Physical Activity Recognition Using Wearable Accelerometers in Controlled and Free-Living Environments

Kostas Konsolakis

Technische Universiteit Delft

TU Delft Delft University of Technology

imec

Challenge the future

Physical Activity Recognition Using Wearable Accelerometers in Controlled and Free-Living Environments

by

Kostas Konsolakis

in partial fulfillment of the requirements for the degree of

Master of Science

in Biomedical Engineering

at the Delft University of Technology,

to be defended publicly on Monday July 16, 2018 at 10:00 AM.

Supervisor:	Dr. ir. R. Heusdens	
Thesis committee:	Prof. dr. ir. W. A. Serdijn,	TU Delft
	Dr. D. M. J. Tax,	TU Delft

Delft University of Technology

The work in this thesis was supported by IMEC (Interuniversitair Micro-Elektronica Centrum) in Leuven, Belgium. Their cooperation is hereby gratefully acknowledged.

To my parents.

Abstract

Physical activity recognition through wearables has enabled the development of novel applications in healthcare. Most of the existing studies focus on predicting activities using wearable sensors, either in a controlled or uncontrolled environment. However, there is not a clear distinction between these two environments. Hence, this thesis aimed to answer the research question *"How accurately can we classify physical activity based on wearable accelerometers placed on the wrist and chest in a controlled and in a free-living environment?"*.

For the data collection phase, two experiments were conducted in the working environment of imec. 40 participants were recruited and were asked to participate in the Controlled and Free-Living Study. The subjects wore two imec wearables, a wrist-worn and chest-worn accelerometer sensor and performed everyday activities. These activities include sitting, dynamic sitting, lying with face up and face down, lying to the left and right, standing, dynamic standing, walking upstairs, walking downstairs, walking, running, and cycling. The Controlled Study showed that most of these activities could be detected accurately using accelerometer data from both sensors with 91.83% F1-score. Similarly, the combination of these two sensors achieved the best performance for the Free-Living Study with 86.98% F1-score. Finally, this work proved that between the two environments a correlation could be possible only for the activity cycling. Consequently, this research concludes that the activity recognition should be explicitly investigated in free-living environments, focusing on real-time activity detection.

ACKNOWLEDGEMENTS

I could never have imagined that moving to the Netherlands to continue with my studies would be such a great and rewarding journey. Not only did I meet incredible people from around the world, including skillful professors and colleagues, but also I made amazing friends. This piece of work is the result of my fruitful collaboration with many of them. Even though expressing my gratitude in words cannot do it justice, I would like to thank a considerable number of people.

First of all, I would like to thank my supervisor Dr.ir. Richard Heusdens for willing to guide and supervise me during the thesis project, even if I had to be abroad. Furthermore, I would also like to express my thankfulness to the professor Dr. Wouter Serdijn who was my master's coordinator in the Bioelectronics group, always providing an elaborated answer to my questions. Together with Dr. David Tax, they were the members of my thesis committee, and thus, I would like to thank them both.

Additionally, I would like to express my gratitude for working at imec in Leuven during the second year of my studies, where I completed my graduation project. Primarily, I would like to thank Dr. Chris Van Hoof and Dr. Walter de Raedt and in particular the Ph.D. candidates Elena Smets and Eric Zhai for being my daily supervisors at imec. I genuinely appreciate the opportunity you gave me to conduct my own experiment and research a topic of my highest interest. Of course, I could never forget the enormous contribution of all the forty subjects, who participated voluntarily in my experiment. For once again, thank you all!

Moreover, I would like to thank my crazy friends from Delft and Leuven for making this study period unique. Makrina, Lucia, Jacopo thank you for constantly being next to me and for all the buonissimo carbonara nights we had; I am looking forward to sharing more craziness with you. Giannis and Giorgos, thank you for guiding me during my first steps in Delft. Ammar, Niovi, Stamatis, Stefanos, Antwnis, Maria, Evaggelia, . . . , dank je wel!

Writing these words, I do feel also obliged to thank Dr. ir. Oresti Banos who has been my supervisor for the last months in the next step of my career, as a Ph.D. candidate. Even if I did not expect that, your contribution for writing this thesis report was significantly valuable.

Finally, I would like to thank my parents for their endless support throughout my academic life and for giving me the opportunity to continue with my studies abroad. Σας ευχαριστώ!

Contents

List of Figures

LIST OF TABLES

1

INTRODUCTION

Wearable health has become one of the most promising research areas, being capable of tracking medically useful information in order to provide diagnostic and monitoring applications. Cutting-edge health devices have been used for diagnostic and monitoring applications by recording physiological, biochemical and motion data [1, 2]. These devices are constantly improving in terms of technology, functionality, and size, with the focus on real-time applications [3, 4].

The use of motion data through unobtrusive wearables plays an important role in the detection of human physical activity and, thus, is commonly available in a wide variety of sensors (e.g., wearables, smartphones). A successful system for human activity recognition aims to monitor and analyze human activities, but also to interpret ongoing activities, regardless the person or the environment where the activities are performed. Physical activity recognition is of major importance for many medical, military and security applications [5]. Specifically, there has been an extensive research focusing on the positive effects of physical activity on health [4]. For instance, Biddle and Asare [6] proved that mental illness is related to physical inactivity. Additionally, Chodzko-Zajko [7] studied the benefits of being physically active in healthy aging, while he also proved that physical inactivity could lead to the development of numerous chronic diseases and conditions. Overall, physical activity has been recommended as a therapeutic intervention for the management and treatment of many chronic conditions and diseases, such as depression and anxiety disorders [8], dementia [9], type 2 diabetes [10], obesity [11], osteoporosis [12, 13], hypertension [14, 15], and coronary heart disease [16, 17].

Recognizing daily activities, such as running, cycling, walking or sitting, allows physicians and caregivers to monitor patients' health and physical behavior. For instance, patients who are advised to be physically active and follow a healthy lifestyle can receive real-time feedback and coaching services. Furthermore, patients with chronic conditions and mental pathologies could be monitored to detect abnormal activities and prevent undesirable behavior. Thus, physical activity recognition has a great number of applications in different areas, such as medical monitoring, assistant living, active living, and rehabilitation, contributing on human behavior understanding and coaching services [4].

The last two decades, there have been two main approaches for activity recognition [18, 19]. The first one is vision-based and requires the use of video cameras as a recording tool for visual sensing facility. The second approach is sensor-based and requires the use of emerging sensor networks. Regarding the vision-based approach, this demands the continuous monitoring of an actor's behavior, as well as the detection of any environmental changes [18]. Hence, this approach lacks portability and may raise privacy concerns. On the other side, the recent advances in sensor technologies, combined with the significant progress of wired/wireless communications and data processing techniques, have enabled the research to shift from a low-level quality of data acquisition and processing towards a high-level quality of integrated information [18]. As a result, there is a growing interest in the field of activity recognition through wearables and there have been many studies related to this research.

The recent technological advances led to the arrival of new pioneer wireless devices, such as wearable sensors and mobile devices, enabling the development of novel applications for activity recognition. The latest wearables that have been designed for activity detection consist of different sensors, such as accelerometers, gyroscopes, magnetometers, and GPS [4]. Specifically, accelerometer sensors have been widely used in many studies for activity recognition, proving that they are adequate for monitoring simple activities, such as walking, running, sitting, etc. [18, 20, 21]. However, physical activity recognition is still a complicated process with many challenges that need to be addressed [4, 19].

Based on recent studies, the development of a system for activity recognition demands new methods in order to improve the accuracy under more realistic conditions [4, 5, 18]. These methods include the following tasks: 1) Selection of necessary attributes for recording data; 2) Selection of appropriate sensors for developing the data acquisition system; the optimal devices should be portable, light-weight, and unobtrusive, with a sufficient memory capacity and low power consumption; 3) Protocol design for acquiring data; 4) Selection of appropriate methods and tools for enhancing the classification performance; this task also involves the processing and analysis of the collected data; 5) Software implementation, considering energy and processing requirements; and 6) System flexibility, by supporting new users without retraining the system.

Most of the existing systems in activity recognition vary on the number of sensors and their placement, the number of subjects who participate in the data collection, the settings where the activities are performed, the number of computed features and the number and type of detected activities [22]. Thus, it is difficult to compare a new proposed method to other existing approaches. Furthermore, there is not a clear investigation for detecting activities in both controlled and uncontrolled environments, examining for instance if data from controlled settings can be used to predict activities in uncontrolled environments.

The purpose of this thesis report is to develop a model for activity recognition, using accelerometer data from wearable devices, placed on the wrist and chest, and answer the following research question: *"How accurately can we classify physical activity based on wearable accelerometers placed on the wrist and chest in a controlled and in a free-living environment?"*. Trying to answer this research question, we will conduct two experiments in both controlled and uncontrolled environments, after reviewing the current trends in activity recognition. It is worth mentioning that the literature review will emphasize on the optimal placement of wearable sensors, trying to find the most significant parameters for data acquisition, and examine the processing and classification techniques that are currently used for activity recognition.

The succeeding chapters will be structured in the following way. Chapter 2 will present the state-of-the-art research in activity recognition, focusing on wearable sensors placed on the user's wrist and chest. Chapter 3 will elaborate our methodology for collecting and analyzing accelerometer data in a controlled and uncontrolled environment. Chapter 4 will evaluate the activities performed in a controlled environment and will present the results for the Controlled Study. Chapter 5 will investigate the activity recognition in uncontrolled settings, providing the related results for the Free-Living Study. Finally, the thesis will conclude with a summary and future recommendations in chapter 6.

2

RELATED WORK

This chapter elaborates the methodology for conducting the literature research and presents the results related to activity recognition through wearable accelerometers.

2.1. METHODS FOR LITERATURE RESEARCH

The literature review is based on three of the most popular web search engines for scientific and academic articles; 'Web of Science', 'Scopus' and 'Google Scholar'. 'Web of Science' is a web subscription-based scientific citation indexing service, which gives access to multiple databases for cross-disciplinary research [23]. Its editors have been monitoring all these databases in order to evaluate and collect journals, without any conflict of interest. Thus, many benefits are provided to the users of this service, such as information about the content, the number of citations and self-citations of the journals. Similarly, 'Scopus' database requires a subscription and covers three main types of sources; book series, journals, and trade journals through high-quality standards [24]. In contrast, 'Google Scholar' is a web search engine that gives access to its bibliographic database without a subscription. Based on its automatic crawler, the user can adjust the query and search among online academic literature, including journals and books, conference papers, theses, technical reports and patents [25].

The literature research was divided into four main categories. First, the use of wearable sensors, as well as their optimal placement on the body, was reviewed. Second, the activity recognition based on accelerometer data was thoroughly examined. According to the results, the activity recognition can be significantly accurate and less obtrusive by placing the sensors on the wrist and the chest. Hence, the third part of the literature research focused on activity recognition using accelerometer wearables on the wrist and the chest. Finally, different preprocessing and classification techniques were examined.

