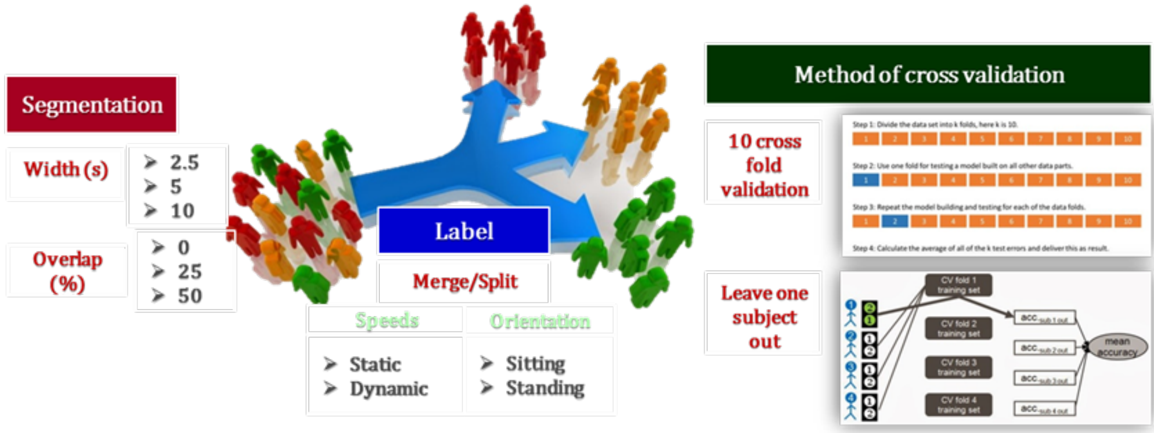


Figure 4.19: Illustration of the different attempts

analyzed, instead in the classification step the rates of the confusion matrices will be examinated.
After that, the label will be modified:

- "dynamic activities "and "static activities" merge together;
- "sitting activities "and "standing activities" merge together;
- All the activities split;

Finally, two different methods of cross validation will be compared: 10-fold cross validation and leave one subject out validation.
The changes will be applied once at the time so it will be possible to see the changes that each modification imply.
So far there is no systematic study evaluating the impact of these steps of the model construction on the features selection and on the performances of the classifier for the detection of free-living activities.
Therefore, the analysis of this work will try to add important information for an optimal choiche of the parameters of the model in the field of activities recognition in the home setting.

4.3.3 Filtering

The denoising of the accelerometers data is a necessary for the good functioning of the detection algorithm, otherwise the features extracted from the signals will not contain relevant information because the output will not just contain the result of the movement of the body but a linear combination with noise.
The noise can come from sources of different natures: internal or external. The mechanical thermal and electrical thermal noise belong to the first category and they are due to passive elements of the electronics such resistors, the effect of the gravity acceleration to the second one.
With the purpose of remove the first source of noise a Chebyshev low pass filter with the following characteristics has been applied:

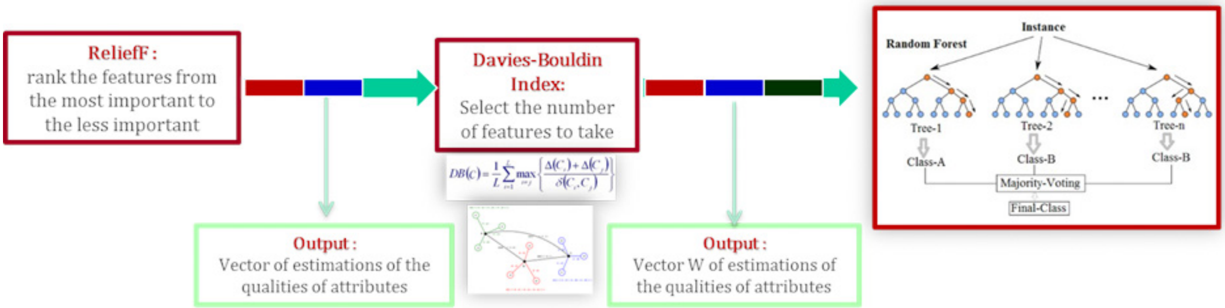


Figure 4.20: Procedure for the features selection and for the classification

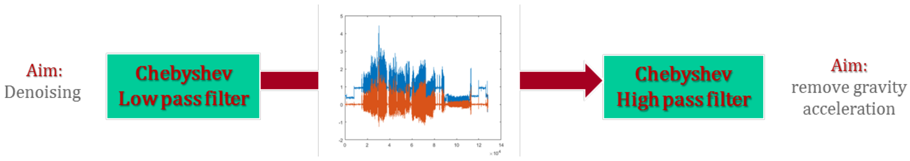


Figure 4.21: Filters chain

- Order: 6;
- Peak-to-peak passband ripple: 0.1;
- Cut off frequency: 20 Hz;

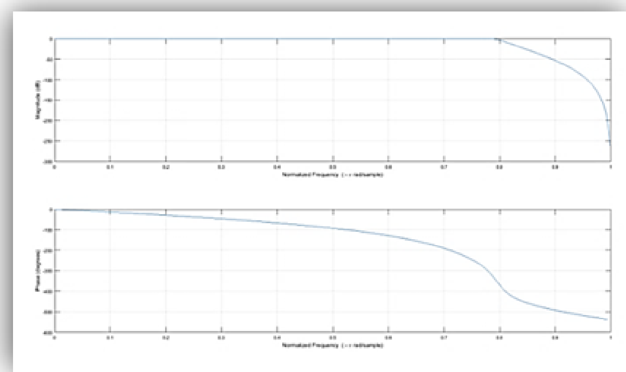


Figure 4.22: Low pass filter

On the other hand, for removing the second type fo noise, the signals were filtered with a Chebyshev high pass filter with these characteristics:

- Order: 6;

- Peak-to-peak passband ripple: 0.1;
- Cut off frequency: 0.5 Hz;

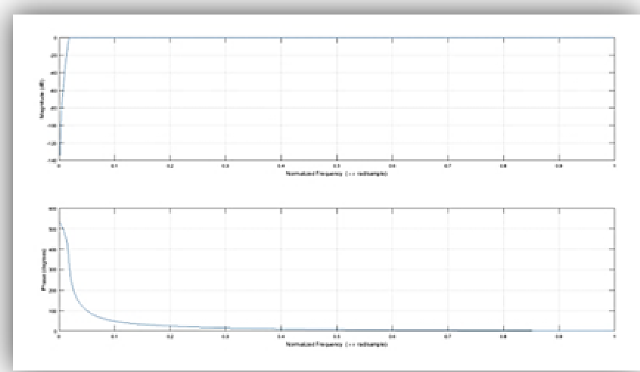


Figure 4.23: High pass filter

In the fig.4.24 the filtered signals from the 3 axes are shown, after the application of both the filters. However, the

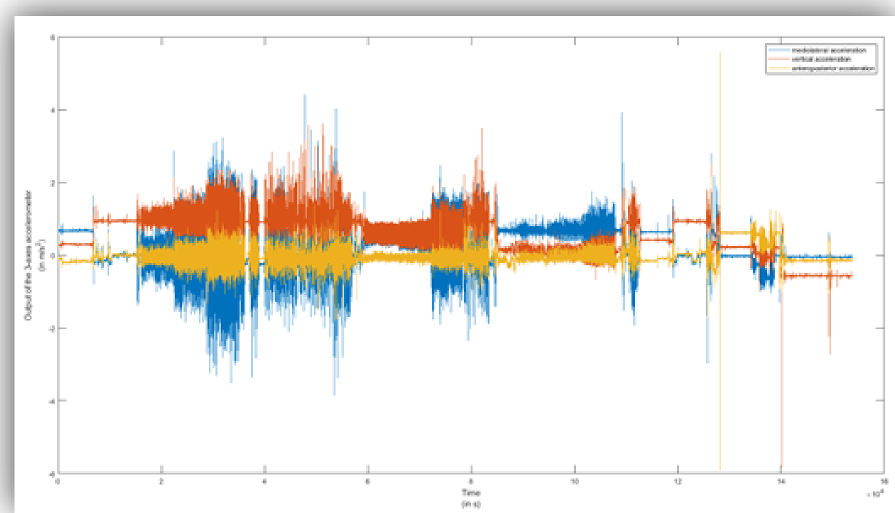


Figure 4.24: Filtered signals

filters don't guarantee a complete separation because of the overlap between the bands of frequency.

4.3.4 Segmentation

The first step for activity detection is divide into smaller time segments (windows) the output of the accelerometers. In this work this processing is off-line and so the features will be extracted separately from each window and then the classifier will be applied sequentially to each sample.

The aim of this part of the work will be understand how different segmentations will affect the features selection and the performance of the classifier and which combination of width and overlap will guarantee the best classification all across the patients and all across the activities.

The most frequently used segmentation techniques are: sliding windows, event-defined windows and activity-defined windows.

The method with event-defined windows requires a pre-processing for detect the events that will be used for the definition of the windows (that will have different sizes); the one with activity-defined windows requires the determination of the instant of change of the activity with, for example, the wavelet analysis (detection of frequency changes); the sliding windows method instead divides the signal into windows of fixed length with no inter-window gaps independently from the signals in input.

The simplicity of the implementation and the adaptability to different input signals make this last segmentation technique the best suitable for this study.

In fact, using a gait event-based segmentation in this case will not lead to good result because of the intra activities and intra patients variability of this system.

Moreover, an approach like that will be convenient also in terms of computational cost compare to wavelets analysis-based algorithms.

One of the most relevant issue to solve applying this method is the choice of the two parameters for define the window: width and overlap.

The choice of these parameters is a trade-off between the amount of information inside each instance and the resolution in the time domain and it depends on the activities that the subject is performing.

Ideally this parameter would be chosen considering the width of the activity that has to be detected. However, for this particular application, there is the necessity of segmenting every activity with the same length.

Looking at the timescale of human movements it has been proved that they go from 160 - 190 ms to around 1-3 s for basic movements. On the contrary, of complex movements like ground walking, stair ambulation and bicycling this duration is indefinite and it depends on the speed at which the activity is done that is correlated with the length of the gate cycle.

If the subject is doing the same activity all the time of the acquisition as it happen during the data collection with instructed activities it is preferable a longer window (10 s) because this will contain more information about that activity and so it will be obtained a better quality of the features estimation and consequently a stronger classifier.

On the other hand, if for example even inside the same activity it is asked to the subject to change the velocity (shorter gate cycle) a longer window can reduce the accuracy of the classification.

This demonstrate that the accuracy of each activity will improve with the choice of a different window. So, theoretically the best option will be segment each activity differently.

However, it is needed a single choice for the window all across the activities that will guarantee good performances in the detection of walking and stairs ambulation in all the possible situations.

In fact, if this classifier has to be applied in free-living conditions, a lot of different activities with different speeds and different durations will be performed. Consequently, it would probably happen that a single window will contain more than one activity and so in this uncontrolled situation the performances of the classifier will get worse because the output will be a media between the values of the features of all the activities contained in the window. The probability of this type of misclassification will decrease reducing the length of the window.
A shorter window indeed will guarantee a higher resolution but on the contrary if the window is too short the output will be more affected by the random fluctuation of th signal.
On the contrary, with a longer window this issue will be reduced thanks to the media between it and the larger percentage of samples of the same activity.
The approach selected for this work is to try consecutive windows of 3 fixed lengths (2.5s,5s,10s) and this choice was based on 2 different motivations:

- the aim of the work is the identification of walking and stairs ambulation that are activities normally performed between 2–10s as a minimum;
- for a clinical purpose it isn't consistent to detect very short pieces of these activities;

The choice of the overlap instead involves the amount of training samples available and the correlation between frame. In this case 3 different attempts were made (25%, 50%,100%). For every window's width and overlap will be evaluated the type and number of features selected, then on the classification step the performances of the classifier will be analyzed.

4.3.5 Exploratory study of the relations between activities

A preliminary exploratory study was done before the feature extraction for a deeper understanding of the similarities and the differences between all the activities, because these characteristics have a strong impact on the signal of the accelerometers.
The aim of this part is to underline different levels of difficulty in the discrimination between activities.
There are a lot of possible differences to take into account but the most relevant are:

- Direction of the motion;
- Intensity (speeds);
- Range of motion;

First of all, it is possible to differentiate the activities looking at the position of the body, so we can divide them between "standing activities" and "sitting activities".
The activities that belong to the first category are: "standing", "walking", "stairs ambulation", to the second category belongs instead "sitting" and "rowing". In the borderline between the two category there is the activity of "bicycling"which is "half-sitting".
Moreover, there is a difference between the intensity of the movement (different speeds): it is possible to discriminate between "static activities" and "dynamic activities".

“Sitting “and “standing” belong to the first category, and all the other activities to the second one. Based on these characteristics of the activities it is feasible to build a network of relation between activities, so that it will be simpler to detect the “most dangerous” borders for the activity recognition.

The exploration of the space of the features helps in the definition of the most critical boundaries in terms of

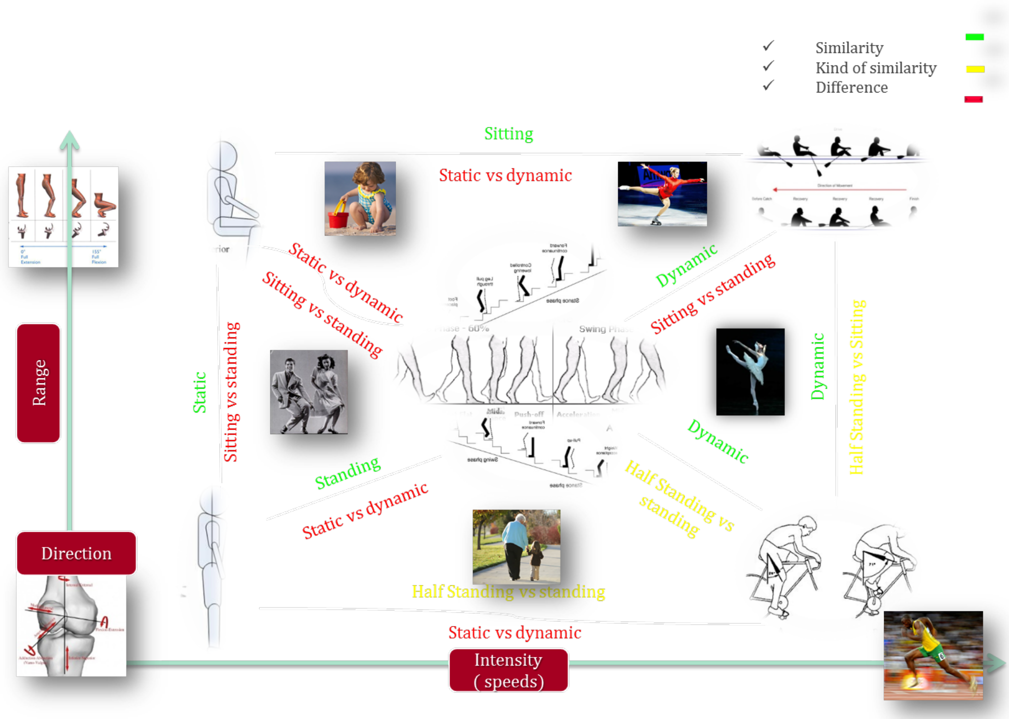


Figure 4.25: Map of the similarities and differences between activities

recognition.

In fact, if for example we look at the border between “sitting” and “standing”, even if both the activities are “static activities”, it is still possible to discriminate between them because of the different orientation of the body.

On the other hand, it is possible to discriminate between “sitting” and “rowing” or between “standing” and “walking” because even if they have almost the same orientation of the body they are characterized by different energies.

The network between all the activities ()4.25)underlined the most critical borders between activities.

The first one is between “bicycling” and the ”activities of interest” because “bicycling” is a dynamic activity and it isn’t a completely “sitting activity” as rowing.

The second one is instead between “standing” and the ”activities of interest” because “standing” is charachterize by a vertical position of the body like ”walking”.

In fact, looking at the sammon map, the borders between these two activities are undefined and so these specific borders will be the most critical for the purpose of the classifier.

For example one solution that will be adopted for avoiding this misclassification between “cycling” and “walking” will be the introduction of the feature of entropy and for avoiding the one between” walking ” and ”standing” the feature of the ”range” of the acceleration.