The primary keywords used for searching in the abovementioned search engines were the following: *'activit*', 'activity recognition', 'activity detection', 'physical activity recognition', 'wearable sensors', 'wearables', 'sensors', 'specifications', 'placement', 'optimal placement', 'physiological data', 'bio-signals', 'motion data', 'acceler*', 'accelerometer', 'wrist*', 'wrist sensor', 'smartwatch*', 'chest*', 'chest sensor'. The secondary keywords were the following: 'experiment', 'data collection', 'data acquisition', 'protocol', 'control experiment', 'laboratory experiment', 'free-living experiment', 'signal processing', 'features', 'feature extraction', 'feature selection', 'window segmentation', 'window size', 'overlap', 'classification performance', 'classification algorithm',* and *'metrics'*. Additionally, the combination of some keywords was also used (see Table 2.1). Based on the four main categories of this literature research, the results were reviewed giving priority to the most highly 'cited by' articles.

A significant number of published articles, journals, papers and other reports exist already in the field of activity recognition. Thus, it is important to select the most relevant search queries. For instance, the search query "acceler*" (including keywords such as accelerometer, accelerometers, acceleration, etc.) returns arti-

Table 2.1: Different search queries on 'Web-of-Science' are presented for the literature research on activity recognition. The range of the publication year is up to 27 February 2017.

Search Query on Web-of-Science	Publication Year: 1900 - 2017	Publication Year: 2000 - 2017
*acceler**	515.196	405.717
accelerometer	27.146	24.725
*activit**	3.437.773	2.494.631
'activity recognition'	63.416	51.938
activit AND acceler**	55.108	46.238
'activity recognition' AND accelerometer	949	949
activit AND (chest* OR (wrist* OR (smart AND watch*) OR smartwatch*))*	11.400	8.915
'activity recognition' AND (chest OR wrist)	311	286
activit AND chest**	6.392	4.726
activit AND (wrist* OR (smart AND watch*) OR smartwatch*)*	5.068	4.245
activit AND (chest* AND (wrist* OR (smart AND watch*) OR smartwatch*))*	60	56
acceler AND (chest* AND (wrist* OR (smart AND watch*) OR smartwatch*))*	44	43
activit AND acceler* AND (chest* AND (wrist* OR (smart AND watch*) OR smartwatch*))*	24	24

cles related not only to activity recognition but also to other studies such as fall detection, stress detection, eating habits, gestures recognition, etc. Table 2.1 describes the returned results based on different search queries, using for instance the 'Web of Science' search engine. Finally, the selected search query is *"activit* AND acceler* AND (chest* AND (wrist* OR (smart AND watch*) OR smartwatch*))"*.

The initially yielded results were 118 articles, journals, and papers, including duplicates and no-related fields such as economics, chemistry, and environmental studies. The process of reviewing and excluding the results from the three web-search engines, is presented in Figure 2.1. In total, 13 articles were related to the purpose of this literature review and will be discussed in the following sections. An overview of this articles is presented in Table 2.2.

2.2. WEARABLE SENSORS

During the last decade, the rapid advancement of microelectronics and micromechanics enabled the development of state-of-the-art sensors, which can sense and measure data in a fast and efficient way by using minimum processing resources and energy consumption. Sensor technologies have been progressively advanced to miniaturized, low power and cost and high capacity wearable sensors, with the main focus on health and fitness applications [26].

There is a significant number of electronic wearable devices, already used for commercial and research purposes, varying from smartwatches and phones, to activity trackers and heart rate monitors. These devices are implemented with different types of sensors, according to the task of monitoring. As a result, wearable sensors, which are sensors positioned directly or indirectly on the human body, can be used for instance to record information related to heart rate, pulse, skin temperature, skin conductance, body position, and movement [27].

Activity recognition is mainly based on motion and physiological sensing technologies, by recording the movement and the physiological state of a person when different activities are performed. Motion signals contain information from inertial sensors, such as accelerometer, gyroscope and magnetometer, and location sensors such as GPS trackers. On the other side, physiological signals contain information based on vital signs data, such as temperature, skin conductivity, respiration rate, heart rate, ECG, EEG, etc.

2.2.1. MOTION SENSORS

The rapid development of MEMS (Micro-Electro-Mechanical Systems) enabled the design of small-size, light-weight and low-cost inertial sensors, being capable of monitoring the movement of the human body in every dimension. However, the number and the location of the inertial sensors on the body varies based on the task and the performed activities [19].

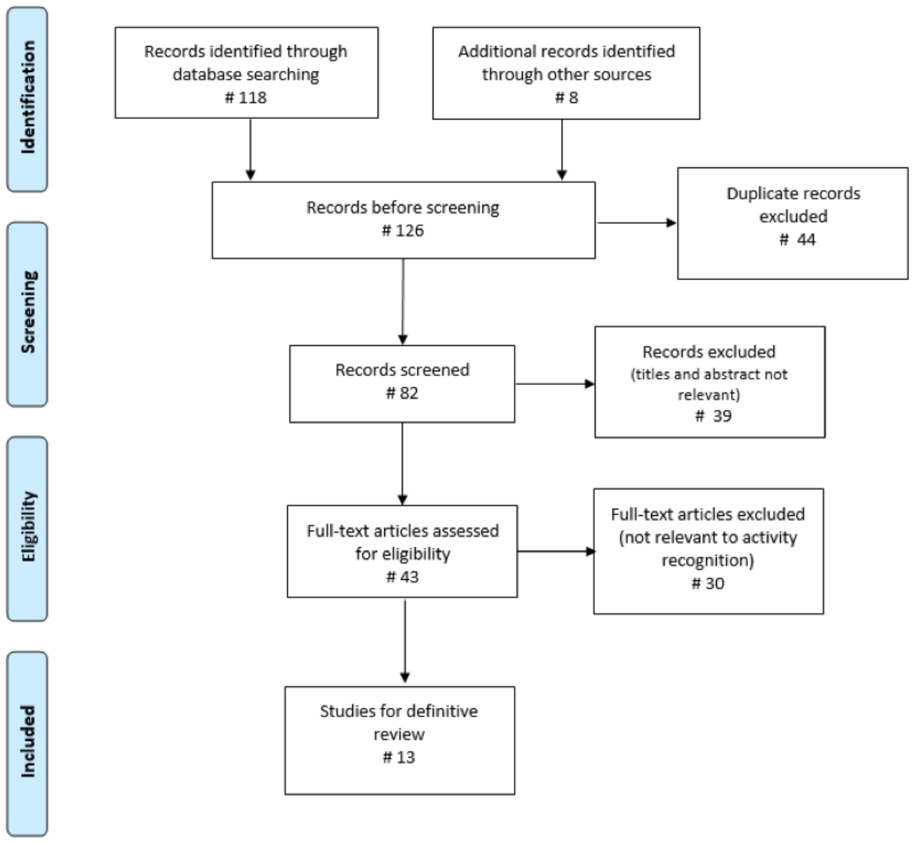


Figure 2.1: Prisma flow chart diagram. Presenting the process of screening the records identified from the Web-of-Science, Scopus and Google Scholar search engines, based on the selected search query. Overall, 126 initially yielded results were limited to 13 for the definitive review.

Accelerometers are the most commonly used wearable sensors for activity detection and are used to measure acceleration by computing changes in position and velocity along a sensitive axis and over a range of frequencies [18, 26]. Three main types of accelerometers, that are used to convert mechanical motion to an electrical signal, are piezoelectric, piezoresistive and capacitive accelerometers [26]. Each one of them follows the same principle where the accelerometer operates as a damped mass on a spring and acceleration is caused by stretching or compressing the spring proportionally. Accelerometers are frequently used to monitor activities such as sitting, standing, walking, climbing stairs, running and cycling [18, 27].

Gyroscopes provide angular rate information and are used to measure changes either in orientation or rotational velocity, while magnetometers determine the absolute orientation, by measuring magnetic fields [28]. Leutheuser et al. [29] proposed a system based on tri-axial accelerometer and gyroscope sensors, placed on wrist, chest, hip and ankle, in order to recognize thirteen activities, including household activities (e.g., washing dishes, sweeping), postures (e.g., sitting, standing, lying), walking behaviors (e.g., walking, running, stairs climbing) and sports activities (e.g., bicycling, rope jumping). They managed to obtain a mean classification rate of 85.8%, and complex activities such as ascending and descending stairs were distinguished successfully. Altun et al. [28] used a tri-axial accelerometer, magnetometer and gyroscope sensors, placed on the users' arm, chest, and leg, and managed to recognize nineteen activities (e.g., standing, lying, climbing stairs, walking, running, cycling, playing basketball) with 87% accuracy.

GPS sensors are widely used for monitoring location-based activities. Specifically, Ashbrook and Starner [30] proposed a system based on GPS data to automatically detect meaningful locations and predict movement in a multi-user scenario. Liao et al. [31] used GPS data logs to detect normal user's behavior (preferable destination and mode of transportation per user) or abnormal behaviors (such as taking a wrong bus). Similarly, Patterson et al. [32] used GPS sensor stream to detect user's behavior by predicting the location and transportation mode (such as traveling and boarding from a specific bus stop). Additionally, Riboni and Bettini [33] studied a system based on accelerometer sensors, placed on the wrist and on phone, combined with GPS data from the user's phone in order to recognize everyday activities (such as walking, jogging, brushing teeth, writing on the blackboard, etc.). They concluded that the use of GPS led to a moderate energy efficient system with a total accuracy above 90%, while some activities such as standing – writing and hiking up – hiking down were misclassified.

2.2.2. PHYSIOLOGICAL SENSORS

Physiological signals (also known as bio-signals) refer to vital signals and have been considered on a few works for activity recognition. For instance, ECG sensors can be related to activity recognition systems, while they are mainly used to extract useful information about the rate and the regularity of heartbeats, contributing to the short-time diagnosis of cardiovascular diseases and stress detection. Tapia et al. [34] proposed a system for activity recognition based on five triaxial accelerometers (placed on the user's ankle, hip, thigh, dominant arm, and dominant wrist) combined with a heart rate monitor placed on the chest (leading to an obtrusive system with high energy consumption). However, they found that heart rate does not contribute to activity recognition systems significantly (80% averaged accuracy), since dynamic activities, such as running, may lead to an increased heart rate level even when the subject is resting (lying or sitting after the activity). Furthermore, Lara et al. [35] studied the use of a chest strap, measuring acceleration and physiological signals (such as heart rate, respiration rate, skin temperature, etc.) in order to recognize five activities (sitting, running, walking, ascending and descending stairs) with 95.7% overall accuracy and 92.8% accuracy considering only accelerometer data.

2.3. OPTIMAL PLACEMENT OF SENSORS

There is a large variety of experiments studying the use of multiple wearables on different parts of the human body, such as the chest, the wrist, the arm, the hip, the thigh, the knee and the foot (see Figure 2.2). It was proved that activity recognition is related to the placement of wearable sensors and can be affected by the location where a sensor is placed on the body, or the way that the sensor is attached to the body [36]. The majority of these studies focused on the placement of multiple wearables, mainly based on accelerometers, on different parts on the body in order to discover the most optimal number of sensors and their suitable placement [5, 36].

Cleland et al. [37] studied everyday activities, such as walking, jogging, sitting, lying, standing and climbing on stairs, based on accelerometer data from six wearables placed on the chest, wrist, hip, thigh, foot and waist, and they concluded that the hip-attached sensor enhanced the classification performance. Pannurat et al. [38] investigated activities such as lying, sitting, standing and walking, by placing accelerometers on the chest, wrist, waist, thigh, head, upper arm and ankle. They concluded that placing the wearables on the thigh and chest can significantly enhance the classification performance.

Gjoreski at al. [39] studied the optimal placement of four accelerometers on the chest, waist, ankle, and thigh for fall detection, indicating that the best results were obtained by combining sensors placed at the chest and ankle, or at the waist and ankle. In a later work, they also studied the detection of activities such as walking, standing, sitting, lying, bending, kneeling, cycling and running and they proved that accelerometers placed on the ankle and thigh perform slightly better than placing on the wrist and chest [40]. Olguin and Pentland [41] evaluated the accuracy of monitoring everyday activities, such as sitting, running, walking, standing, lying and crawling, and proved that accelerometer sensors placed on the wrist, chest and hip could lead to an accurate model for activity recognition. Similarly, Chamroukhi et al. [42] proved that combining accelerometers located at the upper and lower parts of the body, such as chest, thigh, and ankle, can improve activity recognition significantly.

Additionally, many studies demonstrated that wearables placed on the waist could detect everyday activities, such as sitting, standing, lying, climbing stairs, walking and running since the sensors monitor the center of body mass [43–46]. Wrist-worn accelerometers were used to identify activities, such as walking, running, climbing stairs [47–51], to detect fall [40], and to estimate activity levels during sleep [52]. Ankle-attached wearables were used to identify steps, travel distance, velocity, and energy expenditure [53], while head-attached wearables were used to estimate balance during walking [54]. Furthermore, rapid development in technology enabled the researchers to put a spotlight on activity recognition using data acquired from smartphones, where smartphones are placed in different locations on the body (e.g., trouser pocket, t-shirt pocket, belt, etc.) being capable of monitoring simple everyday activities [55].

Overall, it is proved that the chest, wrist, hip, thigh, and waist are optimal locations for wearing accelerometers on the body and can identify everyday activities, such as sitting, lying, walking, running and cycling [36, 41, 56]. Despite the enhanced contextual information, wearing multiple sensors can be obtrusive and uncomfortable for the users. On the one hand, the classification performance could not be enhanced by increasing the number of sensors [37, 41, 58]. On the other hand, the use of a single accelerometer decreases the number of activities that can be predicted [37, 57]. Consequently, many researchers recommend using only two accelerometers, being worn in an unobtrusive location and firmly attached to the body, for detecting different activities accurately.

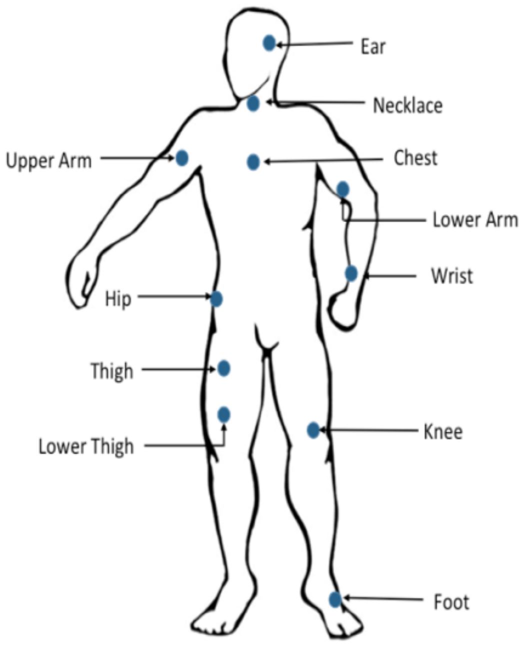


Figure 2.2: Wearable sensors placement [36].

2.4. ACTIVITY RECOGNITION BASED ON ACCELEROMETERS PLACED ON THE WRIST AND CHEST

There is a significant research on activity recognition using accelerometers placed on the wrist and the chest, either separately or in combination with other sensors (see Table 2.2). However, only a few studies are focusing on the combination of solely these two sensing locations.

Chernbumroong et al. [49] examined the activity classification using a single accelerometer on the wrist, detecting five basic daily activities in a controlled environment and using seven subjects, with a total 94.13% accuracy. Similarly, Yang et al. [50] studied the placement of a single accelerometer on the dominant wrist in order to detect eight daily activities in a controlled experiment, with an overall accuracy of 95%. Garcia-Ceja et al. [59] used a smartwatch to record accelerometer data for six long-term activities in a partially free-living environment (user-based annotation; some of the activities were not annotated appropriately), using only two subjects and achieving an overall accuracy of 91.8% (with 88.47% precision).

Lara et al. [35] measured chest-worn acceleration in order to recognize five activities in a free-living experiment (user-based annotation using a mobile application), recruiting eight subjects and achieving 92.8% accuracy. Khan et al. [60] recorded chest-worn accelerometers data to detect fifteen activities from six subjects, in an uncontrolled environment (user-based annotation using a Bluetooth headset), with an average accuracy of 97.9%.

Olguin and Pentland [41] evaluated the accuracy of monitoring three subjects, performing eight everyday activities in a controlled environment and wearing accelerometer sensors on the wrist, chest, and hip, with a 92.13% classification performance. It is worth mentioning that the chest-worn sensor gave 62.45% average accuracy, while a second accelerometer on the hip or the wrist improved the performance by approximately 20%. Arif and Kattan [21] monitored nine subjects performing twelve activities and indicated that accelerometers worn on the chest, wrist, and ankle, can result in 98% (F-score). This dataset (PAMAP2) was studied further from Xu et al. [61] with 93% accuracy. Their study aimed to examine activity data with properties such as nonlinearity and non-stationarity, by investigating the characteristics of the Hilbert-Huang transform (HHT). Furthermore, Kikhia et al. [62] evaluated the accelerometer placement on the wrist, chest, and thigh in a controlled environment, based on ten subjects, and they recognized different everyday activities with an average 85% accuracy performance.

Parkka et al. [20] is the only study focusing on activity recognition based on two accelerometers, one worn on the chest and one on the wrist, in a free-living environment (researcher-based activity annotation using a mobile application). Sixteen subjects participated in the experiment, performing different tasks in different locations, such as lying, sitting, standing, walking, Nordic walking, rowing, running, and cycling. The overall classification was 83.3%, using decision trees. Besides, other signals such as magnetometer, heart rate, and respiration rate were recorded. However, they found that accelerometers provided the most valuable context information and were considered as the most accurate sensors for activity recognition. A limitation of this study was the use of a rucksack strap to place the chest sensor, which was not firmly attached to the body leading to a misclassification on activities such as sitting and standing. Additionally, data were collected only for two hours which could be a limitation for performing activities in realistic conditions.

2.5. ANALYSIS ON SENSOR DATA

Activity recognition is a classification problem and requires processing techniques in order to extract useful information from raw data [4, 63]. For this reason, different steps are involved, such as data acquisition, processing, classification, and evaluation.

2.5.1. DATA ACQUISITION

Depending on the application and the task of monitoring, multiple signals can be recorded. Hence, it is crucial to select the most relevant sensors and their attributes. Garcia-Ceja et al. [59] mentioned that sampling rate could affect the system performance; a sampling rate of 5 Hz decreases the accuracy, while sampling rates above or equal to 10 Hz enhance the activity classification [64].

Furthermore, it is important to consider an effective data collection system which consists of portable, light-weight, unobtrusive, affordable wearable sensors, with sufficient memory capacity and low power consumption [18]. Based on the literature, the experiments for activity recognition can take place either under controlled conditions (also known as the laboratory experiment) or under uncontrolled conditions (also known as the free-living experiment). The former takes place in a controlled environment, where the subjects receive specific instructions and act respectively. The latter takes place in an uncontrolled environment, where the subjects perform different tasks during their daily life, without receiving any instructions. A limitation of the existing studies in uncontrolled environments is the limited amount of collected data and the use of self-reported activity annotation as ground truth [35, 59, 60]. Specifically, the participants are asked to annotate the performed activities by themselves, which might result in wrong labeling and activity misclassification. Overall, most models for activity recognition are validated in controlled settings, while only a few studies focus on uncontrolled environments.

2.5.2. DATA PROCESSING

Raw data are processed in order to synchronize, filter and replace missing values, and extract important features. Extracting features from raw data involve dividing sensor signals into smaller window segments, using different window segmentation techniques. The most widely used windowing technique for real-time applications is the one of window sliding, where signals are divided into fixed-length windows [36]. According to Garcia-Ceja et al. [59], long-term activities are characterized by a sequence of primitives, which can be obtained by dividing data into fixed length windows. Long-sized window segments affect the detection of short-duration movements, such as the transition between sitting and standing [60], and thus, using smaller window sizes with a fixed 50% overlap is recommended [5, 58, 65]. Furthermore, most of the existing activity recognition systems calculate and extract features from raw accelerometer signals based on time and frequency domain features [62, 66]. These features provide information related to body posture, motion shape, motion variation, and motion similarity (correlation). An overview of these features is presented in the following picture (Figure 2.3).

Time-domain features contain basic waveform characteristics and statistical properties of the raw data and aim to differentiate dynamic from static movements [62]. These features include mean, median, min, max, variance, standard deviation, skewness, kurtosis, magnitude, area, and correlation between axes and are proved sufficient for activity recognition systems [21, 35].

Frequency-domain features contribute in distinguishing moderate from vigorous movements, by identifying essential patterns on raw signals [62]. Frequency domain features focus on the periodic structure of the signal and include features such as energy, entropy, peak frequency, Power Spectral Density (PSD), Discrete Fourier Transform (DFT), and other coefficients of the Fourier transform [41, 61].

Inappropriate or redundant features could decrease the classification performance. For this reason, feature selection reduces the number of features, based on the optimal discriminative power between classes, by selecting features that contribute most to the prediction. Other techniques used for features reduction are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Discrete Cosine Transform (SCT) [5, 36].

2.5.3. CLASSIFICATION

There are two main approaches for classification problems in activity recognition systems, related to supervised and unsupervised algorithms [36]. Supervised learning approaches require annotation of the activities and include algorithms, such as Support Vector Machines (SVM) [62], Artificial Neural Networks [50, 61] (also found in unsupervised problems), naïve Bayes [38], nearest neighbor [38], and tree-based modeling like Decision Trees [20, 49] and Random Forests [62]. Unsupervised learning approaches enable activities annotation automatically and include algorithms such as Hidden Markov Models (HMMs) [41, 59]. Overall, supervised classification is mainly used in controlled settings, while unsupervised is preferred in uncontrolled settings where the labels annotation is insufficient.

Kikhia et al. [62] showed that the classification algorithms should be considered wisely in activity recognition systems since their selection depends on the task of monitoring (target activities) and the placement of the sensors on the body. For instance, they proved that the wrist-worn sensor is the best location for classifying various body movements using the Random Forest classifier. Similarly, Pannurat et al. [38] proved that the classification algorithm should depend on the sensor placement. They studied different algorithms, and they found that the nearest neighbor performs better for sensors placed on the thigh, wrist, and arm, while naïve Bayes is preferred for chest, waist, head and ankle worn sensors. Some of the most frequently used algorithms for activity recognition systems are decision tree [20, 49], random forest [40, 62], neural networks [50], SVM [37], naïve Bayes [38], and nearest neighbor [38].