4.3.6 Features extraction

The previous analysis, highlighting the differences and the similarities among all the activities, was the main structure used for decide the kind and number of features to extract.
So the goal of this part was the extraction of the features, based on this architecture.
In the literature it is possible to find a lot of methods for the features extraction but the most important are: features related to the static (like for example max, mean, variance, and energy), FFT and wavelets decomposition and dimensionality reduction techniques like principal component analysis (PCA) and linear discriminant analysis.
The principal types of features are:

- Heuristic Features: they are normally used for discriminate between static and dynamic activities quantifying the amplitude of the acceleration of the body segment caused by postural transition. In fact, when the subject is moving there is a difference (tilt angle) between this dynamic acceleration and the acceleration of gravity (static acceleration).Features like the signal magnitude area, peak-to-peak acceleration and root mean square belongs to this category;

- Time-domain Features: they are mainly statistical features like mean, median, variance and they are directly derived from the window;

- Frequency-domain Features: before their extraction the data inside the window have to be transformed into the frequency domain typically using a FFT (fast Fourier transform) that gives in output a set of coefficients. They represent the amplitudes of the frequency components of the signal and the distribution of the signal energy and can be used for the characterization of the spectral distribution. Median frequency, spectral energy, spectral entropy are examples of features of this type.

In this work 4 time-domain features, 4 frequency-domain and 3 heuristic features (displacement features), were extracted, each one of them from the 3 axis, in total 33 features.
The choice of extracting features from all the 3 axes was justified by the biomechanical analysis of the activities of interest
These features were chosen using as guide map the one in the fig 4.25with the aim of discriminate the border between the clusters.
In particular, these are the categories of differences highlighted from the map:

- Range;

- Direction;

- Intensity;

These categories of features work very well for most of the borders but when it comes to the most critical discrimination (for example between walking and cycling) other features have to be added.
So, from the combination of all these categories together arise other two categories of features:

- Complexity: this category can help in differentiating activities with different complexity of the pattern. In fact, thanks to this type of features, even if walking and cycling are quite similar looking at the intensity, the direction and the range, they can be differentiated from a complexity point of view: the pattern of "cycling" is characterized by a single dominant frequency, instead "walking" has many FFT components (higher frequency domain entropy);

- Abrupt Changes: this type of features are needed to take in consideration if the change of one of these previous categories has happened abruptly.This phenomenon happens especially between activities like "ascending stairs", "descending stairs" and "walking". In fact, even if these activities are similar from an intensity and complexity point of view,they are different from this side. In the ground walking there is no significant abrupt change, instead in "ascending stairs" the abrupt change is higher and becomes maximum in "descending stairs", especially if the subject is almost jumping doing this specific action.

For improve the performances of the classifier some features were calculated on raw signals, others on the filtered ones.

In fact, for example, for the variance it doesn't make any difference extract features from the raw or the filtered ones, instead for features related to maximum, difference between maximum and minimum of displacement and acceleration or displacement have to be extracted from the filtered ones because otherwise there will be the bias of the gravity acceleration.

Furthermore, the time–domain features were calculated on the signals from the accelerometers, instead the frequency-domain features where extracted from the pxx calculated on the acceleration signals and the features of the displacement on the double integrating accelerometer signals.

For discriminate different range of motion the following features have been extracted:

- Peak to peak: it gives in output the difference between the maximum and minimum values of the input signal, extracted from raw data;

- Occupied bandwidth: it returns the 99% occupied bandwidth of the input signal, extracted from filtered data;

For taking into account the different intensity:

- <u>Mean acceleration</u>: it returns the average value of the signal, extracted from raw data;

- <u>Mean frequency of the acceleration</u>: it returns the mean frequency of a power spectral density (PSD) estimate, pxx, extracted from filtered data;

- <u>Power Bandwidth</u>: it represents the 3-dB (half-power) bandwidth of the input signal, extracted from raw data;

- <u>Mean displacement</u>: it gives in output the average value of the displacement, extracted from filtered data;

For detect different orientation of the body:

- <u>Maximum displacement</u>:the maximum value of the displacement, extracted from filtered data;

- <u>Minimum displacement</u>:the minimum value of the displacement, extracted from filtered data;

	Aims	Features	Type of data
Time domain features	Range	Difference between max and min	Raw
	Complexity	Variance	Raw
	Abrupt changes	Max acceleration	Raw
	Intensity	Mean Acceleration	Raw
Spectral features	Intensity	Mean frequency of the acceleration	Filtered
	Intensity	Power bandwidth: the 3-dB (half-power) bandwidth of the input signal	Raw
	Range	Occupied bandwidth :the 99% occupied bandwidth of the input signal	Filtered
	Complexity	Entropy of the acceleration	Filtered
Displacement features	Direction	Maximum displacement	Filtered
	Intensity	Mean displacement	Filtered
	Direction	Minimum displacement	Filtered

Figure 4.26: Features extracted

For recognize different complexity of the pattern:

- Variance: it gives as output the variance of the signal in input, extracted from raw data;
- Entropy of the acceleration: it returns the normalized information entropy of the FFT components, extracted from filtered data;

Finally, for the detection of activity characterize by abrupt changes:

- Max acceleration: the maximum value of acceleration, extracted from raw data.

The table 4.26 resumes of all the features, with their name, category, aim and the type of data from which they were extracted. All the features were normalized with the standard score normalization for avoid that original dimensions with a larger variance would have had more impact on the principal component.

4.3.7 Label

The most common approach in classification problem is assign a fix label to each instance because the underlying thought is that each sample belong to just one class.
However, in real world applications like the one of this work, it is clear that, due to the great amount of different characteristics that we can attribute to the different activities, it is possible to assign each time different labels depending on the different aspect that has to be highlighted.
The hypothesis is that changing the labels of the samples will make the border between clusters changing as well. This will have as a consequence a variation into the performances of the classifier because modify the label implies giving different information as input.

Looking at the relations between activities shown in the map it is possible to identify the following two different labels:

- “static activities” and “dynamic activities”;
- “standing activities” and “sitting activities” ;

Considering also the fact that ”walking”, ”rowing” and ”bicycling” were perfomed at different speed, it is possible also to split or merge the tree different velocities of walking, cycling and rowing.
This exploratory study wants to understand the way in which the classifier will change his performances and if this changes will be fundamental for the detection of the ”activities of interest”.

	Dynamic activities	Static activities
Sitting activities	Rowing	Sitting
Half sitting activities	Bicycling	
Standing activities	Walking , Descending stairs Ascending stairs	Standing

Figure 4.27: Labels of the activities

4.3.8 Features selection

The process for select the most relevant features for achieve the best discrimination for the problem in examination, is an essential step for the construction of a robust model.
Two are the main reasons that make this step necessary.First of all, that in machine learning a large amount of features can lead to overfitting.
Secondly, the issue of the variability intra subject (changes in the iteration inside the same subject) and between subjects (different way of performing the same movement).
So the goal of this step is to find the feature that will not fluctuate very much between different subjects or different repetition of the same subject and that will allow the maximum distance inter-cluster and the minimum distance intra-cluster.
In the literature it is possible to find a lot of different methods for achieve that aim, in this work it has been used the reliefF in combination with the DBI index.
The reliefF is an extension of the relief with the aim of overcome the issue of incomplete data and the limitation of a two-class problems.
In the fig4.28 the algorithm is illustrated. Even though the ReliefF is a very good method for selecting the raking of the features it doesn’t perform very well when there are redundant features because it has no mechanism for eliminating them. In fact, for avoid this problem, it was necessary a preliminary phase of analysis of the best features to extract.
The DBI index is given by the following formula: $DB = \frac{1}{K}\sum_{i=1}^{K} R_{r,qt}$ where $R_{r,qt} = max_{j,j\neq 1} \frac{S_{i,q}+S_{j,q}}{d_{ij,t}}$.
In fact the DBI is the ratio between the within cluster scatter and the between cluster separation [78]. The distance

Algorithm ReliefF
Input: for each training instance a vector of attribute values and the class value
Output: the vector W of estimations of the qualities of attributes

set all weights $W[A] := 0.0$;
for i := 1 **to** m **do begin**
  randomly select an instance R_i;
  find k nearest hits H_j;
  for each class $C \neq class(R_i)$ **do**
    from class C find k nearest misses $M_j(C)$;
  for $A := 1$ **to** a **do**
  $W[A] := W[A] - \sum_{j=1}^{k} \text{diff}(A, R_i, H_j)/(m \cdot k) +$
$\sum_{C \neq class(R_i)} [\frac{P(C)}{1-P(class(R_i))} \sum_{j=1}^{k} \text{diff}(A, R_i, M_j(C))]/(m \cdot k)$;
end;

Figure 4.28: Illustration of the algorithm of ReliefF [77]

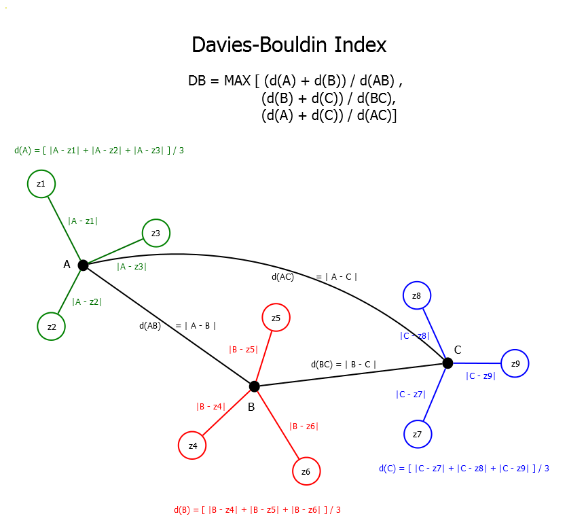


Figure 4.29: Graphic representation of the DBI index [79]

whithin cluster can be calculated in this way: $S_i = \frac{1}{d_i} \sum_{x \in d_i} |x - z_i|$ where d is the euclidean distance.

In this work, the Algorithm of reliefF ranks the features extracted from the most important to the less important for the classification problem and saves the result inside the vector RANKED.

After that, the DBI index has the aim to evaluate how many features use for the discrimination.The DBI index is identified looking at the minimum of the plot in which are represented all the features from the most to the less important. In fact after a certain point, even adding features, the result of the classification will not change and the DBI identify this particular point.

4.3.9 Method of crossvalidation

The aim of this step, performed on the data set split in training and validation set, was to evaluate the predictive accuracy of the model.

The method selected for this validation was crossvalidation: making the data and the test set cross each other is a solution for reduce the problem of the overfitting assuring a relevant amount of data for training the model and for testing it as well.

In particular, in this work, two different partitions of the data set were tested: 10-fold and leave one out cross validation.

In the first approach the dataset is divide in 10 subsets and for 10 times the "hold out method" is repeated using every time 1 fold as test set (validation set) and the other k-1 subsets as training set (4.30).

The power of this method is to reduce the bias because it uses most of the data in the training set and the variance as well because most of the samples are used for the validation. Other important aspects are that the error is obtained doing the average of all the 10 errors in the 10 iterations improving the reliability of the output and that leaves the test sets independent.

The test of 10 fold was made because it has been proved that generally this is the best partition in the classification

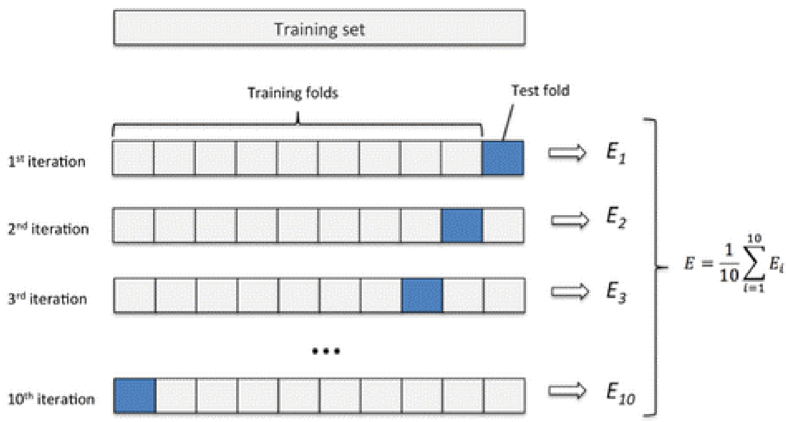


Figure 4.30: 10-fold crossvalidation [80]

problem. In fact, this number f fold is a good compromise between two opposite aspect: increasing the number of fold will assure a generalizable output because almost all the data will be used for train the model but, on the contrary, increasing too much k will increase also the overlap between the folders. Therefore 10-Fold is the best compromise.

In the Leave one out approach, as it is said in the name, one subject is left outside of the training set and this one will be sequentially used as validation set. This process is repeated leaving each time a different subject out for guarantee an identity-independent performance(4.31)

The advantage of this method is the unbiased estimation but on the other hand it introduces high variance and so the output can be unreliable. In this case the error is estimated in the same way.

In this study both the cross validation methods were repeated few times for assure a better quality of the estimation.

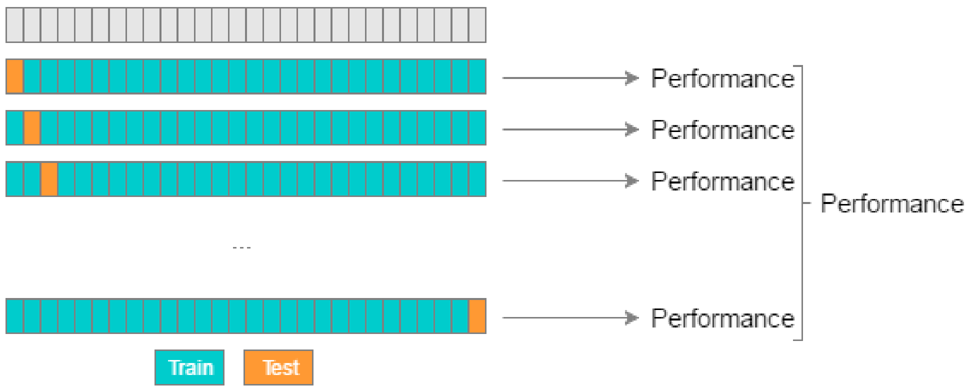


Figure 4.31: Leave one out method [81]

4.3.10 Hierarchic random forest classifier

The classifier implemented in this work is a combination of two methods of classification: a hierarchical approach for realize a cascade of two classifiers for the general structure and a random forest approach for implement each classifier.

The choice of this hierarchical method for the general structure is the optimal one for this application because it is needed to detect the “activities of interest” among the free-living activities (“activities of no interest”) and then, going more in details, recognize inside of them” walking “, “descending stairs” and “ascending stairs”.

Therefore, the features needed in the first classifier will be more general than the ones for the second part. Doing the classification in just one step will make more compact the cluster of the “activities of interest“ that are very similar to each other and at this point it will be very hard to discriminate the activities inside that cluster.

The choice of the random forest classifier was made for having a fast and accurate classification with low bias and relatively low variance.

The idea that underlying this supervised learning method is that combined learning models increases the overall result in terms of accuracy and stability.

This method is called bagging that indicates “bootstrap aggregation” and (with “pruning”) is one of the two approach for avoid the problem of variance.

In bagging this aim is achieved by averaging a set of observations.

Bagging is preferable than pruning because it is possible to reduce variance without adding bias. In the bagging process there are multiple training sets and for each one of them a single tree is built. However, this implies a problem of correlation between tree because for construct different trees it uses the same features and so the trees will have similar prediction.

The Random Forest classifier solve the problem of bagging decorrelating the tree choosing a random subset of the original features for construct the tree[82]. Consequently, this multiclassifier will be formed by an ensemble of decision trees (the predictive models) with different sizes and branches that will make different votes and the final prediction will be the class with the maximum number of votes.

This randomness will improve the performances of the model especially in the contest of a free living environment.

Two types of hyperparameters are useful for define this model: the ones of the decision-tree classifier for the single

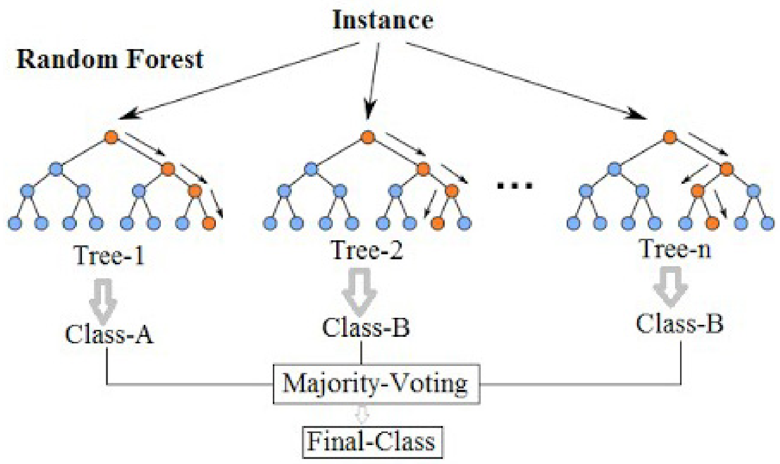


Figure 4.32: Illustration of the random forest algorithm [83]

tree and the ones of a bagging classifier for the ensemble of the trees.
A correct choice of this parameter can improve the power of the predictive model and its velocity.
The following hyperparameters can improve the power:

- nestimators: the number of trees builded from the algorith before the final prediction. The larger this number the better the performances and the stability but the higher the computational
- maxfeatures: the maximum number of features Random Forest can try in a single tree ;
- maxfeatures: the minimum number of leafs required to split an internal node;

On the other hand, the following hyperparameters can improve the speed:

- jobs: If it has a value of “1” tells to the engine to use one processor instead with “-1”it can use an unlimited number.
- randomstate: makes the model’s output replicable
- oobscore: tells the random forest cross validation method.