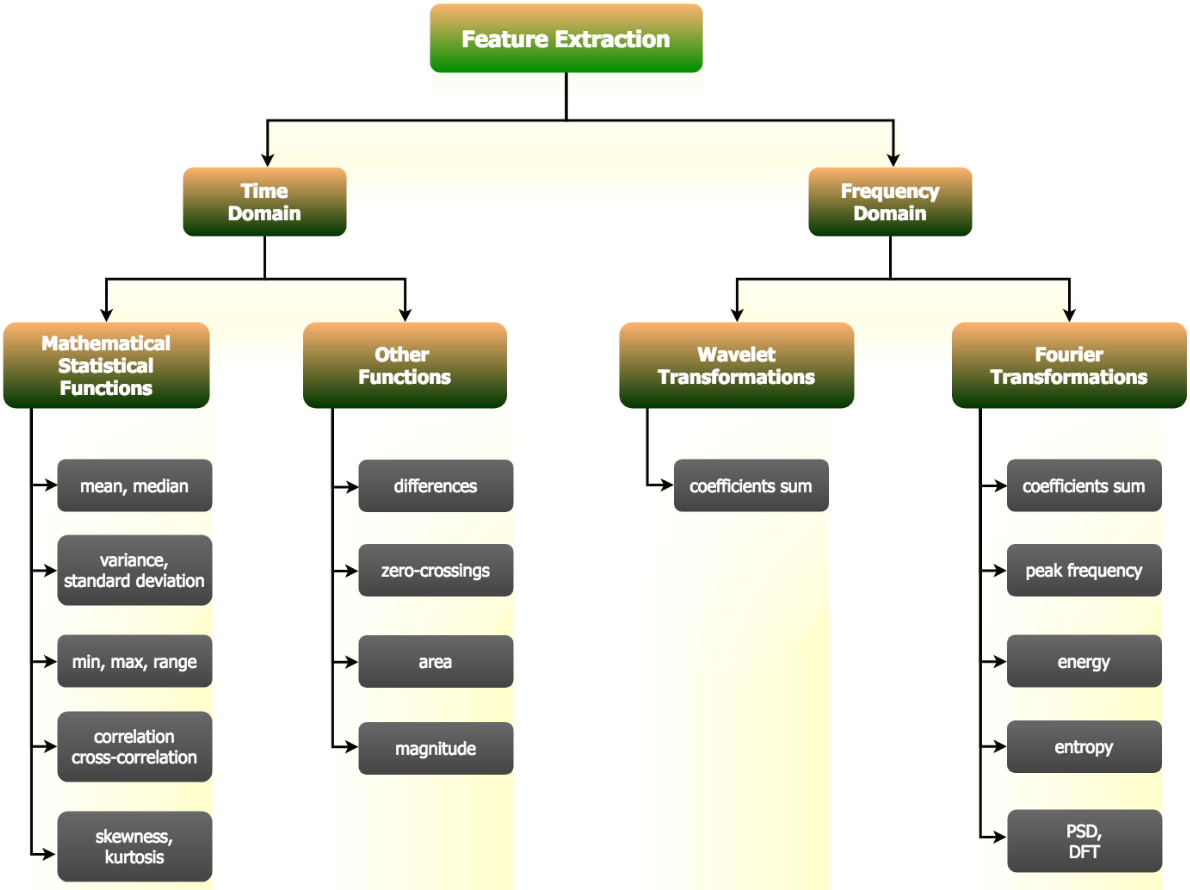


Figure 2.3: A schematic overview of feature extraction.

2.6. DISCUSSION

Activity recognition is a classification problem, which depends on the application and the task of monitoring [4]. Various studies have been already done in this research area. However, it is difficult to compare their performance since different type of sensors (accelerometers, gyroscopes, etc.) are placed on different locations on the body. Even though wearable accelerometers are adequate for activity recognition systems, their placement should be related to the targeted type of activity [62]. For instance, multiple sensors have been used to recognize everyday activities [21, 41, 61, 62], low-intensity activities [66], sports activities [67], stress detection [68] and energy expenditure [69]; hand and wrist-worn sensors have been used to recognize daily hand gestures [70], to detect fall [40] and to estimate activity levels during sleep [52]. However, the data collection of each study is based on different experimental protocols, which vary concerning the nature and the number of the detected activities, the number of recruited subjects, and the total duration of the activities recording. Additionally, different algorithms are trained, using different evaluation criteria (accuracy, precision, recall, f-measure, etc.) and following different validation procedures (P-fold, leave-one-out, etc.).

There is a significant number of studies that proved that accelerometers worn on the wrist, chest, thigh, and hip are adequate when the activities are performed under controlled conditions [21, 41, 49, 61, 62]. Only a few studies have been performed under uncontrolled conditions (free-living experiments), using a limited number of subjects (the average number of recruited subjects is less than twenty), and insufficient recording time of activities [20, 35, 59, 60]. Additionally, the labels annotation in several uncontrolled studies depends on the recruited subjects, who are requested to annotate the performed activities during the experiment, using subjective criteria. The most common method for activity annotation in a free-living environment is based on a mobile application, causing several limitations to the evaluation of the system's performance. For instance, some subjects may annotate the activities inaccurately, while other subjects may not annotate all the performed activities. Nevertheless, a wrong activity annotation could lead to a misclassification of the system. Furthermore, it is questionable if models developed based on controlled experiments can be translated to uncontrolled environments. Hence, activity recognition between controlled and uncontrolled environments needs further investigation.

Considering that the classification performance could not be further enhanced by using more than two wearables, many researchers recommend acquiring accelerometer data from maximum two wearables [37, 41, 58]. Even though wearables placed on the wrist and the chest provide an accurate system for activity recognition [35, 49], only a few studies focus on the combination of these two sensing locations [20], either in controlled or uncontrolled conditions. Consequently, the following research question was generated; *"How accurately can we classify physical activity based on wearable accelerometers placed on the wrist and chest in a controlled and in a free-living environment?"*. This graduation project aims to answer the research question by evaluating accelerometer data for everyday activities, performed both in a controlled and in an uncontrolled environment.

In order to give a complete answer to the above research question, we will also investigate the following sub-questions:

1. *Can 3-axial accelerometer data, from a single wrist-worn or chest-worn sensor, be used to detect simple everyday activities?*
2. *How does combining accelerometer sensors, placed on the wrist and chest, affect the accuracy of activity recognition? Does the classification performance improve for predicting certain types of activity?*
3. *How well does a sensor contribute to detecting both static and dynamic activities, such as static sitting and dynamic sitting, by recording accelerometer data from a single sensor?*
4. *Is it possible to detect activities performed in uncontrolled settings, through a classification model that was trained with data from a controlled environment?*

Table 2.2: Overview of the studies focusing on wearable accelerometers worn on the wrist and the chest.

Authors	Placement of Accelerometer	Experiment	Number of Subjects	Features	Classification Performance (%)	Detected Activities
Olguin and Pentland (2006) [41]	wrist, chest, hip	controlled	3	mean, variance	accuracy: 92.13% (Hidden Markov Model)	Static & Dynamic Activities: sitting, running, walking, standing, dynamic standing, lying down, crawling, and squatting
Parkka et al. (2006) [20]	wrist, chest	uncontrolled	16	mean, variance, median, skewness, kurtosis, spectral centroid, spectral spread, estimation of frequency peak and signal power	accuracy: 83.3% (Decision Tree)	Static & Dynamic Activities: lying down, sitting, walking, rowing, and cycling
Yang et al. (2008) [50]	wrist	controlled	7	mean, correlation between axes, energy, interquartile range, mean absolute deviation, root mean square, standard deviation, and variance	accuracy: 95% (Neural Network)	Static & Dynamic Activities: walking, running, scrubbing, standing, working at a PC, vacuuming, brushing teeth, and sitting
Khan et al. (2010) [60]	chest	uncontrolled	6	mean, standard deviation, spectral entropy, and correlation between axes	accuracy: 97.9% (Autoregressive Modeling)	Static Activities: sitting, lying, standing; Transitions: lie-stand, stand-lie, lie-sit, sit-lie, sit-stand, stand-sit, walk-stand, stand-stand; Dynamic Activities: walking, running, and ascending/descending stairs
Chernbumroong et al. (2011) [49]	wrist	controlled	7	mean, minimum, standard deviation, variance, correlation between axes, difference between axes, spectral energy, and spectral entropy	accuracy: 94.13% (Decision Tree)	Static & Dynamic Activities: sitting, standing, lying, walking, and running

Continued on next page

Table 2.2 – *Continued from previous page*

Authors	Placement of Accelerometer	Experiment	Number of Subjects	Features	Classification Performance (%)	Detected Activities
Lara et al. (2012) [35]	chest	uncontrolled	8	mean, variance, standard deviation, correlation between axes, interquartile range, mean absolute deviation, root mean square, and energy	accuracy: 92.84% (Additive Logistic Regression)	Static & Dynamic Activities: walking, running, sitting, and ascending/descending stairs
Cleland et al. (2013) [37]	chest, wrist, hip, thigh, foot and waist	controlled	8	mean, total mean, standard deviation, total standard deviation, skewness, total skewness, kurtosis, total kurtosis, energy and correlation between axes	accuracy: 96.67% (SVM)	Static & Dynamic Activities: walking, jogging, sitting, lying, and ascending/descending stairs
Garcia-Ceja et al. (2014) [59]	wrist	uncontrolled	2	mean, variance, correlation between axes; and the mean, variance and average derivative of the magnitude	accuracy: 91.8% (Hidden Markov Model)	Complex Activities: shopping, showering, eating, working, commuting, and brushing teeth
Kikhia et al. (2014) [62]	wrist, chest, thigh	controlled	10	mean, standard deviation, magnitude, skewness, kurtosis, energy, fast Fourier transform (FFT)	accuracy: 85% (Random Forest)	Static & Dynamic Activities: walking, running, ascending/descending stairs, sitting, lying and standing, getting dressed, cleaning, and cooking; Body Movement Effort: strong, light, sudden

Continued on next page

Table 2.2 – *Continued from previous page*

Authors	**Placement of Accelerometer**	**Experiment**	**Number of Subjects**	**Features**	**Classification Performance (%)**	**Detected Activities**
Arif and Kattan (2015) [21]	wrist, chest, ankle	controlled	9	mean absolute value, harmonic mean, variance, kurtosis, skewness, root mean square, cumulative length, zero crossing rate, Willison amplitude, and correlation coefficient	F-measure: 98% (Rotation Forest)	Static & Dynamic Activities: lying down, sitting, standing, walking, running, cycling, Nordic walking, ascending/descending stairs, and jumping rope; Household Activities: vacuum cleaning, and ironing clothes
Gjoreski et al. (2016) [40]	chest, right wrist, thigh, and ankle	controlled	10	mean, total mean, area, variance, skewness, kurtosis, quartiles, correlation between axes, quartile range, absolute area, total absolute area, total magnitude, and amplitude	accuracy: 75% (Random Forest)	Static & Dynamic Activities: lying, standing, walking, sitting, cycling, kneeling, running, bending, and transition
Xu et al. (2016) [61]	wrist, chest, ankle	controlled	9	mean, variance, fast Fourier transform (FFT), and wavelet transform (WT)	accuracy: 93.77%, F-measure: 93.53% (Back Propagation Neural Network)	Static & Dynamic Activities: running, walking, cycling, Nordic walking, lying, sitting, standing, ascending/descending stairs, and rope jumping; Household Activities: vacuum cleaning, and ironing clothes

Continued on next page

Table 2.2 – *Continued from previous page*

Authors	Placement of Accelerometer	Experiment	Number of Subjects	Features	Classification Performance (%)	Detected Activities
Pannurat et al. (2017) [38]	chest, wrist, waist, thigh, head, upper arm, and ankle	controlled	12	mean, standard deviation, maximum, minimum, standard deviation magnitude, correlation between axes	varies per sensor: Thigh - accuracy: 99% (Nearest Neighbor) Chest - accuracy: 98.5% (Naïve Bayes) Wrist - accuracy: 80.6% Nearest Neighbor)	Static & Dynamic Activities: sitting, lying on the left/right/back/front side, standing, and walking
this study	***chest and wrist***	***controlled and uncontrolled***	***40***	***52 features***	***varies per sensor (Random Forest)***	***13 Static & Dynamic Activities:*** *sitting, dynamic sitting, lying with face up and down, lying to the left and right, standing, dynamic standing, walking upstairs and downstairs, walking, running, cycling*

3

Methodology

This chapter describes a sequence of steps for developing and evaluating an activity recognition system (see Figure 3.1). These steps include data acquisition, signal preprocessing, signal segmentation, feature extraction/selection and classification. Thus, different state-of-the-art machine learning techniques are presented in order to enhance the classification performance. Specifically, the data collection phase, based on the investigational protocol is described. Furthermore, the proposed methodology for data analysis, including data processing techniques and classification models, is elaborated.

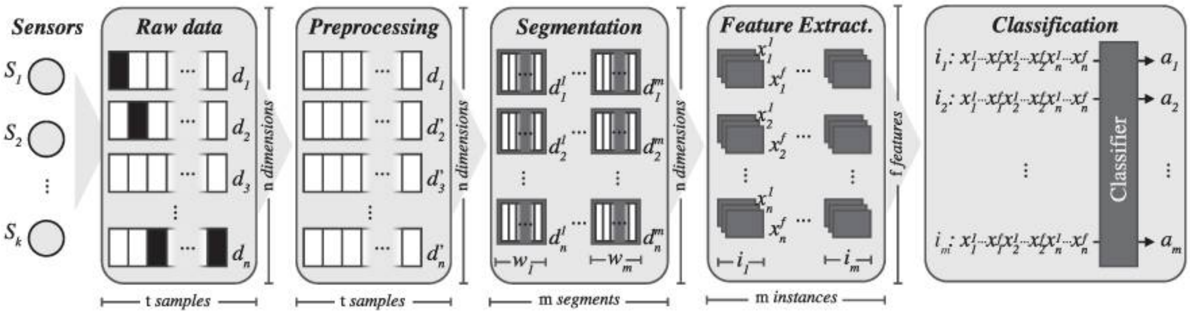


Figure 3.1: Steps involved in activity recognition systems [4].

3.1. Data Acquisition

Triaxial accelerometer data (x-, y-, and z-axes) were collected using two IMEC wearable sensors, one worn on the subjects' wrist and one on the subjects' chest, both firmly attached to the body. According to Attal et al. [36], sensors not firmly attached to the body could lead to a small tilt or misplacement of the sensor, resulting in recording data with large variation. The decision of using exactly two wearable accelerometers and their specific placement on the body was considered according to the literature review but also based on the specific activities that needed to be detected.

The first sensor 'IMEC Health Patch' is a chest patch for the measurement of triaxial acceleration on the chest. For this study, the abbreviation 'SR' will also be used regarding the chest patch sensor. The sensor is designed to be worn on the left chest, close to the sternum and three fingers below the collarbone. The SR consists of a patch and a sensor node, which is designed to record acceleration at 32Hz for seven successive days, continuously. Additionally, electrocardiogram (ECG) signals are recorded, but they are not considered for this study. Data are stored on an internal SD card and uploaded on a central data platform at the end of the experiment. The patch is designed by Delta in Denmark based on a 3M adhesive. It is designed to be biocompatible for measurements up to 7 days. Additionally, for optimal adhesiveness, men were asked to shave or trim their chest the night before the start of the experiment.

The second sensor 'IMEC Chillband' is a wristband, which is designed to record acceleration at 32Hz on the wrist. For this study, the abbreviation 'CB' will also be used regarding the wristband sensor. Additionally, Galvanic Skin Response (GSR) and skin temperature are recorded. Similar to the chest sensor, only accelerometer data are considered for this study, which are stored on an SD card and uploaded at the end of the recording on the central data platform. Participants were asked to wear this sensor during the experiment on their dominant hand.

Both sensors have a capacitive MEMS accelerometer that provides acceleration in correlation to the differential capacitance. The chest patch weights 20g and the Chillband weights 55g. The sampling frequency of the accelerometer is 32Hz, and the dynamic range is ±2g. Before the start of every recording, the sensors' Real Time Clock (RTC) is synchronized to the PC clock.

Additionally, a wearable camera GoPro Hero4 with an extended battery is used for the experiment. The camera was placed on a strap, which was worn on the chest. The camera was adjusted to time-lapse recordings every thirty seconds with 5MP resolution, while the total battery lifetime ensured eight hours of continuous recording. The camera is used for annotating the performed activities during the Free-Living Study, which will be further explained in the next sections.

The three aforementioned sensing devices are depicted in the following figure (see Figure 3.2).

Imec Health Patch (SR)	Imec Chillband (CB)	GoPro Hero4
		
3-axial ACC at 32Hz, using a capacitive MEMS accelerometer	**3-axial ACC** at 32Hz, using a capacitive MEMS accelerometer	5MP picture every 30"
Weight: 20g Long autonomy: 7+ days Storage: internal SD card with capacity up to 30 days of data Biocompatible: made by hypoallergenic materials	Weight: 55g Long autonomy: 7+ days Storage: internal SD card with capacity up to 30 days of data Biocompatible: made by hypoallergenic materials	Weight: ~200g Long autonomy: ~8 hours Storage: 64GB SD card

Figure 3.2: An overview of the selected sensing devices and their specifications.

3.1.1. EXPERIMENTAL PROTOCOL

During the experimental phase, 40 subjects, consisting of 24 males and 16 females and aged from 19 to 45 years old (average age 26.9 ±5.77 years), were recruited voluntarily through emails. Two of the recruited subjects were left-handed, and thus, they wore the Chillband sensor on their left wrist. The admission criteria included employees, interns or other people with a contract at IMEC. During the recruitment phase, all the participants were informed about the experimental procedure and were asked to sign an informed consent form. Exclusion criteria involved subjects with acute health problems (e.g., heart failure) or subjects with physical disabilities (being not able to perform all the activities). Additionally, recruited subjects had the opportunity to stop the experiment at any given time during its course.

Concerning the measurement protocol for activity recognition, the recruited subjects were asked to complete the experiment in two parts. The first part was accomplished under controlled conditions (laboratory experiment), where the subjects were asked to perform thirteen consecutive activities, following the specific instructions given by the researcher. On the other hand, the second part was performed under uncontrolled conditions (free-living experiment), where the subjects did not receive any specific instructions.

For the Controlled Study, forty recruited subjects wore the wristband and the chest patch and were asked to perform a series of thirteen consecutive activities based on the following order:

1. static sitting on a chair,
2. dynamic sitting; including tasks like working on a computer and handwriting,
3. lying with face up,
4. lying with face down,
5. lying to the left side,
6. lying to the right side,
7. static standing,
8. dynamic standing; including tasks like using a mobile phone and writing on a whiteboard,
9. walking upstairs,
10. walking downstairs,
11. walking,
12. running,
13. cycling.

Each activity was performed for four minutes. The dynamic activities, running, and cycling, were performed outdoors. Instructions about the intensity of the dynamic activities, the hand movements or the way of using a mobile phone were not provided. Thus, participants performed all the activities at their own pace and in the most realistic way. However, if a subject struggled to complete any of the aforementioned activities, such as running or walking upstairs/downstairs, these activities were divided into lower time sessions with some repetitions to reach the total mentioned time per activity (four minutes). It is worth mentioning that the researcher was present during the whole recording time in order to annotate the performed activities, but also to provide each participant with the necessary guidance and supervision. At the end of the first part of the experiment, the participants were asked to remove and return the sensors.

For the Free-Living Study, thirty-seven subjects accomplished successfully this part. From the forty recruited subjects, three were unavailable to perform the second part of the experiment. Each participant wore the wristband and the chest patch, as well as the wearable camera, for around eight hours (normal office hours). The subjects were free to perform the activities as usual. Once the experiment had successfully ended, all the sensors and the camera were returned to the researcher. The researcher was not present during this part of the experiment. However, the performed activities were manually annotated by the researcher in a later phase, based on the recorded pictures.

The experiment lasted two days and the total duration was approximately 10 hours. Specifically, the duration was around 2 hours for the first part (Controlled Study) and 8 hours for the second part (Free-Living Study).

3.1.2. DATA PRIVACY

Respecting the privacy issues that may arise from wearing the camera and recording participant's private life, subjects were allowed to remove the camera when they considered it crucial (for example during the time of visiting the restroom) and wore it again afterward. Moreover, subjects were asked to browse through all the taken photos at the end of the experiment and delete those that could be crucial for their privacy, before the researcher assessed the photos. This part was important to ensure that each participant was aware of the provided photos for the activity annotation, and avoid the case of sensitive, embarrassing or private images. Furthermore, it is important to mention that the pictures were assessed only by the researcher. The process of assessing the images was done manually, using a desktop computer. Photos were used for annotating the physical activities in the free-living study. Once the activity annotation was completed, the researcher deleted the recorded photos.

Additionally, the GoPro camera was worn on a chest mount (strap) and was angled at 45 degrees (the camera lens focused mostly on the ground surface) in order to avoid the recording of other employees' faces who could be captured in the photos. Concerning the resolution of these photos, the photos were taken at the minimum resolution, which was 5MP based on the GoPro settings. However, all the photos were stored on a compressed file, where the resolution was downgraded from 2560x1920 to 600x450.

Concerning the acquired data from the wearables, several measures have also been considered to secure the privacy of the participants. The first measure was to use a unique identification code for each subject and not identities. Thus, no personal information was used to link sensors to surveys and annotations. The second measure ensured that the collected data would be stored in a secure data storage, gaining access only to the authorized researcher.

3.1.3. ACTIVITIES ANNOTATION

Different supervised classification techniques are used to make an accurate model for activity recognition. An important aspect of the supervised classification, however, is the use of labeled data related to real annotated activities.