One of the main advantage of that method is the easiness in measuring the relative weight of each feature on the prediction thanks to a tool that compute a score for each feature measuring a features importance by looking at how much the tree nodes reduce impurity across all trees in the forest. Another pro it’s the relatively user-friendly implementation and the fact that the default hyperparameters produce normally god performances and even in the case in which it is required to change them the process is relatively easy and understandable.Furthermore, this algorithm helps avoiding the overfitting problem choosing the right number of tree.
On the contrary, the most critical issue with Random Forest is the time required for the prediction. In fact, the computational time is a function of the number of tree: for a more accurate prediction more trees are required. This is a real problem for real-time predictions but it isn’t the case of this off-line work.

4.3.11 Validation on the unscripted data

Two different evaluations of the performances on the unscripted data will be done: a qualitative and a quantitative analysis.

Qualitative evaluation

The qualitative analysis will be a visual inspection of a vector constructed in the following way: a different colour for each activity recognized and white colour for missing pieces of signal (because for example the subject was going to the bathroom). This vector will be realized for giving an immediate overall vision of the prediction of the classifier. In fact, considering that the subjects were wearing the Kromm in a working environment, the prevalent colour in the vector has to be black.
Another important focus of this visual analysis will the transitions between activities.
For quantify this visual inspection two different parameters will be defined: percentage of stability and percentage of correct transition.
The percentage of stability evaluates the consistency of the output. This parameter is evaluated for avoid the problem of an unreliable output or, on the contrary, an ouput poor of details as well.
At this stage the stability can be calculated in a qualitative way just as preliminary test of the performances of the classifier on the unscripted data. In particular this evaluation will be done giving 1 to one element of the vector if the previous and the following one are equal (for the windows of 10 s), same argument but evaluating +2 and -2 for the window of 5 s and +4 or -4 for the window of 2.5s.Then, after that, all the stable instances of the signal are summed, and this value is the stability.
The percentage of correct transition will be calculated based on a system of score 4.34that rates the coherence of the recognition. Particularly the output is considered "coherent" if the sequence of the transition is meaningful.For example, there is an high probability that the recognition of "walking "will be followed by "ascending stairs" or "descending "stairs, on the contrary it's impossible that the activity of "sitting" is followed by stair ambulation. Based on the consideration a network of possible transition was created. In particular, two action with a high probability of transition were connected by a line with a consistent thickness, instead a low probability was indicated by a less thick line and an impossible transition with no line.
The system of score built on that network was defined as follow: the initial score was the number of instances and for each detection of a low probability transition 1 point was detracted from that sum, for an impossible transition 2 points. Another evaluation was conducted looking at other sings of a good prediction:

- the sequence of speed: in the data collection was asked to the subject to perform walking slow, medium and fats in this order;

- the sequence in the stair ambulation: the subjects were on the 3rd floor so a good prediction has to show descending stair follow by walking and then ascending stair or nothing, it's very unlikely that a signal of ascending stairs is followed by descending stairs, this could have happened only in the part of the day in which the app asks to the subjects to perform a certain activity;

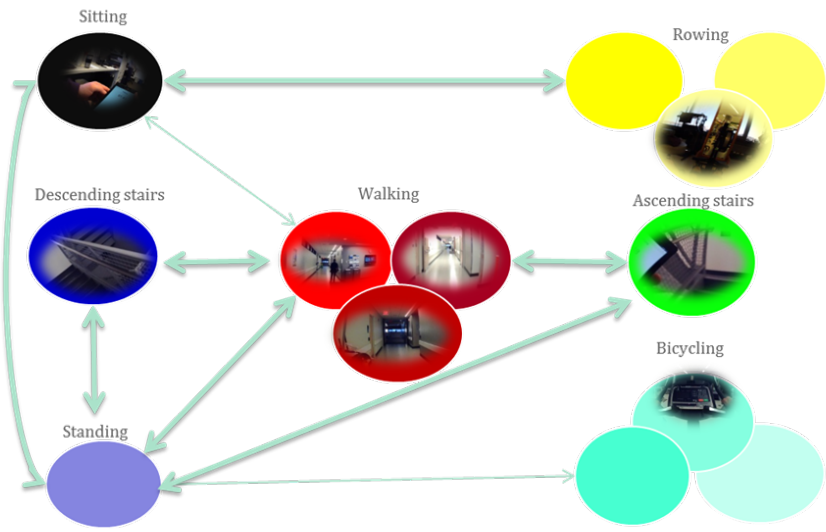


Figure 4.33: Probability of transition between activities

Initial score
total number of
samples

Line width	Probability of transition	Score
	Impossible transition	Minus 2
	Low probability of transition	Minus 1
	High probability of transition	0

Figure 4.34: Method of scores

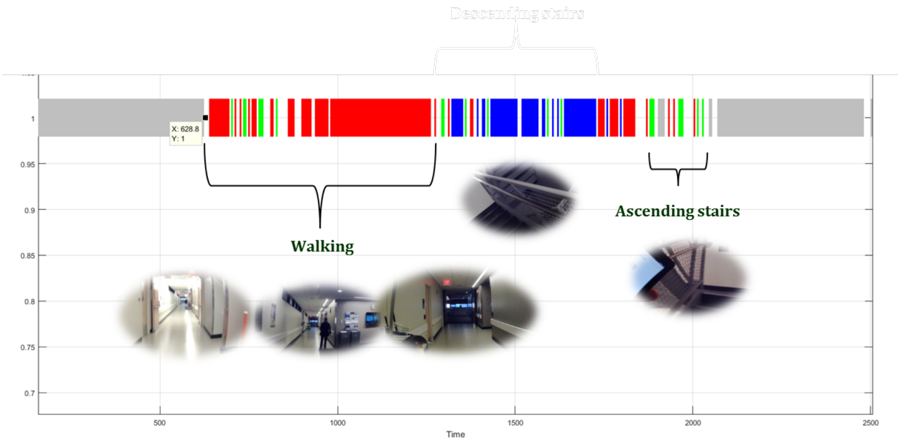


Figure 4.35: Visual inspection

- the sequence between standing and sitting: if standing is recognized after sitting or vice versa it means that the algorithm is working well;
- the sequence between stairs and walking: find short pieces of walking in the middle of the stairs take into account the pier.

Quantitative evaluation

For a deeper evaluation of the perfomances of the classifier also a quantitative evaluation will be conducted using as a standard goal the output of the camera. The goal will be to calculate the accuracy and all the confusion matrix rates also from the unlabelled data with the "label" provided by the photos. In particular, both the output of the classifier and the photos of the camera will be translated in the time domain for a direct comparation and so for constructing the confusion matrix. So, for this aim two vectors will be created, with the length of the seconds inside the almost 9 hours of acquisition.
Moreover, the photos from which it will be impossible to determine the actual activity were deleted from the vector for avoid to have an unreliable standard goal.

Chapter 5

Results

The aim of this section is understanding how and how much the segmentation, the label and the method of cross validation affect the features selection and the performance of the classifier for building an accurate and flexible detection algorithm.
The starting point will be a model will be the following characteristics:

- a window of 5 s with no overlap;
- different label for all the activities;
- 10 cross fold as method of crossvalidation;

Then, changing one parameter at the time, the results of the different attempts will be showed.
Initially, the segmentation will be change in terms of window width and overlap and it will be observe how this changes actually affect the features selection and the perfomances of the first and of the second classifier. After that, the label of the instances that are use for training the hierarchical classifier will be modify and the perfomances will be evaluated for each one of the different configuration.Finally, the two method of crossvalidation 10-fold and leave one out) will be compared for both the classifiers.
All the performances of the scripted data will be evaluated comparing the output of the first and of the second classifier with the actual output given by the labels.These two vectors will be compare inside the confusion matrices and the number and type of missclassified for each class will be analyzed.
In particular, the accuracy, the specificity, the sensibility, the precision and the recall will be the rates of the matrices that will be taken into account for the optimal choices for the final model.
After this session of results, a qualitative evaluation will be conducted on the unscripted data for understandig which configuration would give the best performance on this data even if the same configuration will have given slightly worse results on the scripted one.
After this two steps, the final parameters will be selected and the final model will be evaluated. First, the features selection step will be showed, with the plot of the DBI index for the first and the second classifier and the features selcted for each one of them.After that, the PCA of the two classifiers before and after the step of the features selection will be analyzed for demonstrate the validity of the features selected. Then, the performances of the final model in terms of rates of the confusion matrices will be inspected for both the classifier of the hierarchical

structure.
For making the result of the confusion matrices more tangible different visual plots will be showed. In particular, one will illustrate the differences between the label and the output of the classifier, and the other one will show the missclassified between the most critical borders.
After the quantitative evaluation of the perfomances on the scripted data, a final analysis will be effectuate on the unscripted ones using as label the output given by the camera. In particular, the stability of the output and the distribution of false negatives and false positives has been evaluated for understanding if the misclassification is negligible or relevant for the purpose of this work.
The analysis on the scripted data will lead to the choiche of the optimal segmentation, label and method of cross-validation comparing the rates of the confusion matrix and the PCA of the different models.
In particular, it will be showed how the number of instances, the type of instances and the correlation between them affect the features selection and the perfomances of the classifier.
At this stage the main risk is to avoid the overfitting.In fact, even if the algorithm has to design define borders around the clusters of the ”activities of interest”, it has to be flexible and so to adapt itself to the changes in the training set for having good perfomances also on the unscripted data.

5.1 Impact of different segmentations

On the scripted data the results showed a consistent impact of the segmentation in terms of window overlap and window width both on the feature selection and on the classification step.
In particular, as it is possible to see from figure 5.1, on the step of the feature selection, they affect the number and the type of features.
First of all, considering the number of features it is possible to observe that for both the classifiers there is a decreasing number of feature selected increasing the width of the window.
Moreover, changing the overlap instead affects mainly the feature selection for the second classifier but it has no relevant impacto on the first classifier. Considering instead the type of features selected changing the width and the overlap,it is possible to observe that they have an impact on both the classifiers.
Specifically, a window of 10 s will highlight features about the intensity or the range of the action but it will not be specific in the detection of abrupt changes. Therefore, even if this type of window will allow the activity detection with an high accuracy on the scripted data because of the fixed length of the data recorded, it will not perform well on a new data set with rapid and instantaneous variations of the activities. However, the pro of this window length is to allow the selection of features that describe very well the parameters of a specific activity despite the random fluctuation of the signal.
Considering the window of 2.5 s it is possible to see that most of the features selected are mainly oriented for spike detection. In fact, this width of window is too short to give reliable information about the activities dynamics but it can only detect abrupt changes.If features of this type are very important for discriminate inside the cluster of the activities of interest they will lead to bad perfomances on the first classifier because they don’t give enough information about characteristic like the range and the complexity of the signal and so they will not help in dis-

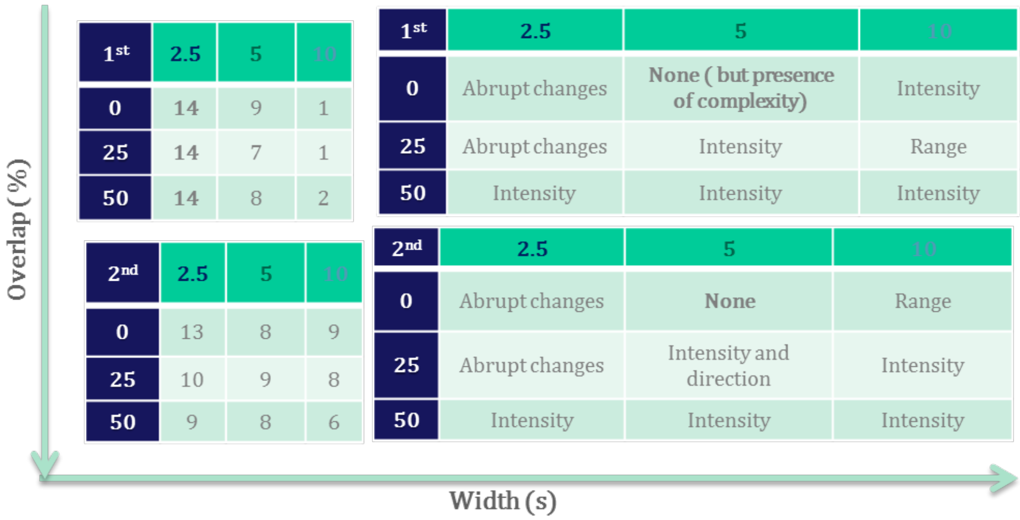


Figure 5.1: How changing the segmentation affect the features selection: on the left it is illustrates how the window width and the overlap affect the changes in the number of features, on the rigth how these changes affects the types of features selected

criminating between tha activities of interest and the activities of no interest.

Given this considerations, a window of 5s seems the best suit for the purpose of these works because it selects features from almost all types, without giving any preference to one specific kind of features. The selection of different types of feature makes the algorithm flexible and adaptable to a new data set and so this window length is the best performing on a new data set.

Looking at the fig.5.1 it is also possible to underline that increasing the overlap decrease the variation in the types of features so for all the reasons analyzed above, the optimal choice for the overlap is no overlap.

After the examination of the impact on the step of features selection the evaluation of the impact on the perfomances of the classifier was conducted.

Looking at the rates of the confusion matrix obtained from different segmentation it is possible to observe that the windows of 2.5s has the worst performance both on the 1st and on the 2nd classifier. On the contrary, comparing the windows of 10s and 5s it is shown that they have almost the same false positive rate and the same sensitivity in both the classifier but on the first classifier the 5s window has slightly better performances and on the 2nd the 10s is slightly better in the detection of walking.

Considering the performances on stair ambulation the differences between the two segmentation become higher: 84% and 87% accuracy with the window of 5 s against 90% and 91% of the window of 10s.Therefore, at this stage, the window of 10s seems the best suitable for the application.

After the analysis on the impact of the different segmentation on the scripted data, the hypotesis formulated during this part of the result were tested with a qualitative evaluation of the impact of different segmentation on the perfomances on the unscripted data. In fact, this visual inspection of the plots of the output of the classifier with different window widhts and overalps was observed for have an understanding of the possible future result on the scripted data.

In particular, the analysis reveals that a window of 2.5s is too short to be consistent because looking at the band

1st classifier	No overlap			25% overlap			50 %overlap		
	2.5s	5s	10s	2.5 s	5s	10s	2.5 s	5s	10s
Accuracy	99.53%	99.42%	95.53%	88.60%	99.86%	90.51	88.71%	99.88%	90.66%
Sensitivity	99.08%	98.51%	94.38%	83.48%	99.89%	86.08%	83.37%	99.90%	86.03%
Specificity	99.76%	99.87%	96.09%	91.18%	99.85%	92.73%	91.39%	99.87%	92.99%
Precision	99.52%	99.74%	92.28%	82.58%	99.70%	85.53%	82.94%	99.75%	86.12%

Figure 5.2: How changing the segmentation affects the perfomances of the first classifier

2nd classifier	No overlap			25% overlap			50 %overlap		
	2.5s	5s	10s	2.5 s	5s	10s	2.5 s	5s	10s
Accuracy	93.3%	98.54%	98.53%	93.51%	97.15%	98.34%	94.17%	97.27%	98.13%
	92.64%	98.20%	97.97%	93.19%	96.23%	97.99%	93.75%	95.92%	97.78%
	92.69%	97.98%	97.18%	93.11%	95.81%	96.80%	93.92%	95.65%	96.61%
Sensitivity	95.70%	95.80%	98.90%	96.59%	98.10%	99.26%	96.80%	98.41%	98.33%
	76.98%	84.20%	93.71%	77.85%	86.22%	91.61%	80.92%	87.04%	94.00%
	81.83%	84.11%	93.48%	78.19%	91.92%	91.98%	80.20%	89.66%	91.62%
Specificity	93.21%	99.06%	97.96%	88.02%	95.71%	96.72%	89.52%	95.57%	97.79%
	81.98%	99.14%	98.90%	96.40%	98.46%	99.29%	96.41%	97.92%	98.58%
	83.79%	98.98%	98.15%	96.55%	96.89%	97.95%	97.14%	97.32%	97.82%
Precision	95.70%	95.89%	98.72%	93.47%	97.18%	98.17%	94.22%	97.08%	98.70%
	76.98%	84.21%	94.90%	81.91%	92.60%	96.32%	83.39%	90.37%	93.38%
	81.83%	85.82%	92.97%	83.91%	89.10%	91.41%	86.80%	90.28%	91.34%

Figure 5.3: How changing the segmentation affects the perfomances of the second classifier

Figure 5.4: Predicitions of the classifier on the unscripted data with different segmentation. It is possible to observe how changing the window width and the overlap affect the output

it is possible to observe that the colour of the instances changes very fast from one window to another and so it is very hard to find relatively long recognition for a specific activity that can be used as relevant information for clinical support.