For the Controlled Study, activities were manually annotated by the researcher during the recording session. In total, 52 minutes of continuous recording were annotated for each subject. Thus, each activity was annotated for 160 minutes for all subjects, resulting to a balanced dataset of 9600 samples (160 * 60 = 9600 seconds) per activity.

For the Free-Living Study, the researcher annotated the performed activities manually based on the pictures (taken from the wearable camera). Most of the subjects performed this part of the experiment during their working time at IMEC, from around 9 am to 5 pm. Approximately, 480 minutes (around eight hours) of continuously recording were annotated respectively, based on 960 taken pictures per subject (a picture was taken every 30 seconds). Due to the environment limitations, subjects did not perform the lying activities (with face up, face down, to the left and to the right). Furthermore, activities such as running and cycling were performed only by a few subjects, leading to an unbalanced dataset for the free-living study (see Figure 3.3). Consequently, this problem of imbalanced data should be addressed respectively in the later phase of classification.

Another significant limitation of the free-living annotation was the case where the taken pictures from the camera were insufficient for the activities annotation. Indeed, camera's pictures were sufficient for activities annotation such as sitting, standing, walking, ascending and descending stairs, cycling and running. However, activities such as sitting and dynamic sitting, or standing and dynamic standing were not always annotated sufficiently (see Figure 3.4). For instance, the camera lens was placed very close to the subject's desk, and despite the obvious activity, which was sitting, it could not be significantly differentiated from static sitting to dynamic sitting (e.g., working on computer or handwriting a document). Additionally, pictures were recorded every thirty seconds. For instance, if a subject was sitting and standing for multiple times during the thirty seconds, the activity annotation was based on the taken picture, which did not include all the performed activities.

It is worth mentioning that all the activities performed in the Free-Living Study were annotated based on the 13 completed activities from the Controlled Study. For instance, activities such as driving, eating (on a table) were annotated as dynamic sitting, while activities such as presenting a speech, writing on a whiteboard and standing on an elevator were annotated as dynamic standing.

An overview of all the performed activities for the Controlled and Free-Living Study is presented in Table 3.1.

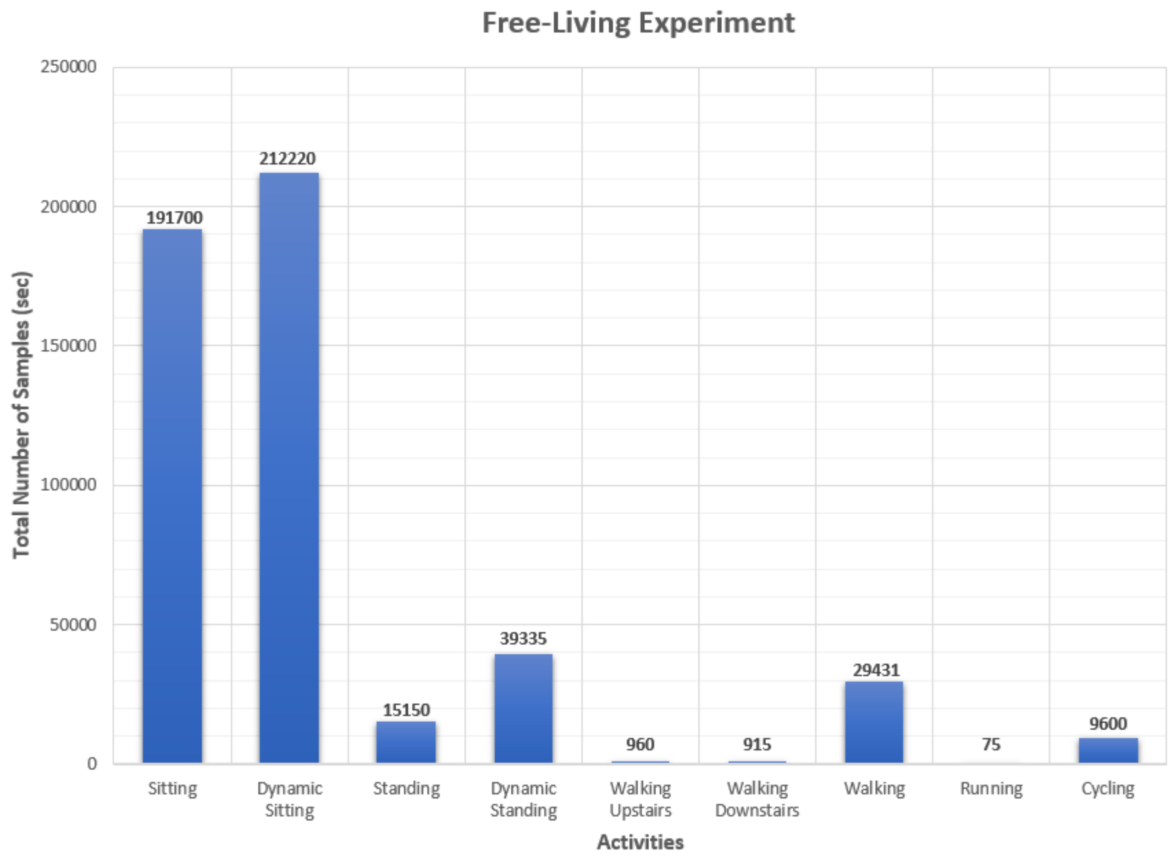


Figure 3.3: An overview of the collected data, performed by 37 subjects, in the Free-Living Study. Each type of activity consists of a different number of samples, while activities such as walking upstairs, walking downstairs and running were sparsely performed.

Figure 3.4: Annotation for sitting activities; (top left) static sitting, (top right) static sitting, (bottom left) dynamic sitting - typing on keyboard, (bottom right) dynamic sitting -texting on a smartphone.

Table 3.1: An overview of the performed activities for each subject. All activities as mentioned as (1) static sitting, (2) dynamic sitting, (3) lying with face up, (4) lying with face down, (5) lying to the left, (6) lying to the right, (7) static standing, (8) dynamic standing, (9) walking upstairs, (10) walking downstairs, (11) walking, (12) running and (13) cycling.

Participants	Controlled Study		Free-Living Study	
	Activities	Time (min)	Activities	Time (min)
Subject01	all	52	1, 2, 7, 8, 9, 10, 11, 13	453 (~7.5 hours)
Subject02	all	52	1, 2, 7, 8, 9, 11, 13	418 (~7 hours)
Subject03	all	52	1, 2, 7, 8, 9, 10, 11, 13	477 (~8 hours)
Subject04	all	52	1, 2, 7, 8, 9, 10, 11, 13	446 (~7.5 hours)
Subject05	all	52	1,2, 7, 8, 9, 10, 11, 12	493 (~8 hours)
Subject06	all	52	1, 2, 7, 8, 9, 10, 11, 13	475 (~8 hours)
Subject07	all	52	1, 2, 7, 8, 9, 10, 11, 13	475 (~8 hours)
Subject08	all	52	1, 2, 7, 8, 11, 13	396 (~6.5 hours)
Subject09	all	52	1, 2, 7, 8, 9, 10, 11, 13	513 (~8.5 hours)
Subject10	all	52	1, 2, 7, 8, 10, 11	456 (~7.5 hours)
Subject11	all	52	1, 2, 7, 8, 9, 10, 11, 13	386 (~6.5 hours)
Subject12	all	52	1, 2, 7, 8, 9, 11	440 (~7.5 hours)
Subject13	all	52	1, 2, 7, 8, 9, 10, 11, 13	525 (~9 hours)
Subject14	all	52	1, 2, 7, 8, 9, 10, 11, 13	462 (~8 hours)
Subject15	all	52	1, 2, 7, 8, 9, 10, 11, 13	466 (~8 hours)
Subject16	all	52	1, 2, 8, 11	407 (~7 hours)
Subject17	all	52	1, 2, 7, 8, 9, 11	483 (~8 hours)
Subject18	all	52	1, 2, 7, 8, 9, 10, 11, 13	462 (~8 hours)
Subject19	all	52	2, 7, 8, 11, 13	89 (~1.5 hours)
Subject20	all	52	1, 2, 7, 8, 11, 13	359 (~6 hours)
Subject21	all	52	1, 2, 7, 8, 11, 13	493 (~8 hours)
Subject22	all	52	-	-
Subject23	all	52	1, 2, 7, 8, 9, 10, 11, 13	418 (~7 hours)
Subject24	all	52	1, 2, 7, 8, 9, 10, 11	468 (~8 hours)
Subject25	all	52	1, 2, 7, 8, 9, 10, 11	478 (~8 hours)
Subject26	all	52	1, 2, 7, 8, 9, 10, 11	467 (~8 hours)
Subject27	all	52	1, 2, 7, 8, 9, 10, 11	466 (~8 hours)
Subject28	all	52	1, 2, 7, 8, 9, 10, 11, 13	482 (~8 hours)
Subject29	all	52	-	-
Subject30	all	52	1, 2, 7, 8, 9, 11	465 (~8 hours)
Subject31	all	52	1, 2, 7, 8, 9, 10, 11	416 (~7 hours)
Subject32	all	52	1, 2, 7, 8, 9, 10, 11, 13	442 (~7.5 hours)
Subject33	all	52	1, 2, 7, 8, 9, 10, 11, 13	464 (~8 hours)
Subject34	all	52	1, 2, 8, 11, 13	472 (~8 hours)
Subject35	all	52	1, 2, 8, 9, 10, 11, 13	485 (~8 hours)
Subject36	all	52	1, 2, 8, 9, 10, 11, 13	485 (~8 hours)
Subject37	all	52	1, 2, 7, 8, 9, 10, 11	543 (~9 hours)
Subject38	all	52	-	-
Subject39	all	52	1, 2, 8, 9, 10, 11	495 (~8.5 hours)
Subject40	all	52	1, 2, 8, 9, 11	436 (~7.5 hours)

3.2. DATA PROCESSING

After data collection, raw data are processed in order to extract relevant features. This phase includes different steps, such as signal preprocessing, segmentation, feature extraction, and selection. The data processing, as well as the whole procedure of analyzing data in this study, is performed using the programming language Python (version 2.7) and particularly the python library 'scikit-learn' [71].

3.2.1. SIGNAL PREPROCESSING

Signal preprocessing is the first important step in data analysis. Different actions are involved in this phase in order to synchronize the acquired data from the wristband and the chest sensor, based on their timestamps, but also to filter and remove data that do not correspond to the performed activities from subjects (data not related to the labeled activities were removed).

3.2.2. SIGNAL SEGMENTATION

In order to extract useful information from the collected data, it is important to divide the raw data into smaller segments using the sliding window approach with overlap. This approach is based on a fixed size window that moves across the sensor-stream data (see Figure 3.5). By decreasing the window size, a faster activity detection can be achieved with low computational cost [61]. However, the window segment may not contain the complete cycle of the performed activity. In contrast, increasing the window size results in the detection of more complex activities, with the drawback of increasing the computational cost, too. Thus, increasing the window size is not recommended for real-time applications. According to literature, the most commonly used window segments in activity recognition systems, are based on smaller window sizes with a fixed 50% overlap [5, 58, 65]. Thus, eight different window segments with 50% overlap were evaluated and presented in Figure 3.6. In total, thirteen window segments with varying window size and overlap were investigated and are elaborated in the next chapter (see chapter 4).

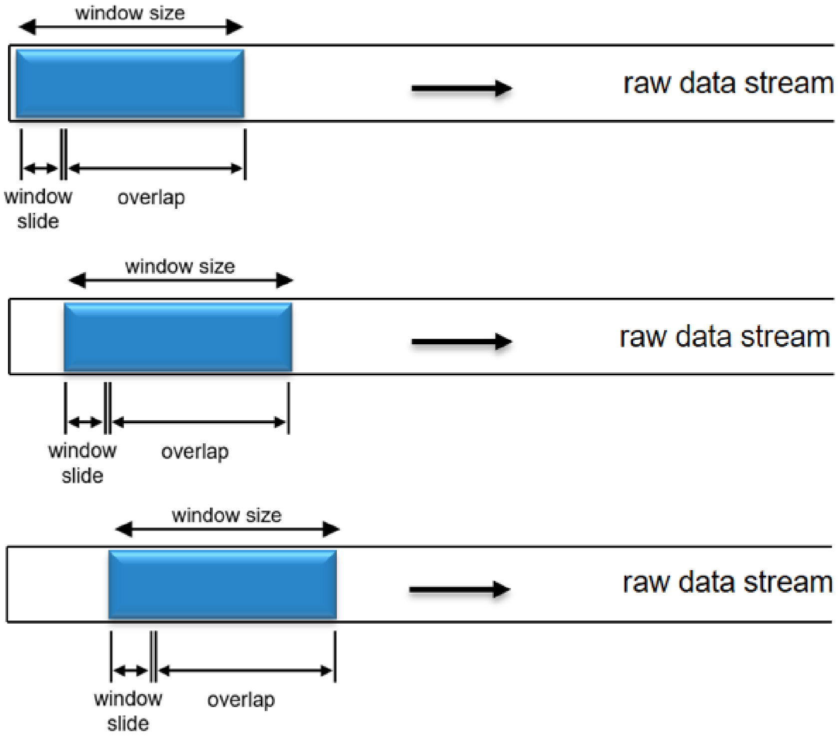


Figure 3.5: Signal segmentation based on a fixed size window with overlap.

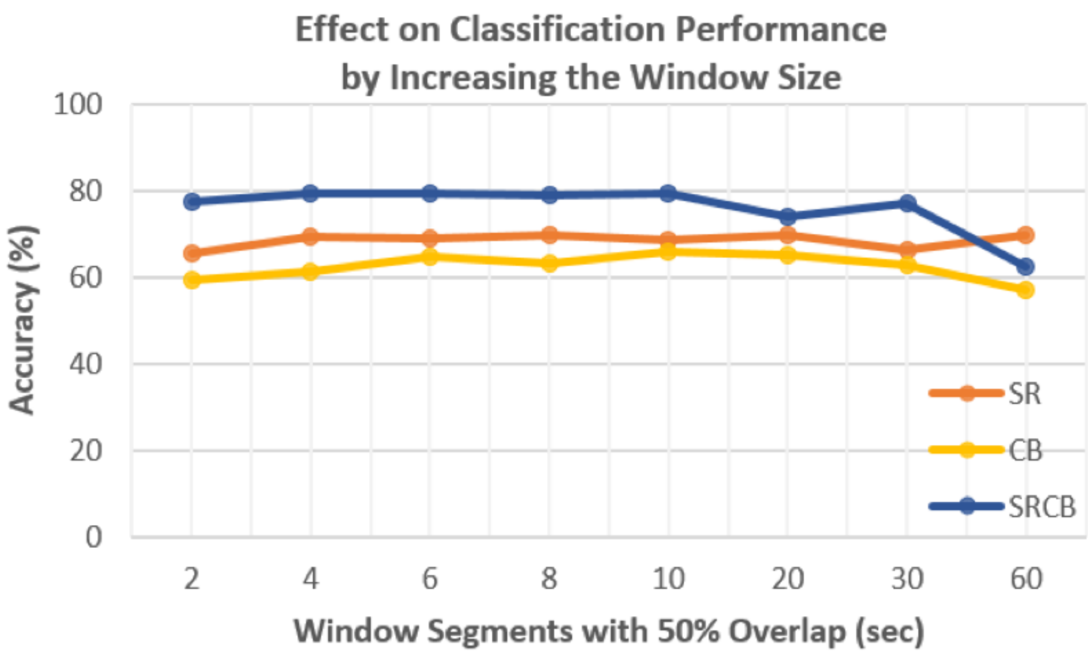


Figure 3.6: Different window sizes with 50% overlap are presented for three sensors; SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

3.2.3. FEATURE EXTRACTION

The process of transforming large input raw data into a reduced set of features is called feature extraction. The purpose of this processing step is to extract the most important characteristics of a window segment, by representing accurately the raw data stream. Two of the most widely used types for calculating features in activity recognition systems are time and frequency domain features. Most of the studies calculate time domain features, such as mean, variance and standard deviation [72], since frequency domain features require extra Fourier transform calculation, resulting to extra computational complexity [73].

In this study, 52 features are calculated based on time and frequency domain and are summarized in Table 3.2. Hence, 52 features are extracted for the Stingray (SR) dataset, 52 features for the Chillband (CB) dataset, and 104 for the Stingray- Chillband (SRCB) dataset. Features related to body posture measurements are represented by the mean, median, area, and meandistance; features related to motion shape measurements are represented by absmean, absarea, magnitude, entropy, skewness and kurtosis; features related to motion variation measurements are represented by variance, std, amplituderange and interquartilerange; and features related to motion spectral content measurements are represented by signalpower and fftfreq [74].

3.2.4. FEATURE SELECTION

Once features are extracted, the phase of feature selection aims to filter the feature set and remove any redundant features, reducing the dimensionality and improving the overall classification performance. The use of redundant or inappropriate features could decrease the classification performance. Thus, feature selection is defined as a process of searching a subset of informative features from the original feature set, resulting to less computation time and complexity during the classification phase, while enhancing the estimators' performance [36].

Table 3.2: Overview of the calculated features.

Feature Name	Number	Domain	Description
mean (x,y,z)	3	time	computes the mean; the average of all sample values in a sample window [73]
absmean (x,y,z)	3	time	computes the absolute mean; the absolute average of all sample values in a sample window [73]
median (x,y,z)	3	time	computes the median; the value that divides the higher half from the lower half in a sample window [73]
variance (x,y,z)	3	time	computes the variance; the average of the squared differences of the sample values from the mean in a sample window [73]
std (x,y,z)	3	time	computes the standard deviation; the square root of variance in a sample window [73]
max (x,y,z)	3	time	computes the min; the lowest number of all sample values in a sample window [36]
min (x,y,z)	3	time	computes the max; the highest number of all sample values in a sample window [36]
magnitude	1	time	computes the magnitude; by adding each one of the squared axes in a sample window, and calculating the square root of the sum [74]
skewness (x,y,z)	3	time	computes the skewness; the asymmetry of the distribution of the sample values around the mean in a sample window [74]
kurtosis (x,y,z)	3	time	computes the kurtosis; the shape description of the distribution of the sample values in a sample window [74]
meandistance (x-y, x-z, y-z)	3	time	computes the mean distance; the differences between the mean values of the x-y, x-z and y-z in a sample window [74]
amplituderange (x,y,z)	3	time	computes the amplitude range; the difference between the maximum and minimum sample values in a sample window [74]
interquartilerange (x,y,z)	3	time	computes inter quartile range; the difference between quartiles[1] Q3 and Q1, and describes the dispersion of the acceleration signal [74]
area (x,y,z)	3	time	computes the area; the sum of the sample values in a sample window [74]
absarea (x,y,z)	3	time	computes the absolute area; the sum of the absolute sample values in a sample window [74]
entropy (x,y,z)	3	frequency	computes the entropy; the degree of distortion in a sample window (discriminate activities that have the same Power Spectral Density[2] but different patterns of movement) [36]
signalpower (x,y,z)	3	frequency	computes the signal power; the non-normalized sum of the Power Spectral Density in a sample window [36]
fftfreq (x,y,z)	3	frequency	computes the Fast Fourier Transform peaks; the frequency related to the highest computed Power Spectral Density in a sample window [36]

[1] Quartiles are calculated by partitioning the sample values of a sample window into four quarters, each one contains 25% of data, with Q1=25%, Q2=50% and Q3=75% [72].

[2] Power Spectral Density (PSD): the squared sum of its spectral coefficients, normalized by the number of the window slide [36].

By transforming high-dimensional data into a meaningful representation of lower dimensionality, feature selection phase provides three main benefits:

- enhanced accuracy: misleading features are removed, resulting to a better classification performance,
- less training time: training based on fewer data results to a faster learning model, and
- less overfitting: redundant features are removed, resulting in less prediction due to noise.

In this study, feature selection is achieved through a two-phase process. At first, highly correlated features are removed based on the Pearson's correlation coefficient, which measures a linear correlation between features. For each feature, the correlation coefficients are calculated and ranked according to the other features, starting from the lowest graded feature. Hence, if the selected feature has a correlation coefficient higher than the threshold (the threshold is equivalent to the absolute value of 0.80) with at least one feature, then the features are removed. For instance, the 'mean' features are correlated with the 'median' features, and thus, only one of these two is kept.

Secondly, non-informative features, with low information gain, are removed. Each feature is ranked based on the gain ratio, which is measured respecting the contribution of each feature to the accurate prediction. For this phase, low information gain is calculated based on four main methods; removing features with low variance, removing features based on univariate statistical tests, removing features recursively and removing unnecessary features [75].

'Removing features with low variance' is based on features with identical or almost the same value in all samples. 'Univariate feature selection' keeps features that contribute the most to the target variable, using the classification estimator f_classif to calculate univariate scores and p-values. 'Recursive feature elimination' assigns weights to the features and recursively removes the attributes that do not contribute to the target attribute's prediction. 'Feature selection using SelectFromModel' keeps highly important features either using linear estimators, such as linear SVC (Support Vector Classifier) or using tree-based estimators, such as Random Forest Classifier. These techniques are evaluated and presented in the following chapter.

3.3. CLASSIFICATION

After data processing, the dataset is divided into training and test set. Supervised classification techniques are used to build a learning model based on the training set and detect the labels (activities) accurately on unseen data. During the learning phase, different supervised algorithms are evaluated in order to select the classifier that contributes the most to the model's performance. In this study, four classification algorithms will be investigated. Each algorithm uses a different classification method related to either distance-based approach (e.g., k-nearest neighbors), or statistical approach (e.g., naïve Bayes), or kernel approach (e.g., SVC) or decision-tree based approach (e.g., as Random Forest).

kNN (k-nearest neighbors) algorithm is based on the assumption that data can be grouped into different classes, respecting their similarities and their geometric properties [4]. This algorithm detects the class labels by measuring the distance from a new instance to the instances of the training set, in order to assign the new instance to the class label that is closest to the k neighboring instances [40].