The output in fact isn't stable and this lack of stability is an issue also for the future possibility of a future correction of the false positives and false negatives. On the contrary, even if a 10 seconds it guarantees a stable output, it is weak in the recognition of the start and of the end of the activities because it has low recognition and the media that it naturally does on the signal will make loose information about some activities if their are perfomed for less than the window width.

A window of 5s is instead a right balance between these two issues: the recognition of the activities is very likely to the one with 2.5s but it avoids a very sensible recognition or the excessive filtering media of 10s window. As it is possible to observe from the fig.5.4the colour is more uniform than 2.5s window but less than 10s window.

Furthermore, this visual inspection highlights the coherence all across the prediction even with different segmentation which is an ulterior corroboration that the classifier is working well.

A further evaluation was made on the possible performances on the unscripted data in order to assest the conclusions made after the analysis on the scripte one. In particular, using the system of scores described in the previous section, it was possible also to observe that the score of sense reached the pick for a 5s width, that the sensibility of walking gets higher with a larger window, the sensibility in the detection of the stair ambulation get higher with a shorter window and finally that a larger overlap improves the stability, especially in the detection of the stair ambulation.

A window of 10s hasn't enough resolution for detect different speeds(different length of the gait) because normally a gait last less than 10 s even in patient with knee OA. Therefore this window will have a lot of misclassified because it gives relevance especially to the energy of the activiies but not to information about the specific shape of a gait cycle.

The weakness in the recognition of detecting stair ambulation is hidden inside the lack the media that this segmentation does smoothing the spikes that characterize the differences between these activities.

Moreover, for the same reason of the low resolution, this window doesn't select features about complexity so it's not able to memorize difference between walking and activities with the same energy but different pattern like for example "walking" and "cycling".

5.2 Impact of different labels

The results of this section will show that giving different labels to the instances in the training set change both the features selection and the performances of the classifier because different labels generate different dimensions and shapes of the clusters in the traininig set.
In fact, different features are required for different shape of the clusters in the training set and so different label will affect the features selection step. Furthermore, because this last step affect the performances of the classifier on the training set, it will be showed how giving different labels affect this step too. Both the PCA scatter plot and the confusion matrices will lead to focus the learning of the classifier along the most critical borders, directing the final choice of the label and consequentally the type and number of features to extract.
First of all, the classifier has to be fed with instances of static and dynamic activities to learn to discriminate between the "activities of interest" and activities like "sitting" or "standing". Moreover, inside the cluster of the dynamic activities, the classifier has to learn how to discriminate between different level of dynamism and different complexity of patterns. As it is possible to see from the scatter plot of the PCA (fig.5.5) the borders between the static and the dynamic activities are well defined.
These first projections were obtained with the basic set of features, without features about the complexity of abrupt changes. In fact these last features were added in a second step for better discriminate along the most critical border but they were excluded from the first plots. The aim of the first projects was to identify the issue in the discrimination, between the "activites of interest" and the "activities of no interest" and inside the cluster of the dynamic activities itself.
Looking at the PCA scatter plot it is possible to see that there are few overlaps. The labels to give to the instances has to be chosen with the aim to make the classifier learn the differences between the two activities of the overlap.
After that, another attempt was made in order understand shape of the clusters of the " sitting activities" and

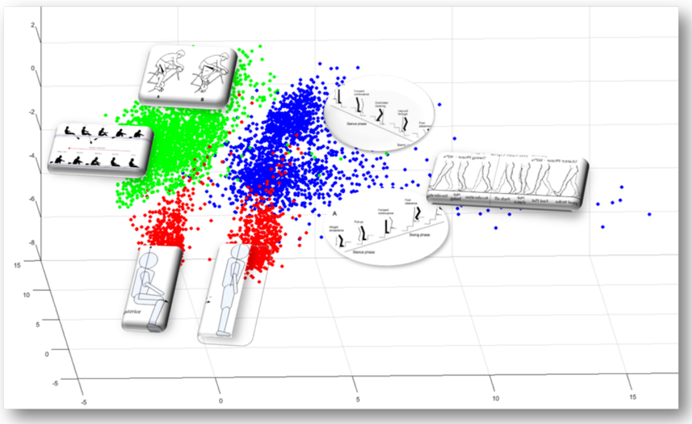


Figure 5.5: PCA scatter plot with 3 different labels: static activiies (red), dynamic activities(green)and activities of interest (blue)

" standing activities"(fig.5.6.Also in this case the basic set of features was able to create a compact distribution. However, with just this set of features there are few overalp along the most critical border between cycling which is an half-standing activities and for example walking that has almost the same orientation of the body and almost the same energy. The overlap between cycling and sitting because is negligible instead because even if these two activities has the same orientation, they have a different level of energy.

After the analysis of the critical borders using just the basic set of features, a further analysis was conducted for

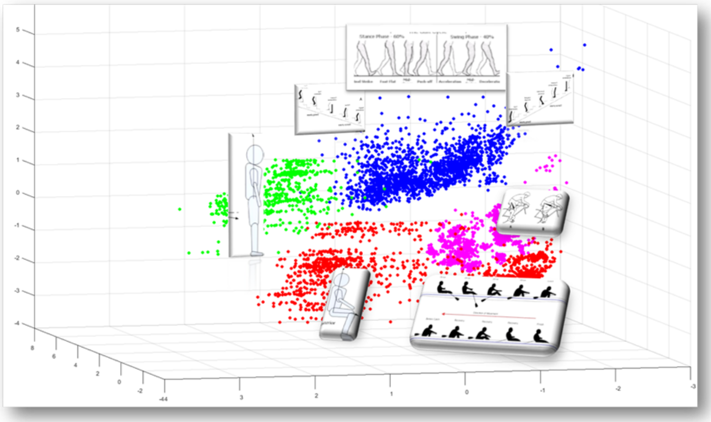


Figure 5.6: PCA scatter plot with 4 different labels: sitting activiies (red), standing activities(green), half-standing activities (purple) and activities of interest (blue)

seeing the effect that the choice of adding other ”special” features implied along the most critical borders.

The analysis of the clusters of cycling and the activities of interest reveal another critical border, this time due to

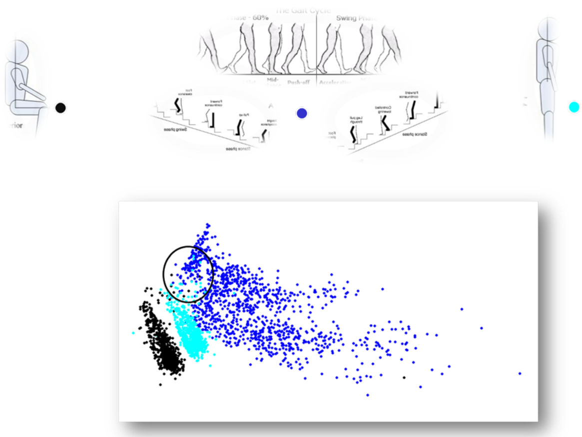


Figure 5.7: Focus on the border between the activities of interest and the static activities

both similarities in the orientation of the body and in the dynamism. The border between rowing and the activities of interest is well-defined and this is not a critical one. There is only a small overlap due to the same level of energy of some instances. Moving the analysis inside the cluster of the activities of interest an ulterior fine discrimination is required and so a deeper examination of the principal direction of variance. In this cluster the discrimination is more difficult because the differences between the activities are reduced.

Specifically, starting from the border between walking slow and the activities of interest the projections showed 5.10 that the component responsable for the discrimination between these two activities is the first principal component which explaines the 54.46% of the variance. So not only this component allow the discrimination but it

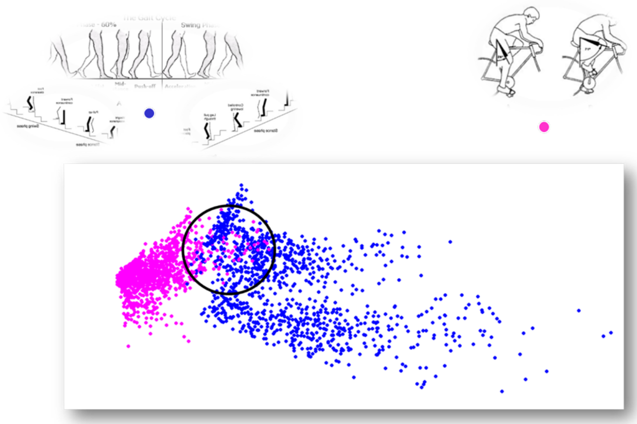


Figure 5.8: Focus on the border between cycling the activities of interest

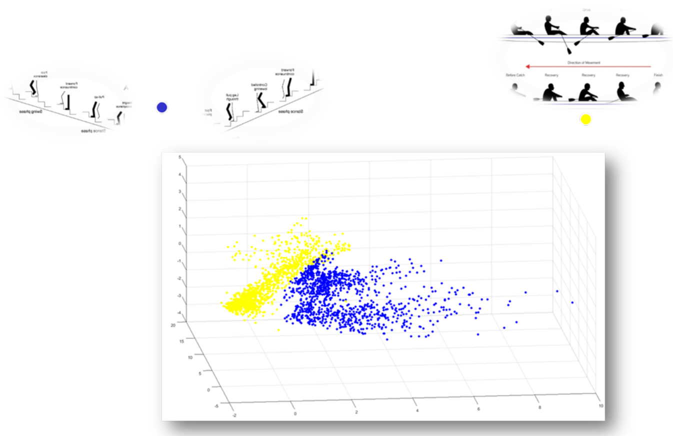


Figure 5.9: Focus on the border between rowing and the activities of interest

also explain half of the total variation.
In particular, the feature that has the greatest weight in this recognition is ’Peak2PeakRawY’ which take into account the abrupt changes along the vertical direction. In fact this feature was extracted because even if walking slow and ascending stairs are similar for the energy and the orientation of the body, they are different because of the presence of a peak in the signal of the stair ambulation.

Looking in details at the border between walking fast and descending stairs 5.11 it is possible to see that the third principal component separates these two activities and that this component explains the 17.07% of the variance. Therefore the variables loaded in the third principal component explain the differences between the instances but their effect is smaller compared to the other components.
In particular, the features that is more responsable for this discrimination is ’MeanRawz’ because of the greater movement back and forward of the device during stair ambulation.

If the PCA showed how different label affect the shape and the overlap between cluster, the confusion matrix showed in a quantitative way how these changes affected the performances of the classifier. The result shows a change in the type of feature selected. Specifically, merging activity with different speeds will affect the informa-

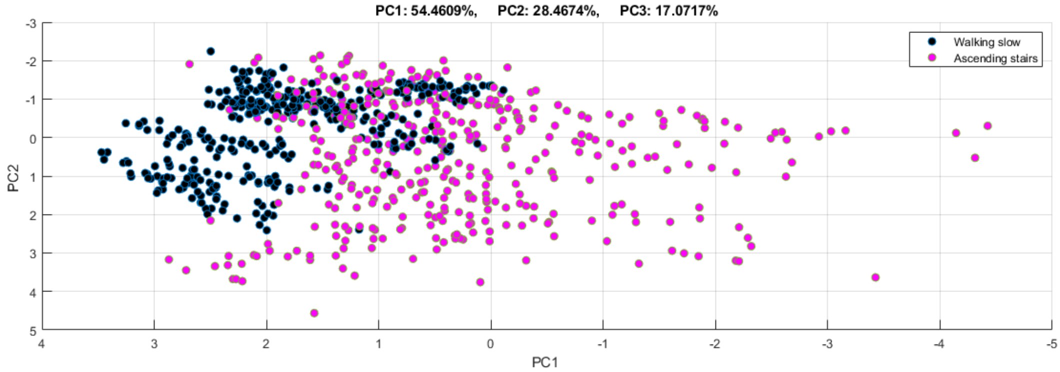


Figure 5.10: Focus on the border between walking slow and ascending stairs

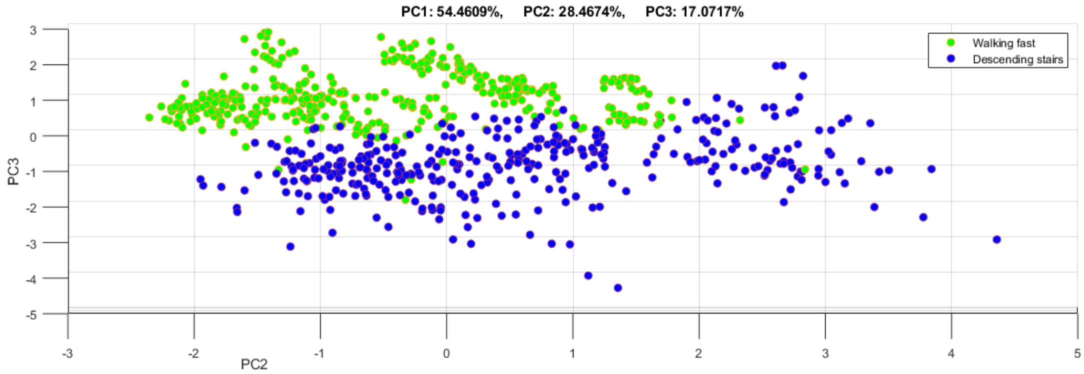


Figure 5.11: Focus on the border between walking fast the descending stairs

tion about the intensity and so the features of this category that will lose relevance.

On the other hand, splitting or merging ”sitting activities” ans ”standing activities” influences the information about the orientation, and so the features that help in discriminating the position of the body. Removing one of the clusters will have relevant consequences for the performances of the second classifier inside the cluster of the activities of interest.

In particular, removing the cluster of “walking slow” will make decrease the accuracy of “ascending stairs”, removing “bicycling” will have an impact on the accuracy of “walking at medium speed” and removing “walking fast” will make decrease the accuracy of “descending stairs”. Therefore, feeding the classifier with a solid training set, with types of activities will build a map of interconnection that will help the classifier in having good perfomances on the new data set. Instead, choosing just a limited number of activities and features will make the classifier lose information about the most critical border and even if it will perfom well on the scripted data it wouldn’t be able of a fine discrimination on the unscripted one. On the other hand, if the classifier has been trained on discriminating these differences on the scripted data it will be able to recognize the pattern of the clusters even with a new data set with unbalanced instances and unbalanced duration of the activities.Specifically, as it is possible to see from the fig. 5.13, the analysis reveals that the labels of the three different speeds of walking are the ones that will help in discriminating the borders inside the activities of interest and between the activities of interest and bicycling, and

Label	Features affected
Split /merge Interest and no interest	Range and complexity
Split / Merge Static and dynamic (different speeds)	Intensity
Split/merge Sitting and standing (different direction)	Direction

Figure 5.12: How different labels affect the features selection step

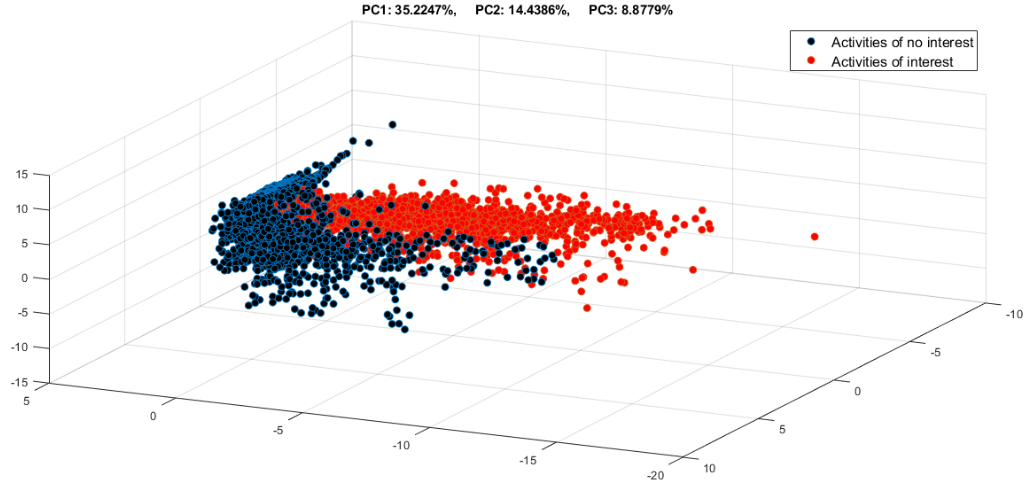
so they will improve the performances of the second classifier. In particulaar the accuracy of ”descending stairs” increased of almost 4%, the sensitivity of 3% and the specificity of 2%. For ”ascending stairs” it was possible to observe the same improvements: 2.5% for the accuracy, almost4% for the sensitivity and almost 2% for the specificity.

2nd classifier	Descending stairs		Ascending stairs	
	1 cluster of walking (speed merged)	3 clusters of walking (speeds split)	1 cluster of walking (speed merged)	3 clusters of walking (speeds split)
Accuracy	94.44%	98.20%	95.36%	97.98%
Sensitivity	81.11%	84.20%	80.36%	84.11%
Specificity	97.32%	99.14%	97.19%	98.98%

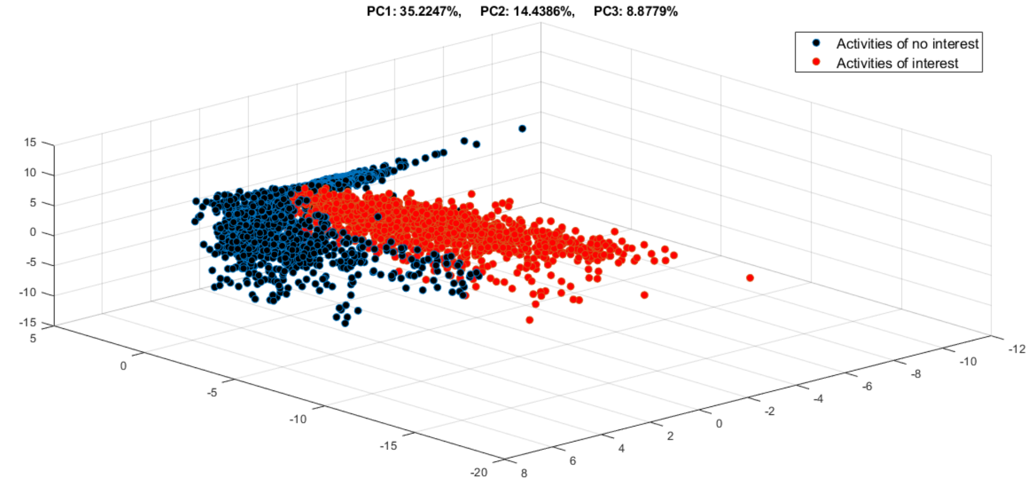
Figure 5.13: How merging or splitting the three clusters of different speeds of walking affect the detection of stair ambulation

5.3 Impact of different methods of crossvalidation

The results on the scripted data showed better rates for the 10 - fold cross validation but this is due to overfitting.In fact, another aspect to take into consideration choosing the method of cross validation is the generalization ability of the model.On the other hand, the leave one subject out method helps finding the common pattern between subjects, removing each time a different subject that can be the outlier. In fact, the former consists in using repeatedly the whole training set but one point, computing many estimators and combining them at the end. This combination should lead to a new estimator whose bias is low.



(a) PCA with 10-fold method

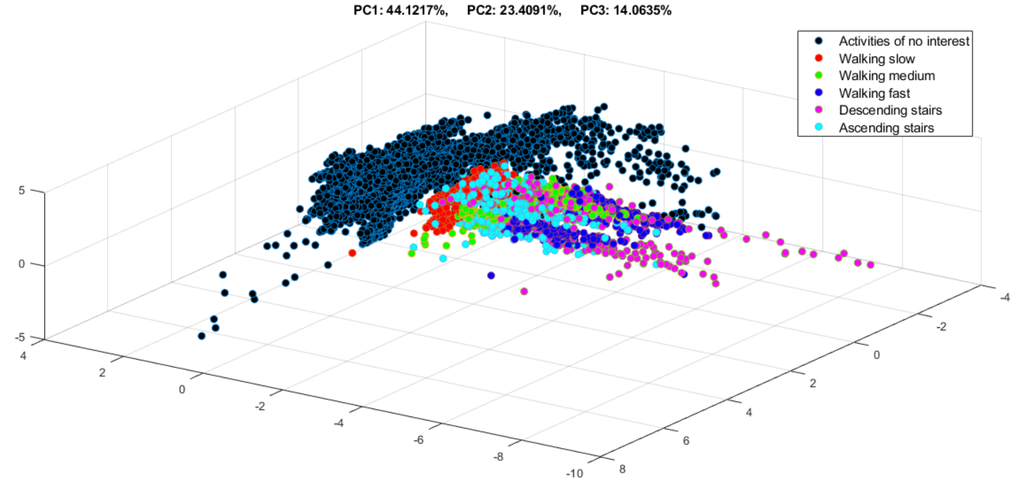


(b) PCA with leave one out method

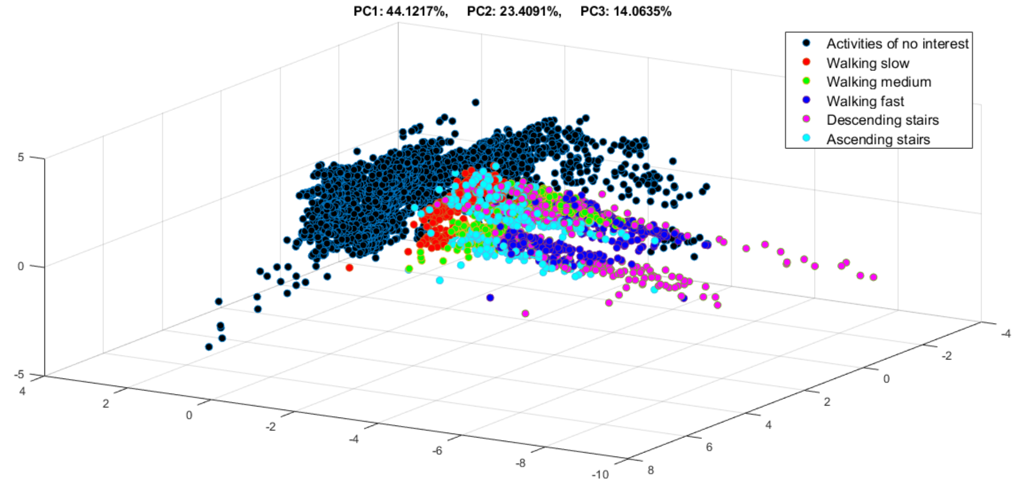
Figure 5.14: First classifier, PCA scatter plot

RATES	Interest	
Accuracy	99.00%	99.44%
Recall	99.50%	98.64%
Specificity	99.87 %	99.85%
Precision	99.76%	99.71%
Fscore	99.5%	99.17%

Figure 5.15: Rates of the first classifier with 10 fold (blue) and leave one(green) method



(a) PCA with 10-fold method



(b) PCA with leave one out method

Figure 5.16: Second classifier, PCA scatter plot

RATES	Walking		Descending stairs		Ascending stairs	
Accuracy	96.03%	96.15%	95.15%	95.18%	94.57%	94.50%
Recall	97.25%	97.05%	84.92%	85.19%	87.47%	87.47%
Specificity	94.18%	94.12%	97.44%	97.47%	96.54%	96.43%
Precision	96.17%	96.10%	88.19%	88.46%	87.47%	87.07%
Fscore	96.70%	96.57%	86.52%	86.79%	87.47%	87.27%

Figure 5.17: Rates of the second classifier with 10 fold (blue) and leave one out(green)method

5.4 The final model

The analysis of the perfomances of the classifier with different configuration of segmentation, labels and method of crossvalidation led to the choice of the best model.
The optimal configuration selected was the following:

- Windows of 5s with no overlap;
- Different labels for the three different speeds of walking;
- leave one subject out as method of cross validation;

The segmentation of 5s and no overalp allows the selection of features of different types and number and from different axes. In particular, for the first classifier the following 9 features 5.1 were selected. They are from both raw and filtered and from all the axes and they contains information about the range of the signal, the the complexity and the abrupt changes.

Table 5.1: Perfomances of the first classifier on the scripted data

Selected Features	**Meaning**	**Axis**	**Type of data**
$MeanRawZ$	Mean acceleration	Anteroposterior	Raw
$MeanFreqFiltZ$	Mean frequency	Anteroposterior	Filtered
$MeanRawX$	Mean frequency	Mediolateral	Raw
$Peak2PeakRawX$	Difference between the max and min of the acceleration	Mediolateral	Raw
$MaxRawZ$	Max acceleration	Anteroposterior	Raw
$MaxRawX$	Max acceleration	Mediolateral	Raw
$Peak2PeakRawZ$	Difference between the max and min of the acceleration	Anteroposterior	Raw
$Peak2PeakRawY$	Difference between the max and min of the acceleration	Vertical	Raw
$MaxRawY$	Max acceleration	Vertical	Raw

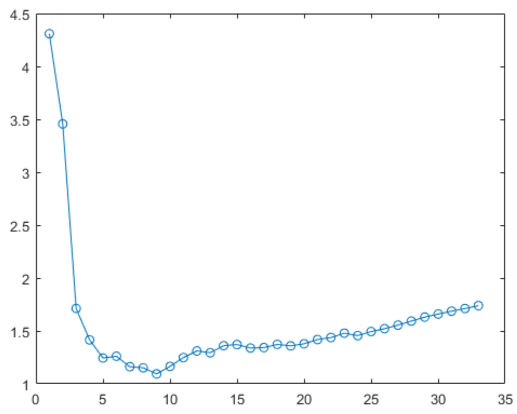
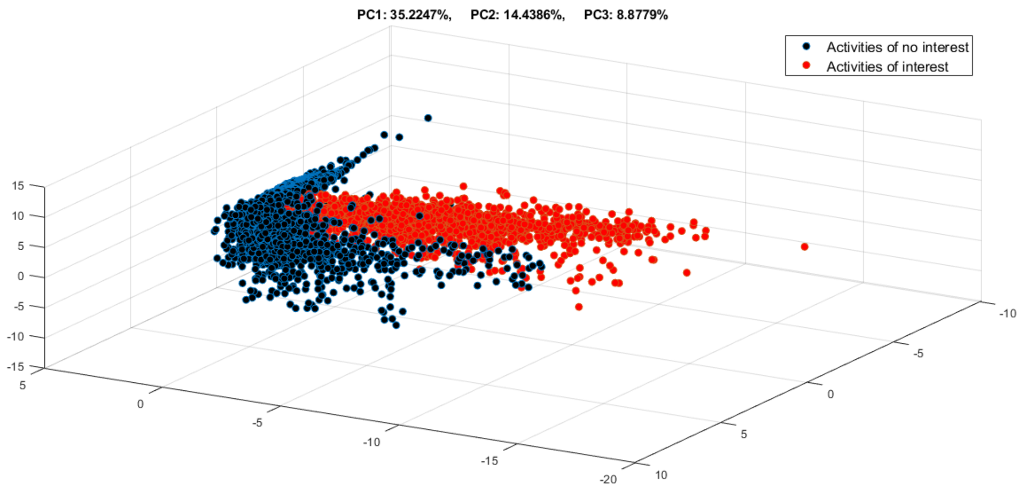
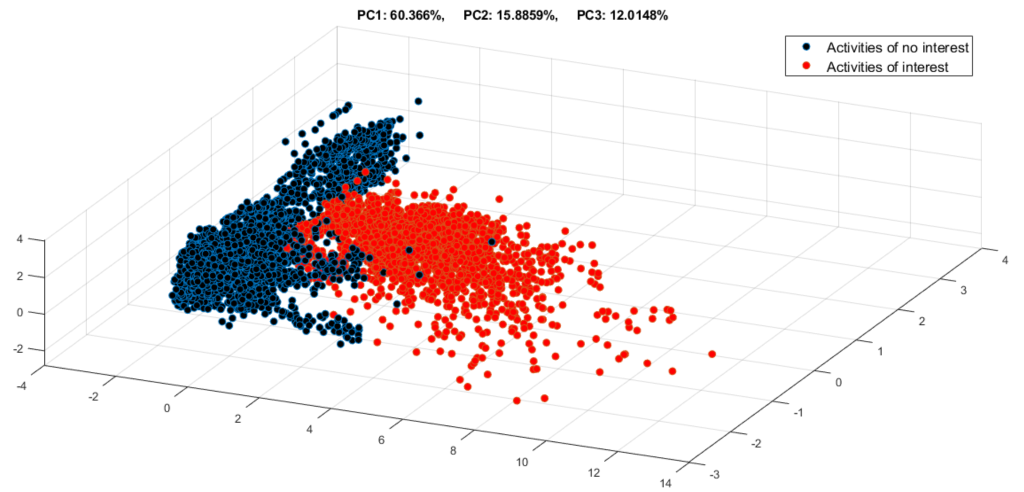


Figure 5.18: DBI 1st classifier



(a) PCA before features selection



(b) PCA after features selection

Figure 5.19: classifier 1, PCA scatter plot

For the 2nd classifier instead the following 8 features were selected as it is showed in the tab. 5.2. They are from both and filtered data and they take into account different aspect of the signal like the mean frequency, the range and the complexity. The result of the features selection on the space of the features it is showed in the fig. **??**. Looking at the scatter plot of the pca in fact it is clear that after the features selection the cluster are more separated and define, so the features selected can well discriminate the activities of interest.

First of all, looking at the figure 5.22and at the tabel 5.3 perfomances of the first classifier of the hierarchical structure will be evaluated.This classifier has the aim of discriminate between the ”activities of interest” and the ”activities of no interest”and it requires and higher accuracy because all the missclassified at this step will be added to the missclassified of the second step.

From the table 5.3 it is possible to observe that the perfomances are charachterized by an high accuracy of 99.42% for the activity of interest. Considering that the two classes have a very different number of instances, for under-

Table 5.2: Perfomances of the first classifier on the scripted data

Selected Features	**Meaning**	**Axis**	**Type of data**
$MeanRawX$	Mean acceleration	Mediolateral	Raw
$MeanRawZ$	Mean acceleration	Anteroposterior	Raw
$MeanRawY$	Mean acceleration	Vertical	Raw
$MeanFreqFiltY$	Mean frequency	Vertical	Filtered
$Peak2PeakRawY$	Difference between the max and min of the acceleration	Vertical	Raw
$VarFiltX$	Variance	Mediolateral	Filtered
$MaxRawY$	Max acceleration	Vertical	Raw
$ObwFiltX$	Occupied bandwidth	Mediolateral	Filtered

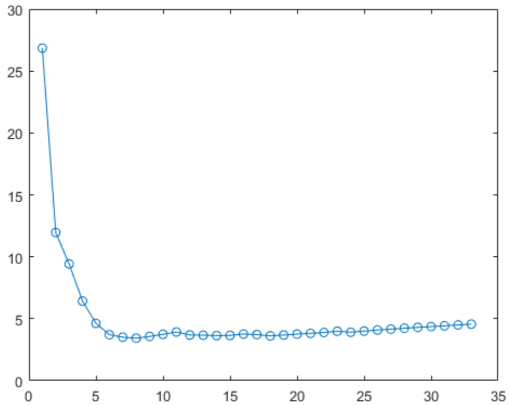
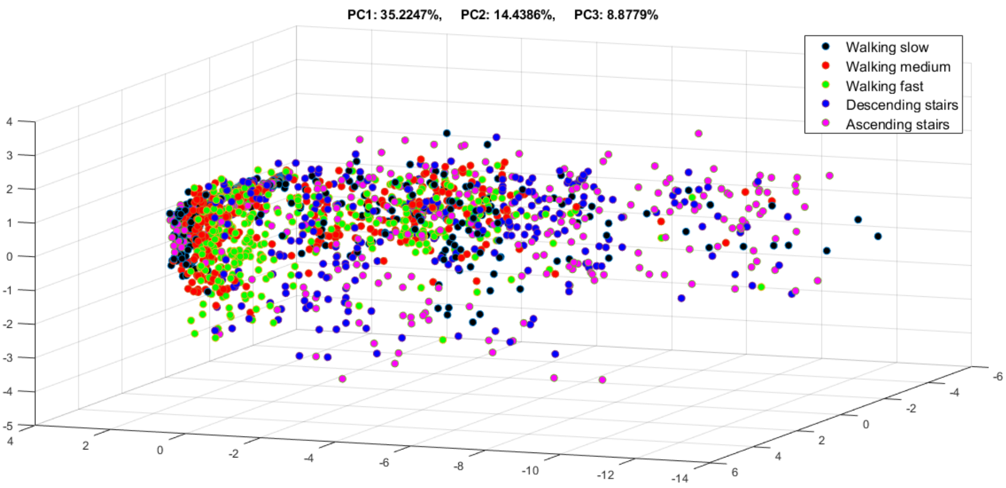


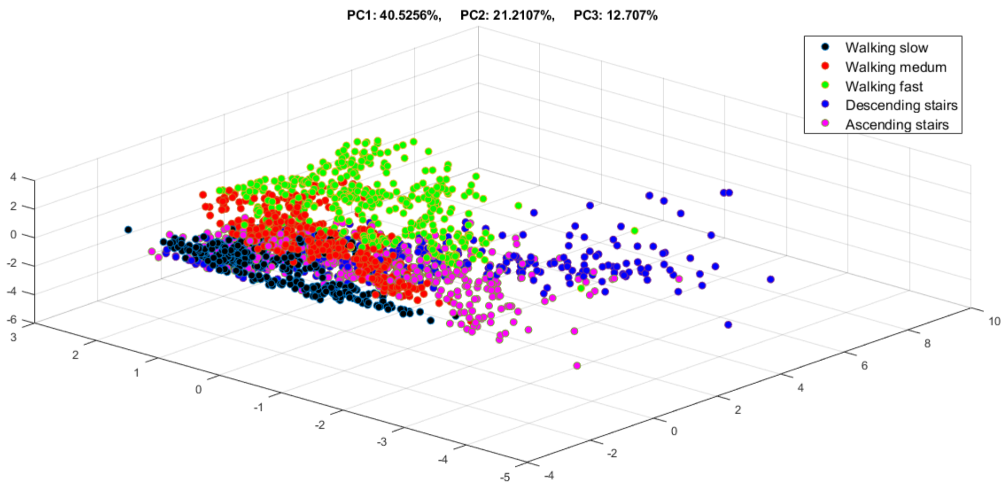
Figure 5.20: DBI 2nd classifier

standing the real perfomances of the classifier also the other rates has to be evaluated. In particular, the classifier is slightly worse in terms of sensitivity (98.51%)and recall (98%) for the class of the " activities of interest" and in terms of specificty fot the class of the "activities of no interest". However all the rates are above 98.5% so the performances are good as expected.

After that, an evaluation of the second step of the hierarchical classifier was made. This classifier has the aim of discriminate between "walking", "descending stairs" adn "ascending stairs". The classification at this step was more difficult because it was required a define and precise discrimination between very similar activities that are also unbalanced in terms of number of instances. In particular, the figure 5.23 will exhibit the performances of the second classifier inside the cluster of the activities of interest. In the confusion matrix are showed all the different speeds of walking as separate classes before the final merge andthe table 5.4 the rates for each class are shown. The result of the tabel 5.4 showed good perfomances for all the different activities and a balance in the number an type of missclassified. In particular, the accuracy reaches the highest value for walking fast (97.7%) but even the lowest values of ascending stairs (94.23%) is above 94%.However, to have a better understanding of the real perfomances of the classifier, also the other rates have to be deeply evaluate. In fact, especially when the number of the instances in each class is different, there will be the possibility that the accuracy of the smallest classes is higher because they are neglected from the model.



(a) PCA before features selection



(b) PCA after features selection

Figure 5.21: classifier 2, PCA scatter plot

In this contenst, looking at the specificity it is possible to observe that it has an overall value of 98.1% for the walking class,97.51% for descending stairs and 96.15% for ascending stairs, so the number of false positive (considering walking as first class) is low and the classifier will not recognize as class walking instances from the other classes. Looking at the sensitivity instead the value slightly decrease for all the classes but the lowest one (85.35%) is still an index of reliable output. The precision (positive predictive value) showed almost the same values with a decreased value for walking and ascending stairs and an increased one for descending stairs.Finally, the recall showed an overall value of 93% for walking an overall value of 86% for stair ambulation.Looking at the missclassified it is possible to see that most of them occur between ”ascending stairs” and ”descending stairs” and between ”ascending stairs” an ”walking slow” as it was expected because they were the most critical borders. The lowest number of missclassified are instead between ”walking slow” and ”descending stairs” and between ”walking medium” and ”descending stairs” because normally the activity of ”descending stairs” is perfomed much faster than the normal speed of walking. From this confusion matrix it is also possible to observe that the number of true

Scripted data , 1st classifier	Predicted class			
Actual class	All the subject s	Activities of no interest	Activities of interest	TOT REAL
	Activities of no interest	3904	5	3909
	Activities of interest	29	1923	1952
*every instance is equivalent to 5 second	TOT PREDICTED	3933	1928	5861

Figure 5.22: Confusion Matrix of the classifier 1

Table 5.3: Perfomances of the first classifier on the scripted data

Rates	**Activities of no interest**	**Activities of no interest**
$Accuracy$	99.42%	99.42%
$Sensitivity$	99.87%	98.51%
$Specificity$	98.51%	99.87%
$Precision$	99.26%	99.74%

positive and true negative is balanced for each class. It is also interesting to look at the missclassification between the different speed of walking because this can be an efficient parameter to measure the ability of the classifier of discriminate between diffent level of energy and so different levels of acceleration.

Looking at the results of both the classifier it is clear that, thanks to the segmentation, the features selected and the method of crossvalidation it learns the main significant variation of the acceleration of different signal. Specifically it learns these variation in terms of different dynamic of the signal in the time domain and frequency components. In fact, the activities of the training set represent different shades in the range of motion and in the complexity of the pattern. Sitting and standing have almost zero acceleration for all the directions, walking and bicycling have the highest acceleration, the signal of descending and ascending stairs has peak inside the vertical and the antero-posterior direction and the structure of the model took into account all these informations. The only information that it can't learn are the random fluctuation of the signal that generate random, inevitable missclassified but all the other type of missclassification are reduced.

After the separated analysis on the two classifiers, in the fig 5.24, it was showed a final confusion matrix comprehensive of both the result of the first classifier and of the second one and so the final result of the hierarchical structure.

Scripted data	Predicted class						
Actual class	All the subjects	Walking slow	Walking medium	Walking fast	Descending stairs	Ascending stairs	TOT REAL
	Walking slow	339	13	0	3	14	369
	Walking medium	7	366	11	3	4	391
	Walking fast	0	16	368	4	3	391
	Descending stairs	9	2	4	303	37	355
	Ascending stairs	17	4	3	29	364	417
*every instance is equivalent to 5 second	TOT PREDICTED	372	401	386	342	422	1923

Figure 5.23: Confusion Matrix of the classifier 2 with the different speeds of walking

Table 5.4: Perfomances of the second classifier on the scripted data

Rates	Walking slow	Walking medium	Walking fast	Descending stairs	Ascending stairs
Accuracy	96.72%	96.88%	97.87%	95.27%	94.23%
Sensitivity	91.87%	93.61%	94.12%	85.35%	87.29%
Specificity	97.88%	97.72%	98.83%	97.51%	96.15%
Precision	91.13%	91.27%	95.34%	88.60%	86.26%

Scripted data, all the subject s	Predicted class					
Actual class	All the subject	Activities of no interest	Walking	Descending stairs	Ascending stairs	TOT REAL
	Activities of no interest	3904	3	1	2	3910
	Walking	15	1120	13	20	1168
	Descending stairs	6	17	304	34	361
	Ascending stairs	7	24	33	339	403
*every instance is equivalent to 5 second	TOT PREDICTED	3932	1164	351	395	5842

Figure 5.24: Final confusion matrix of the scripted data

Table 5.5: Perfomances of the classifier on the scripted data

Rates	Activities of no interest	Walking	Descending stairs	Ascending stairs
Accuracy	99.43%	98.4%	98.2%	97.98%
Sensitivity	99.80%	95.80%	84.20%	84.11%
Specificity	98.50%	99.06%	99.14%	98.98%
Precision	99.84%	95.89%	84.21%	85.82%

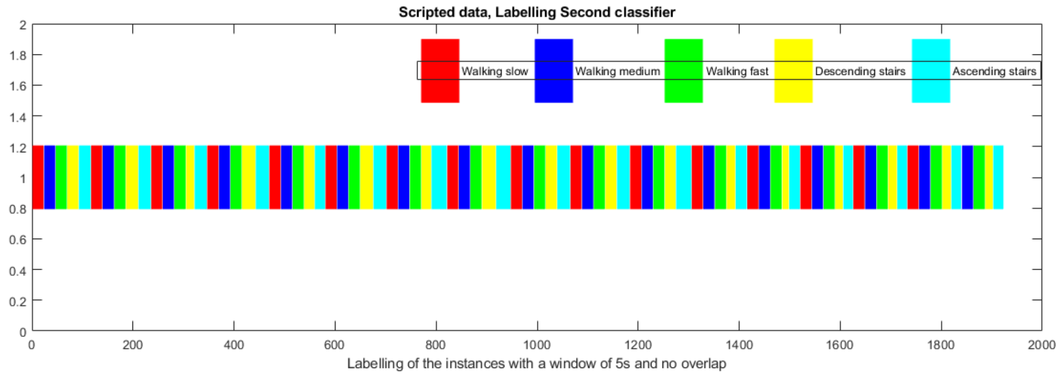


Figure 5.25: Scripted data, Labelling of the activities of interest

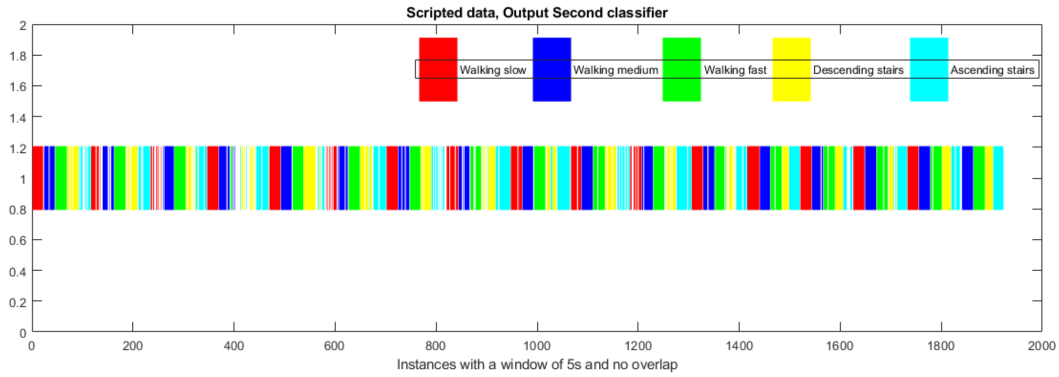


Figure 5.26: Scripted data, Output of the second classifier

The classification's results underlined that the most frequent missclassification and so the most dangerous for the final purpose of the algorithm are the one between the two types of stair ambulation and the one between ”walking” and ”ascending stairs”.
Therefore, these two missclassifications were plotted with the aim of seeing the distribution of them along the signal because a random distribution will minimize the effect of the false negatives and false positives.
In the fig.5.27 the ”descending stairs”samples classified as ”ascending stairs” and the ”ascending stairs” samples classified as ”descending stairs” are showed.In the fig.5.28 instead the ”walking”samples classified as ”ascending stairs” and the ”ascending stairs” samples classified as ”walking” are showed.

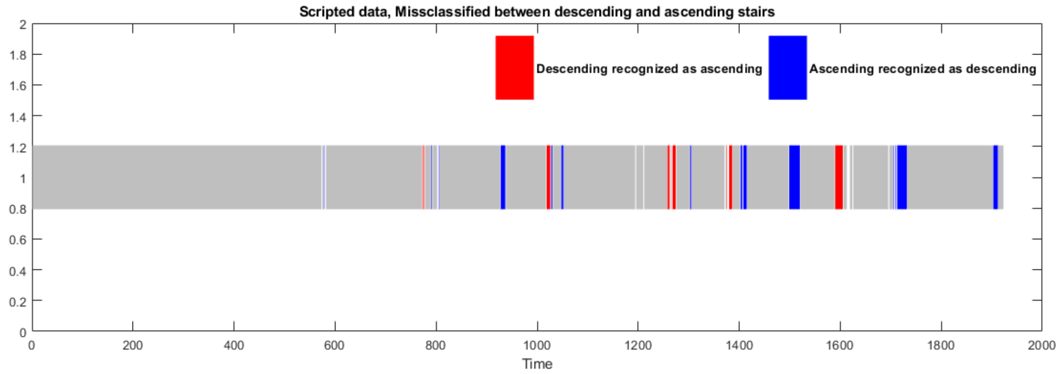


Figure 5.27: Scripted data, missclassified between descending and ascending stairs

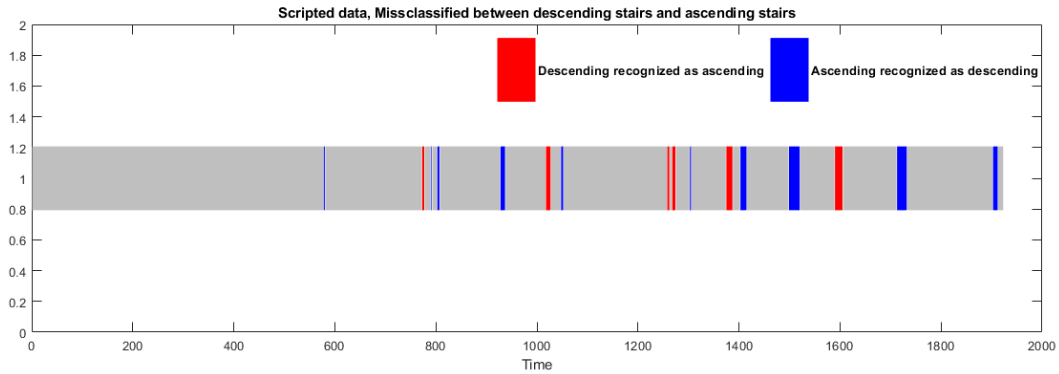


Figure 5.28: Scripted data, missclassified between walking and ascending stairs

Chapter 6

Discussion

The results support the hypothesis of this study that the choices of the segmentation, the label and the method of cross validation have a significant impact on the features selection and on the performances of the classifier. Moreover, this model will guarantee detailed and, at the same time, a consistent information, following the variation of the activities but also giving a stable output.
This goal was reached with an exploratory study that led to the choice of the optimal parameters for this particular application:

- the window for the segmentation of the signal in terms of width and overlap;
- the labels for the instances;
- the method of cross-validation;

These choices have an impact on the performances of the classifier because they affect these three different aspects:

- the number of instances;
- the correlation between instances;
- the information carried by every single instance;

In particular, the number of instances and the correlation between them affect the stability of the output, meaning the possibility of identify a consistent information of the activity that the patient is doing. In fact, if the algorithm is too sensible to every little variation it will no give a stable output.
The information carried by a single instance has an influence instead on the ability of following the fluctuation of the signal and so the algorithm power in detecting the start and the end of every activity. These choices were made in order to build a model with the following characteristics:

- Informative: it has the ability of detect even the shortest activity and it is precise in the recognition of the start and the end of the activities;
- Consistent: it gives a consistent ouput without being to sensible to the variation of the signal;
- A random distribution of false positive and false negative: a random distribution will still allow to recognize the activity,a focus one instead will lead to miss relevant information;

- Flexibile: it overcome the inter-subject variability guaranteeing almost the same accuracy with subjects with diffent ways of doing the same activity and the intra-subject variability guaranteeing almost the same accuracy with different daily routine;

Each one of these aspects will be very important for the application on patients. In fact, every patient has different characteristics and, during the theraphy, even the way the same patient performs the same activity changes due for example to the fatigue or the progression of the disease.
The first part of the project was limited inside the controlled environment of the laboratory in which there was less variability among the data of the same activities considering that they were recorded following a specific protocol with instructed activities.The main risk of this part of the project could have been the overfitting of the training set that would have led subsequently to a bad recognition on a future new data set.
Another difficulty was the use of only one accelerometer that requires to extract very specific features based on a very precise analysis of the different signals.
The rates of the confusion matrix achieved with the data collected inside the lab showed the potential of the detection algorithm with a hierarchical structure. In particular, the results proved that the hierarchical model is better than a flat one.The main reason behind this result is that in the case of the hierarchical model a simpler discrimination it is performed between "activities of interest" and "activities of no interest" and then more defined cluster are constructed around the classes of "walking" and "stair ambulation".
On the contrary, with a flat structure, the fine information that will help discriminating inside the cluster of the "activities of interest" would have been lost and a lot of useless information around the other clusters would have been given as an input. Moreover, this kind of structure would allow to use again the classifier for the detection of other types of activities for future studies. Furthermore, the choice of the random forest algorithm makes the classifier identity independent, knocking off the rough edges of the clusters of the decision trees that tunes the detection on the subject specific executions.
Another key point during the analysis was to discover the type and distribution of misclassified instances because they are responsible of the definition of the border between activities. Another important choice was to reduce these most frequent misclassifications. The solution adopted was to maintain the label of different speeds that helps in the reduction of the misclassified between "walking" and other activities along the borders of this cluster. In particular, the crucial borders that this ploy helps were the one between "walking medium" and "cycling", "walking slow" and "ascending stairs" and "walking fast" and "descending stairs". In fact, training the classifier with different speeds helped improving the sensibility of the output of the accelerometers which is very similar for these tree couples of signals, especially during the transition from one step to the other.
The performances on the unscripted data will decrease compare to the ones on the scripted one. Part of the reason why this happen is the distribution of the classes in free-living environment. In fact, the data set for training the model presents an equal distribution of the samples, instead in the daily-life, considering that the subjects are in the work environment, most of the samples owns to "sitting". Different observations will be made on the scripted and on the unscripted data taking into account the different environment in which these data were collected. In particular, it was preferred a solution that would have given more coherent results on the unscripted data even if it was generating slightly worst results on the scripted data in terms of accuracy and precision.
Train the model also on daily-life data could improve the accuracy but only if the classifier will be apply only on

a different type of patients the problem will come again so the choices adopted in this work were made with the aim of realize the most user independent algorithm.Some misclassification in fact could happen in the free-living environment and these false negatives and false positives will be different depending on the life style of the patients.For seeing effective potential and limitations of this work the results of the present study have to be tested on different data set and on a larger population.

6.1 Choice of the segmentation

The results showed significant changes of the performances with different segmentation even if the classifier was robust enough to contain them. Specifically, the width of the window influenced the number of instances and also the number of different activities contained inside a single instance. The overlap instead affected the number of the activities and the correlation between instances.

The evaluation of the performances of the different windows was done also looking forward at the possible results on the unscripted data. In fact, even if with a fixed length of activities there is no significant difference between the perfomances of the 10s and the 5s window this gap will increase in a real world environment. The window of 10s indeed is often longer than the duration of the activity and so it will include signals of the other activities or of the resting periods. Therefore, the calculation of the features would contain an error due to the average of two different signals and this could possibly lead to misclassification, in particular to false negatives. Therefore, from this perspective, a shorter window could be more appropriate for calculate parameters related to a specific activity with the aim of reducing the false negatives. However, make the window too much short like the one of 2.5s could lead to the opposite risk of too many false positives because for example the transition activities that are normally very short and that are not relevant for this study could be misinterpreted as actual activities.

Furthermore, looking just into a short frame of the gait, two activities that are different can look similar. For example, ”ascending stairs” and ”descending stairs” both present a peak of acceleration when the foot is touching the next step of the stairs but this peak happens in two different times of the gate. However, if the window is too short, the features are extracted from the peak and there will be an high missclassification between these two classe.

Therefore, even if the window of 2.5s is more precise in the detection of the start and of the end of the activities and of the shortest pieces of them, it has to be discarded because is too short to realize the media that will take into account all the information of the tipical pattern of the activity to detect. In fact, the results of the application of this window reveal an overall accuracy of 92.8%, so almost 6% lower than the other two windows and almost 10% lower sensitivity and specificity for stair ambulation in respect to the windows of 5s and 10s.

It is possible to see these differences between different segmentation looking at the number and at the types of feature selected by each window. In fact, the window of 10s selected features related to a general information about the signal, the one of 2.5 s features related to abrupt changes. Both of these window have the limitation of underling just a particular aspect of the signals, losing information that will help in a more define discrimination.

The choice between a shorter or a longer window is even more complicated because a shorter window implies a longer time of processing and this will a problem because the identification will be on data recorded on an entire

day.
In this contest is the window of 5 second the one that offer a good compromise in terms of:

- Type of features selected;
- Number of features selected;
- Detection power;
- Stability of the output;
- Processing time;

Moreover, for the final purpose of this work, which is the clinical evaluation of the changes in the pattern of the knee flexion and extension, only pieces of signal with the minimum length of 5s are relevant for the monitoring so this window will be also the best fit from a clinical perspective.Looking at the results the two window of 5s and 10 s are comparable but the impact of the window of 5 s will become greater on the unscripted data.
On the data recorded during the data collection the performances of the window of 10 s are better for example for the sensitivity of the stair ambulation because this segmentation is applied on instructed activities with fix length and so a high resolution is not required but is preferable instead consistence. However, when it comes to a free-living situation this window lead to a higher misclassification because it will miss instances at the start and at the end of the activity and the shortest activity. Moreover, in a free condition the subject is often doing activities like ”walking” for a shorter period for going to grab, for example, an object in a corner of the office. Therefore the results showed that the best compromise is a width of 5s that guarantee good performance on the scripted data set but also on the future unscripted data. This window achieves this goal for different peculiarities that will make the detection mechanism:

- Flexibile: in the step of features selection it guarantees a balanced number and variety of types of features;
- Informative: the window will be able to detect also short pieces of signal and activities with a shorter gate (higher speed);
- Consistent: the window is long enough for being able to characterize the activity and to do the media for delete the contribute of transition activities and abrupt changes;

The choice of the overlap was important as well and became very critical for the discrimination between ascending and descending stairs. In fact, the pattern of these two activities is very different if the signal is segmented in a way that every window will include the exact gait of the activity, otherwise the two can be mix up.
After the analysis of the results, it was clear that change the length or the overlap of the window affects the detection of every activity and in particular that every window is the best for improve the accuracy of one activity.This variability doesn’t let one window to be the perfect fit but the one with 5s width and no overlap is the one more adaptable to changes and so the one preferable for this particular study.

6.2 Choice of the label

The systematic exploration of how the boundaries of the clusters move depending on the different labels is another key point of the analysis.In fact, the choice of the label is highly correlated with the final shape of the cluster and so with the performances of the classifier.In particular, the label influenced the DBI index because adopting a lot of labels lead to choose a lower number of features because of the multiples relations of similarities and differences between activities that constrain the distances between them.

First of all, the analysis of the relation between “dynamic activities “and “static activities” and between “standing activities and “sitting activities showed the overlap between these cluster.For the discrimination of these clusters also the basic set of features was enough because separation between them depend on the the similarties and the differences between these activities that connect or divide them and for these activities these characteristics are well defined.

The hard issue is instead the definition of the boundaries inside these cluster when it comes to very simmilar activities. For avoiding the risk of a concentration of missclassified along these borders, the addition of more label, and more features are required. For example, as it is possible to see from the result it is very simple to discriminate the cluster between ”walking” and ”sitting” because these activities are different both for energy required and orientation of the body.

Defining the borders between cycling and walking instead is more complicated because they have almost the same energy and the same orientation. However, the most difficult definition of the borders is inside the clusters of the activities of interest because of the multiple similarities in the pattern of ” walking”, ”ascending stairs” and ”descending stairs”. So, in the first step of the hierarchical classifier are selected different features from the one selected for the second discrimination inside the cluster of the activities of interest.It is clear again the benefit of this structure that allows to tailor the needs of the first and of the second classifier for the two different type of discrimination. Instead, with a flat model the features selected woul have been a compromise between the goal of define the cluster of the activities of interest and meanwhile define the borders inside this cluster.

The exploratory analysis of giving different labels to the training set and looking at the change of the accuracy reveals how the type of instances have an impact on the accuracy of the classifier. In fact, the classifier needed to be trained to recognize the most difficult borders between very similar activities. These border were the most difficult to define and the analysis of their movements led to the choice of the best option of label for a “shape” simpler to “cut” for the random forest algorithm. In particular, the analysis reveals how leaving separately the three labels of the different speeds of walking improved the performance of the classifier for two reason.

The first one is related to the problem of an unbalanced data set that affects the performances of the 2nd classifier because the instances of walking are three times bigger than the instances of ”ascending stairs” and ”descending stairs”. The second reason is the similarity in the orientation of the body between ”walking slow” and “ascending stairs” and between ”walking fast” and ”ascending stairs”. Merging all the three labels together led instead to loss of information for the classifier and so to a greater amount of misclassified along these borders.

6.3 Choice of the method of crossvalidation

The comparison between the two methods was conducted on different levels.First of all, the results were evaluated looking at the overall accuracy, specificity, precision, recall and Fscore that are comparable for the two methods.After that, it was evaluated the possibility of overfitting and for this reason leave one out was preferable because with 10 cross-fold validation and this data set the small cross-validation group would have led to overfitting and so to low performances on the unscripted dataset.

Leave-one-out cross-validation is approximately unbiased, because the difference in size between the training set used in each fold and the entire dataset is only a single pattern. However, while leave-one-out cross-validation is approximately unbiased, it tends to have a high variance (i.e. very different estimates if the estimated is repeated with different initial samples of data from the same distribution). As the error of the estimator is a combination of bias and variance, whether leave-one-out cross-validation is better than 10-fold cross-validation depends on both quantities.

The variance in fitting the model tends to be higher if it is fitted to a small dataset (it is more sensitive to any noise/sampling artefacts in the particular training sample used). This means that 10-fold cross-validation is likely to have both higher variance and bias if the amount of data is limited, as the size of the training set will be smaller than for leave one out. So in case the size of the training set is small, leave one out appears to be most appropriate.Therefore the method of cross validation selected was leave out.

6.4 Final Model

The results on the final model showed that the classifier is robust and overcomes the issue of the most critical borders designing well defined borders around the clusters of stair ambulation and ground walking.

In fact, the probability that the random forest learns the noise is lower than other algorithms with a large data set because the bagging approach, thanks to ensemble of different models, contains this risk.

In particular the results showed a well-defined discrimination between the cluster of "static activities" and "dy-

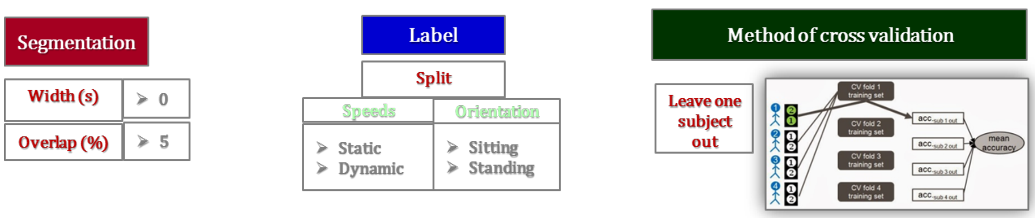


Figure 6.1: The final choices for the model

namic activities" and between "sitting activities" and "standing activities".

Moreover they reveal also the most dangerous borders:

- between "walking medium" and "cycling";
- between "ascending stairs" and "descending stairs";

- between ”walking slow” and ”ascending stairs”;

- between ”walking fast” and ”descending stairs”;

Specific types of features related to the abrupt changes of the signals and to the complexity of the pattern were selected for the most difficult discriminations. In particular, the discrimination of walking slow and ascending stairs and the one between ascending stairs and descending stairs was possible thanks to features to detect the spikes of the signals.
The plots of the DBI showed an efficient number of features for the final model: 9 more general features for the first classifier and 8 more detailed features for the second one. The reliefF ranked as most important features, features of al the types so they will allow the definition of the borders between activities very similar, like walking and stair ambulation, and very different like walking and sitting. The final choices adopted (fig.6.1) for the model made it flexible and adaptable for a new data lager and different data set.The perfomances of the hierarchical classifier were higher compared to the previous studies with the almost the same set of activities.
Starting by considering the first classifier it is possible to observe that the discrimination between the activities of interest and the activities of no interest led to an overall accuracy of 99.42%, a precision in the recognition of the activities of interest of 98.51% and a 1% higher specificity and precision. These results were expected because it was preferred to adopt choices especially with the aim of reducing the false positives.
The performances of the second classifier were slightly lower but this was expected to because the discrimination inside the cluster of the activities of interest is more difficult due to the multiple overlap in the space of the features.The activities of interest are indeed similar for level of energy, peaks of acceleration and the variability among the features extracted from all this instances produced a bias that create noise in the observation of the differences between different activities. Despite that, the classifier guarantee a robust, precise and accurate discrimination also inside the cluster of the activities of interest.
In particular, the overall accuracy in the detection of “ground walking” was 97.1 %, with the highest value for the activity of walking fast. The overall accuracy in the recognition of stair ambulation is lower (94.75%) but it is still above 94%, specifically is 95.27% for “descending stairs” and 94.23% for “ascending stairs”. Moreover, also for the second classifier the values of the specificity are higher than the sensitivity and of the precision. There are a lot of factors could have limited the performances of the classifier.First of all, extracting the same features from all the activities for different type of discriminations and segmenting all the signals with the same window. In fact, segmenting in different ways the data from different recorded activities and choose the best features for each recognition will have been the perfect choice for the scripted data but looking forward at the future application on the unscripted data, where there is no label, and it isn’t know a priori which instances belong to a specific class, this is the best choice.
Another issue has been the choice between doing the identification in one step or in two with a hierarchical structure. In fact the first solution put higher the accuracy and the specificity but lower the sensitivity because separating all together all the clusters will make the dimension of the distance inside the cluster of the activities of interest reduced. In fact, the presence of the other clusters makes the cluster of stair ambulation and ground walking more compact.
Another issue was the unbalanced data set between walking and stair ambulation that was solved thanks to the

solution of mantaining separately the label of different speeds of walking. Moreover, most of the study of activity recognition focus the attention only on the scripted data for selecting the optimal parameters and so the model isn't flexible for the application on a new data set.The choice for this algorithm instead were made also looking on the unscripted data. Therefore, the results reveal, with a good recognition of walking and stair ambulation among the 13 activities performed in the lab, the feasibility of the application.

Informativeness and stability of the model: ability of the detecting

The informativeness is the ability of detecting of the model.Looking at the plot of the output of the classifier and comparing it to the actual labelling of the activities it is possible to establish whether the classifier is missing relevant information or not.
On the other hand, in the same plots it is possible to evaluate the stability of the output, that is the ability of giving a consistence output. In particular, the parameter of the stability of the output, measured looking at the variation between an istance and the following five one, reveal that for each subject the classifier gives a strong and stable output.

Flexibility: ability of handling the inter-subject and intra-subject variability

Looking at the results it is possible to observe that there is a balanced distribution of missclassified between subjects. This will have a relevant impact on the unscripted data set when the subjects will have a slight variation in their daily routine even if they are similar because they are sitting most of the time, walking for going to the other offices or to the lunch room, doing the stairs for going to other floors. However each one of them can decide whether take the elevator or not, how many times go to the bathroom, how many times go to another office to discuss with a colleague.Therefore, the data set of each subject will be different in terms of the time spent in the activities of interest and of the time spent doing walking and stair ambulation.
The aim will be that the overall accuracy of the classifier on the subjects will not be very influenced by these differences. Moreover, there wouldn't be a better rate of recognition in a specific part of the day: even if in the morning the subject is less tired than in the evening this aspect doesn't affect the perfomances of the algorithm. So the model will overcome both the issue of the inter-subject variability and of the intra-subject variability.

Distribution of false positives and false negatives

The main problems in this detection are: identify as interest pieces of signal that are actually a no activity (false positive) or don't identify pieces of signal that are actually walking or stair ambulation (false negative).
The first missclassification could lead to the identification of a change in the pattern when it's actually another activity, the second one instead could make miss these changes. So the choices were made for reducing at the minimum the number of false positives and false negatives but also trying to have a homogeneous distribution of them. In fact, with a homogenous distribution of them it will be still possible to infer something about the daily changes. The false positives and false negatives in fact have not to be systematically due to a wrong structure of

the model but to random fluctuations of the signals.

It is also important to notice that the results on the first classifier were evaluated looking at the possible further performances on the unscripted data of healthy control subjects and unhealthy subjects in a less controlled environment.

Looking at the missclassified between ascending stairs and descending stairs an between walking and ascending stairs, it is possible to observe that the missclassified are distributed uniformily along the day and their number is almost the same.This kind of distribution allows the identification of the activities of interest because even if inside a consistent piece of walking or stair ambulation there are some random missclassified it will be still possible to detect the presence of the main acitivity.A concentration of missclassified instead will have led to the identification of that group of missclassified as an actual consistent acitivity and so this will have led to the loss not only of some instances but of an entire piece of signal.

On the other hand, also the number of missclassified plays an essential role because a greater number of false negative will have implied a loss of information and a greater number of false positive will have included into the clinical evauation pieces of signal that were actually from another activities. In particular the damage of this last type of missclassification is the worse because loosing some instances along an entire day is not so relevant than have a wrong information about the condition of the patient due to false positive.The model selected avoided both this risks with a low number of missclassified instances and a uniform distribution of them.

Chapter 7

Conclusions

It has been shown that this work reached his goals of detecting ground walking and stair ambulation and so it brings significant improvements to the current state of the art.

In fact, the results support the aim of the study: to have a tool for analyze the daily changes in the pattern of flexion and extension in the knee kinematics and so let the cliniciand build a personalized rehabiliation, specifically tailored for each patient.

Therefore, the Kromm has to detect the ground and grade walking among free-living activities and so, in particular, the algorithm has to construct flexible and defined decision boundaries around the walking and stair ambulation clusters.

For achieving this aim an exploratory study was done looking at the impact that the segmentation, the labels and the method of crossvalidation have on the features selection and on the perfomances of the classifier both on scripted data (data of instructed activites of the same duration perfomed by 17th subjects inside the lab) and on the unscripted data (data recorded during the daily routine at work of 4 of these 17 subjects).

The results of the exploratory study led to the choice of a window of 5 seconds, to a different labels for the three different speed of walking, to a hierachical structure using the random forest algorithm with 100 trees and to the leave one out method for the crossvalidation.

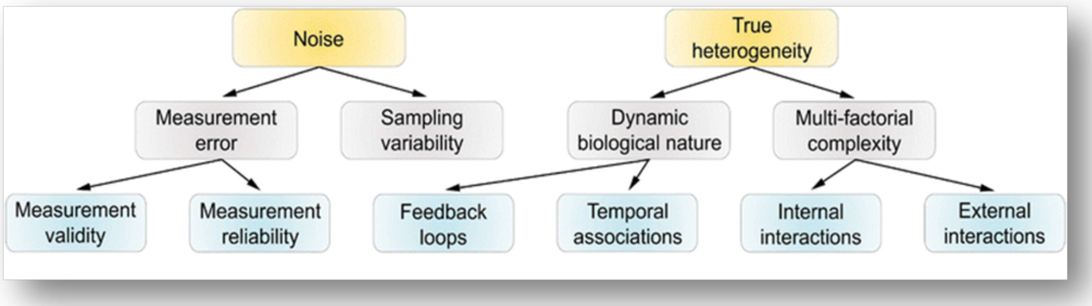


Figure 7.1: Principal factors of error of the classifier

The choiche on the scripted data were adopted looking forward to the future application on the unscripted data. More specifically, the aim was reaching good performances avoiding the overfitting which would have led to a rigid model with worse perfomances on the unscripted data.
Therefore, thank to these choices,the machine learning method will have a reliable output also on a complex and heterogeneous environment intra and inter subjects variability.
The selected model offer the flexibility, and overall accuracy above 97% for all the activities. Its output is informative because it has the ability of detecting the start and of the end of the activities and it is also consistent without being too sensible to the random fluctuations of the signal, which is one of the main problems of a correct classification.
This noise creates both measurement errors and sampling variability and consequentially a random distribution of false positives and false negatives which is inevitable but the fact that this distribution is random still allow to recognize the activities of interest.
The other issue is due to the heterogenity of the space of the features, this complexity and diversification could have led to an incorrect construction of the model an so to a focused missclassification with the consequence of missing relevant information of a specific part of the signal.
The model selected fixed the seconf types of error and limited the first one. Moreover, it allowed a significant reduction of the false positives that are the most critical type of missclassification for clinical purpose.
Two main issue that need still to be solved are the potential break of the device, especially during the night and the problem of synchronization between the app and the device.
However, at this step the goals of develop an accurate and robust classification classifier for activities detection in lab environment and free-living conditions is achieved.
The differences between this work and most of the studies about activity recognition is the focus on a particular subset of activities ”of interest” specifically selected for this target of patients instead of a broad paronama of activities.
The activities of the training set are collected not for a further recognition of them but because the different information hidden inside each one of them is essential for defining the borders around the cluster of the activities of interest. Different and similar speeds, ranges, directions are the connections inside a complex and heterogeneus map that can changes his shape during the day and between patients but that can remember the original proportions and so adapt them to the new data set.

Chapter 8

Future works

A lot of challenges are open for the future but the main goal will be continuing with the study on patients with knee OA in the home environment for realizing a personalized rehabilitation.
For this target of subjects, it is expected that the performances of the classifier will be worse because of the increased variability intra-subjects due to the progression of the disease and to the freedom of the home environment and the increased variability inter-subject due to the different types of patients and to the different ways they will perform the tasks.
For this purpose, the robustness of the detection mechanism could be improved following different ways:

- the implementation of a semi-supervised model with a training on specific data of the patients;
- a new data collection for insert inside the system other types of activities;
- a new data collection with OA patients in the home environment;
- the implementation of an Hidden markov Model for taking into account the probability of transition among states

The hierarchical model will be the perfect fit for future implementations because this particular structure will allow simpler modifications for adapting the algorithm to the new model, especially in the features selection step.
Another key aspect that will be fondamental for future applications is the flexibility of the selected model that will allow good perfomances also on a new heterogeneus data set. Moreover, the daily routine of these new subjects is similar to the one of the subjects inside the lab or even more sedentary for people with the last stage of this disease so the classifier will more likely perform well also on the new data set.
Furthermore, another important improvement could be conduct an analysis to detect postural transitions because they provide valuable information on the functional ability of the patient and on the other hand correct the identification of small movements like slight readjustment of posture while sitting.
In the home setting there will be also the need of a more robust structure of the device and a more comfortable one because the optimal output of the classifier it is related to a correct fit of the sleeve and because the breaking of the device will make lose significant data.
Finally, the implementation of a finite state machine based on the score systems presented in this work could be

implemented for calculate the probability of transition from one activity to the other witha more quantitative and systematic approach.In fact, in the home environment, there will not be a reliable output like the one given by the camera and so the perfomances of the classifier will have to be evaluated using not an external output as label but inferring something inside the signal itself.For this purpose, the exploratory study done in this work will be a solid background for a deep understanding of the reliability of the predicted output.

However, looking at the results on the semi-unsupervised environment of the lab, it is possible to assest that the model built in this work will identify the activities of interest among the free living activities with a small random error due to the fluctuation of the signal and that it will recognize the movements made by the patients even with their differences and even if they perform these activities for a short period of time.

Nowadays a pilot study on the home environment is starting with the aim of collecting more data from patients, all the million possibilities offered by the Kromm are ready to be discovered.

Bibliography

[1] A. Li and D. Meyre. Jumping on the train of personalized medicine: A primer for nongeneticist clinicians: Part 3. clinical applications in the personalized medicine area. *Current Psychiatry Reviews*, 2014.

[2] http://www.lumir.info/new/olink-proteomics.awp.

[3] https://en.wikipedia.org/wiki/Personalized_medicine.

[4] W. K. Redekop and D. Mladsi. The faces of personalized medicine: A framework for understanding its meaning and scope. *Science direct*, 2013.

[5] https://www.cancer.gov/publications/dictionaries/cancer-terms/def/personalized-medicine.

[6] F. R. Vogenberg, C. I. Barash, and M. Pursel. Personalized medicine, part 1: Evolution and development into theranostics. *Pharmacy and therapeutics*, 2010.

[7] http://blogs.nature.com/genome_istock_thinkstock.jpg.

[8] A. K. et al. A personalized, intense physical rehabilitation program improves walking in people with multiple sclerosis presenting with different levels of disability: a retrospective cohort. *BMC Neurology*, 2015.

[9] https://http://walk-again.de/armeo-spring-armtrainer/#lightbox[group-1219]/2/.

[10] X. F. Michelle J Johnson, L. M. Johnson, and J. M. Winters. Potential of a suite of robot/computer-assisted motivating systems for personalized, home-based, stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 2007.

[11] M. J. et al. Potential of a suite of robot/computer-assisted motivating systems for personalized, home-based, stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 2007.

[12] S. L. W. et al. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke. *JAMA*, 2006.

[13] K. J. et al. Eular recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a task force of the standing committee for international clinical studies including therapeutic trials. *Ann Rheum Dis*, 2003.

[14] https://www.123rf.com/photo_stage-of-osteoarthrit.html.

[15] http://www.punarjanis.com/asset/osteoarthritis-2.jpg, .

[16] J. P. et al. Managing knee osteoarthritis: The effects of body weight supported physical activity on joint pain, function, and thigh muscle strength. *BMC Neurology*, 2014.

[17] S. P. Messier, D. J. Gutekunst, C. Daviv, and P. DeVita. Weight loss reduces knee-joint loads in overweight andobese older adults with knee osteoarthritis. *ARTHRITIS and RHEUMATISM*, 2015.

[18] S. Biswal, B. Medhi, and P. Pandhi. Longterm efficacy of topical nonsteroidal antiinflammatory drugs in knee osteoarthritis: Metaanalysis of randomized placebo controlled clinical trials. *The Journal of Rheumatology*, 2004.

[19] E. R. et al. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee. *Rheumatology*, 2004.

[20] D. K. R. et al. A mechanical hypothesis for the effectiveness of knee bracing for medial compartment knee osteoarthritis. *US National Library of Medicine*, 2007.

[21] A. K. et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine*, 2008.

[22] http://www.drugsclaim.com/blog/wp-content/uploads/2016/08/total-knee-replacement-surgery-methods.jpg, .

[23] https://www.pthealth.ca/app/uploads/2017/07/Osteoarthritis.jpg.

[24] A. J. B. et al. Increased knee joint loads during walking are present in subjects with knee osteoarthritis. *Elsevier Science*, 2002.

[25] I. H. et al. Effects of knee orthoses on walking capacity and biomechanics in patients with knee osteoarthritis. *Journal of NeuroEngineering and Rehabilitation*, 2014.

[26] https://https://www.researchgate.net/figure.

[27] N. D. Reeves and F. L. Bowling. Conservative biomechanical strategies for knee osteoarthritis. *Medscape*, 2018.

[28] https://www.medscape.org/viewarticle/736532_4, .

[29] https://www.mayoclinic.org/diseases-conditions/osteoarthritis/diagnosis-treatment/drc-20351930.

[30] M. A. H. et al. Lateral trunk lean explains variation in dynamic knee joint load in patients with medial compartment knee osteoarthritis. *Elsevier Science*, 2007.

[31] K. et al. Characteristics of trunk lean motion during walking in patients with symptomatic knee osteoarthritis. *Elsevier*, 2008.

[32] M. Hun and J. Takacs. Effects of a 10-week toe-out gait modification intervention in people with medial knee osteoarthritis: a pilot, feasibility studys. *OARSI*, 2014.

[33] `https://ard.bmj.com/content/annrheumdis/66/10/1271/F1.large.jpg`, .

[34] `https:www.medscape.org/viewarticle/736532_3`, .

[35] E. A. S. et al. Effect of a home program of hip abductor exercises on knee joint loading, strength, function, and pain in people with knee osteoarthritis: A clinical trial. *Physical Therapy*, 2010.

[36] M. Billinghurst and T. Starner. Wearable devices, new ways to manage information. *Computer*, 1999.

[37] `https://www.researchgate.net/profile/Heres_Arantes_Junqueira/publication/322261039/figure/fig5/AS:579249787674624@1515115319835/Different-types-of-wearable-technology.png`.

[38] M. Ermes and J. P. et al. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, 2008.

[39] S. Lee and K. Mase. Activity and location recognition using wearable sensors. *IEEE pervasive computing*, 2002.

[40] K. Kunze and P. Lukowicz. Dealing with sensor displacement in motion-based onbody activity recognition systems. *Embedded Systems Lab*, 2003.

[41] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. *Massachusetts Institute of Technology*, 2004.

[42] E. M. T. et al. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. *IEEE International*, 2007.

[43] F. FOERSTER and J. FAHRENBERG. Motion pattern and posture:correctly assessed by calibrated accelerometers. *Behavior Research Methods*, 2000.

[44] Subramanya, A. Raj, J. Bilmes, and D. Fox. Recognizing activities and spatial context using wearable sensors. *arXiv.org*, 2012.

[45] J. Wu and A. O. et al. Recognizing activities and spatial context using wearable sensors. *IEEE 11th International Conference on Computer Vision*, 2007.

[46] L. X and Y. et al. Single-accelerometer-based daily physical activity classification. *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009.

[47] T. Brezmes1, J.-L. Gorricho2, and J. Cotrina2. Activity recognition from accelerometer data on a mobile phone. *Journal of agricultural*, 2004.

[48] S. P. et al. Feature selection and activity recognition from wearable sensors. *International Symposium on Ubiquitious Computing Systems*, 2006.

[49] M.-T. et al. Activity classification based on inertial and barometric pressure sensors at different anatomical locations. *Physiological Measurement*, 2014.

[50] F. Massé1 and R. R. G. et al. Improving activity recognition using a wearable barometric pressure sensor in mobility-impaired stroke patients. *Journal of NeuroEngineering and Rehabilitation*, 2015.

[51] E. G.-C. et al. Long-term activity recognition from wristwatch accelerometer data. *Sensors*, 2014.

[52] G. E. et al. An activity monitoring system for elderly care using generative and discriminative models. *Pers Ubiquit Comput*, 2010.

[53] L. T. V. al. Semi-markov conditional random fields for accelerometer-based activity recognition. *Applied Intelligence*, 2011.

[54] T. G. et al. epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition. *2009 IEEE International Conference on Pervasive Computing and Communications*, 2009.

[55] T. et al. Comparison of fusion methods based on dst and dbn in human activity recognition. *J Control Theory Appl*, 2011.

[56] Huynh, Blanke, and B. Schiele. Scalable recognition of daily activities with wearable sensors. *International Symposium on Location- and Context-Awareness*, 2007.

[57] D. M. K. et al. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, 2006.

[58] M. J. M. et al. Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, 2004.

[59] A. B. et al. Detection of type, duration, and intensity of physical activity using an accelerometer. *Medicine and Science in Sports Exercise*, 2009.

[60] F. C. et al. Joint segmentation of multivariate time series with hidden process regression for human activity recognition. *Elsevier*, 2013.

[61] G. L. et al. A description of an accelerometer-based mobility monitoring technique. *Medical Engineering Physics*, 2005.

[62] G. et al. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Elsevier*, 2014.

[63] S. I. L. et al. Activity detection in uncontrolled free-living conditions using a single accelerometer. *IEEE*, 2015.

[64] `https://medtechboston.medstro.com/wp-content/uploads/2014/07/KROMM_MedTechBoston2-830x404.png`, .

[65] http://srh-mal.net/wp-content/uploads/2015/03/Untitled.png, .

[66] *IMU User Guide*, .

[67] *LogAndStream for Shimmer3 Firmware User Manual Rev 0.9a*, .

[68] R. C. DIAS and J. M. D. DIAS. Impact of an exercise and walking protocol on quality of life for elderly people with oa of the knee. *Physiotherapy Research International*, 2003.

[69] G. D. D. et al. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee: A randomized, controlled trial. *Annals of internal medicine*, 2000.

[70] https://clinicalgate.com/wp-content/uploads/2015/03/B9780323056694100148_f14-01-9780323056694.jpg.

[71] R. S. H. et al. Delayed onset of quadriceps activity and altered knee joint kinematics during stair stepping in individuals with knee osteoarthritis. *Annals of internal medicine*, 2000.

[72] J. D. C. et al. Alterations in lower extremity movement and muscle activation patterns in individuals with knee osteoarthritis. *Science Direct*, 2003.

[73] M. P. M. S. et al. Range of joint motion and disability in patients with osteoarthritis of the knee or hip. *Rheumatology*, 2000.

[74] https://www.concept2.com/files/images/indoor-rowers/muscles-used/the-drive.jpg.

[75] https://taphysio.files.wordpress.com/2012/08/pedal_2915652a.jpg?w=460.

[76] http://scott.fortmann-roe.com/docs/docs/BiasVariance/biasvariance.png.

[77] M. R. et al. Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning Journal*, 2003.

[78] D. Davies and D. Bouldin. A cluster separation measures. *IEEE transactions on pattern analysis*, 1979.

[79] http://www.turingfinance.com/wp-content/uploads/2015/01/Davies-Bouldin-Index1.png.

[80] https://cdn-images-1.medium.com/max/1600/1*me-aJdjnt3ivwAurYkB7PA.png.

[81] G. C. Cawley and N. L. Talbot. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern recognition*, 2002.

[82] L. BREIMAN. Random forests. *Machine Learning*, 2001.

[83] https://cdn-images-1.medium.com/max/592/1*i0o8mjFfCn-uD79-F1Cqkw.png.