Naïve Bayes is a probabilistic classifier that uses estimated conditional probabilities from the training set to calculate posterior probabilities and assign the highest one to every class on the test set. Following a naïve assumption that all features are conditionally independent of each other, it applies the Bayes' theorem to detect the class labels [4, 40]. However, this assumption is not always sufficient due to the high correlation among features. For instance, acceleration data produce highly correlated features that may affect the classifier's performance [5].

Support Vector Machines (SVM) algorithms, which are also known as Support Vector Classifiers (SVC), depend on kernel functions to project all instances to a higher dimensional space. Trying to define decision boundaries and detect class labels, they produce hyperplanes that categorize data into different parts [40].

Random Forest classifier follows the main principle of Decision-Trees algorithms using a sequence of decisions to classify labels, in which attributes are equivalent to edges and nodes, while branches represent feature values and leaves represent class labels [4, 5]. Random Forest consists of several weighted decision trees, randomly constructed from different subject of features, combining the majority of the various predictions made by the decision trees to predict the class labels [40].

3.4. EVALUATION

Our model for activity recognition is evaluated based on the Leave-One-Subject-Out Cross Validation (LOSOCV). This technique is used in order to avoid overfitting, by excluding data from subjects that were used for training the classifier (learning model), and include only unseen data for validating our model. In order to measure the classification performance, different metrics are used. The most commonly used metrics in activity recognition problems are the classification accuracy, F-score, precision (also known as positive predictive value), recall (also known as sensitivity, hit rate, or true positive rate), specificity (also known as true negative rate) and the Cohen's kappa. Additionally, F1-score can be calculated as a weighted average of precision and recall and refers to the balanced F-score.

These metrics can be easily calculated through a confusion matrix. A confusion matrix is a summary of correct and incorrect predictions classified for each class, compared to the actual labels. Specifically, the elements at the main diagonal represent correct classifications, while the other elements represent incorrect classifications. A confusion matrix consists of true positives (correct classifications of positive examples), true negatives (correct classifications of negative examples), false positives (incorrect classifications of negative examples to positive class), and false negatives (incorrect classifications of positive examples to negative class). In a multi-class problem, the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) can be calculated based on Figure 3.7.

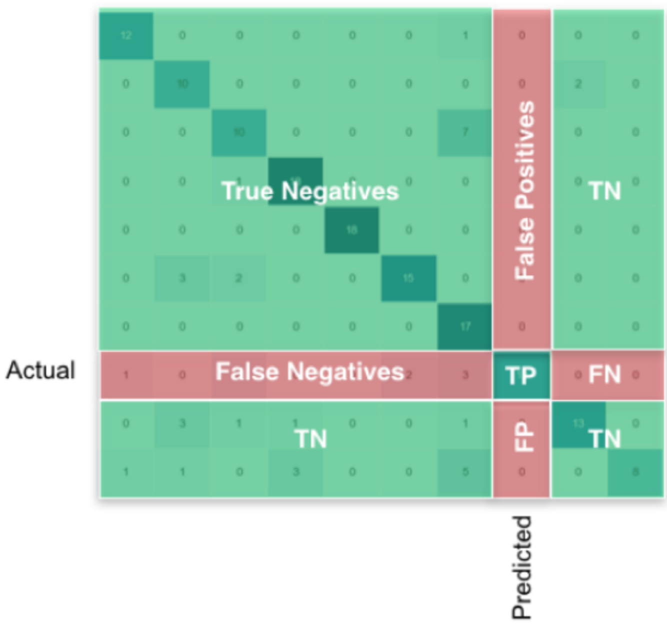


Figure 3.7: Example of a confusion matrix [76].

For this study, the following metrics are calculated:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

$$F1 - score = 2 * \frac{precision * recall}{precision + recall} \tag{3.4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.5}$$

$$BalancedAccuracy = \frac{Sensitivity + Specificity}{2} \tag{3.6}$$

$$Cohen'skappa = \frac{P_o - P_e}{1 - P_e} \tag{3.7}$$

where:

P_o = accuracy
$P_e = \frac{(TP+FN)*(TP+FP)+(FP+TN)*(FN+TN)}{(TP+TN+FP+FN)^2}$

Cohen's kappa refers to a score that measures the level of agreement between two annotators on a classification problem. This score is recommended as a metric, especially, for unbalanced data [77, 78]. Po represents the probability of overall agreement, over the label assignments between the classifier and the real labels, and is equal to accuracy. Pe represents the chance agreement over the random assignment of labels. Pe is defined as the sum of the proportion of examples assigned to a class times the proportion of true labels of that class in the data set.

On balanced data, all the metrics mentioned above perform significantly well and are essential for the evaluation of the classification model. However, on imbalanced data, some of these metrics perform poorly due to their dependence on how rare some labels are. For instance, classification accuracy can be a misleading metric on the free-living study (imbalanced dataset) when an unusual activity (e.g., running) appears in 1% of the test set and a trivial classifier always declines the prediction of this activity. In that case, the classification accuracy will misleadingly be 99%. Additionally, averaging precision over many labels (e.g., 13 activities) can be misleading when specific activities unfairly dominate the score [79].

Trying to overcome the problem of the imbalanced dataset, the balanced accuracy, instead of the classification accuracy, will be used for models' evaluation. It will be calculated through the confusion matrixes. Furthermore, the weighted F1-score, precision, and recall from sklearn [80] will be used to calculate the score for each label and find their average, weighted by support (the number of true instances for each label), and respecting the label's imbalance.

3.5. DISCUSSION

This chapter elaborates our methodology for collecting and analyzing data, focusing on developing a model for activity recognition, in order to answer the research question. A further explanation of these methods, combined with the results from the Controlled and Free-Living Study, is presented in the chapter 4 & chapter 5.

It is worth mentioning that the answer to our research question can be given without performing power or sensitivity analysis (in order to calculate the recommended number of participants). Overall, activity recognition, as a classification problem, demands a sufficient amount of data in order to train the classifier. Thus, our decision to recruit and collect data from forty subjects exceeds the current trends in activity recognition systems, where few participants are recruited, and focuses on enhancing the classification performance. In the following figure (Figure 4.1), we present the impact of the number of subjects (training set) on the overall classification performance, and we show that an increase of the training set enhances the performance in activity recognition.

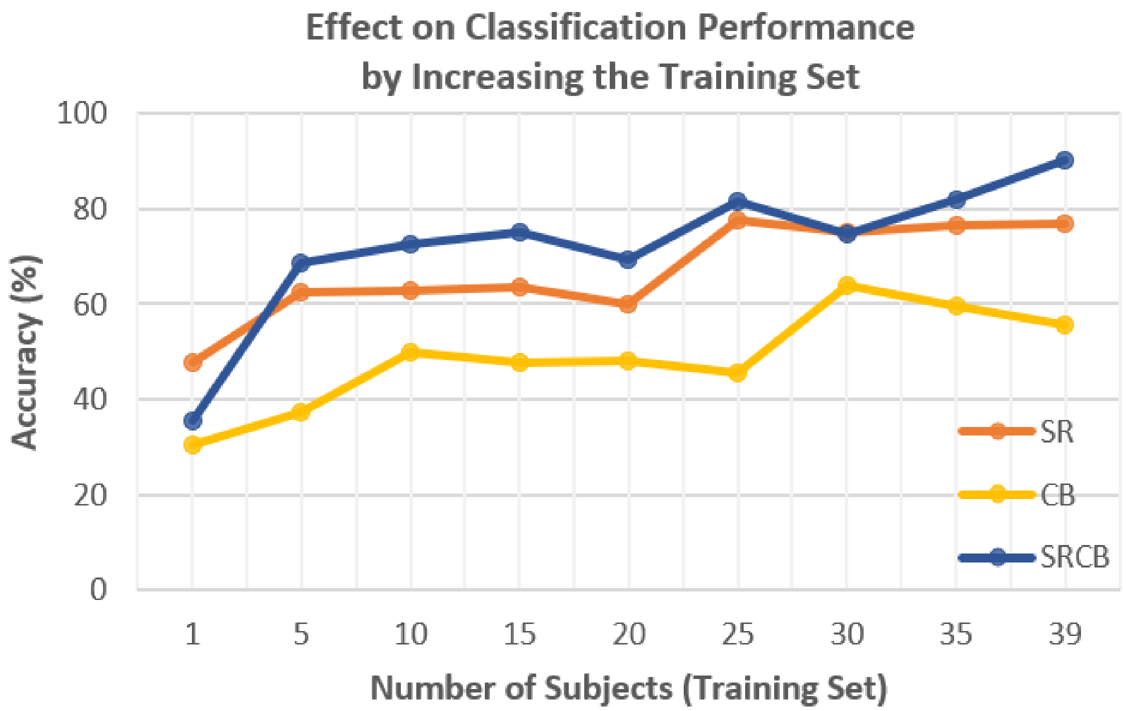


Figure 3.8: The effect on classification performance by increasing the training set. The classifier is trained for detecting 13 activities based on the Controlled Study for a different number of subjects. Overall, 39 subjects are used for the training set, while the last subject is used for the test set. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

4

CONTROLLED STUDY

In this chapter, the dataset from the Controlled Study will be investigated in order to implement a classification model for activity recognition systems in a controlled (laboratory) environment. This dataset consists of raw accelerometer data from 40 subjects, in a controlled environment, performing 13 different activities. The acquired accelerometer dataset includes data from the stingray (SR) sensor, the chillband (CB) sensor and the combination of these two stingray-chillband sensors (SRCB), and are investigated separately.

In order to evaluate our activity recognition system, data from 15 subjects are used for optimizing the learning model, while data from the other 25 subjects are used to validate the classification performance. In addition to recognizing 13 activities, we will also evaluate the model for predicting 7 main activities, including sitting (both static and dynamic sitting), standing (both static and dynamic standing), lying (with face up, face down, to the left and to the right side), walking on stairs (both ascending and descending stairs), walking, cycling and running.

4.1. LEARNING MODEL

Data from 15 subjects are analyzed, focusing on enhancing the performance of the classification model. These subjects will be used in order to define the window segment, the features, and the classification algorithm. Based on these parameters of the learning model, the classification model will be validated for predicting the activities accurately on unseen data (the other 25 subjects). For the evaluation of the learning model, two different approaches will be investigated. The first is based on predicting 13 activities and the second on predicting 7 activities.

4.1.1. DATA PROCESSING

Since different activities have different periodic signals, the procedure of determining the best window segment is challenging. Trying to define the most optimal window segment for the current study, different window sizes with different window slides were evaluated. In total, thirteen different window segments were evaluated and are presented in the following figure. The validation was performed on fifteen subjects of the controlled study, using Random Forest as the default classifier, and based on the Leave-One-Subject-Out Cross Validation (LOSOCV). The best classification accuracy was found for the window size of 5 seconds with 1 second window slide and 4 seconds overlap.

It can be seen in Figure 4.1 that classification performance is significant the same for more window segments, with varying window sizes from 4 to 10 seconds. The decision of selecting the window segment of 5 seconds length with 1 second step (which is equal to 4 seconds overlap), was based on literature where smaller window sizes are recommended, but also on the fact that the selected window segment performed better when the predicted activities were limited to 7, including sitting, standing, walking, walking on stairs, lying, running and cycling.

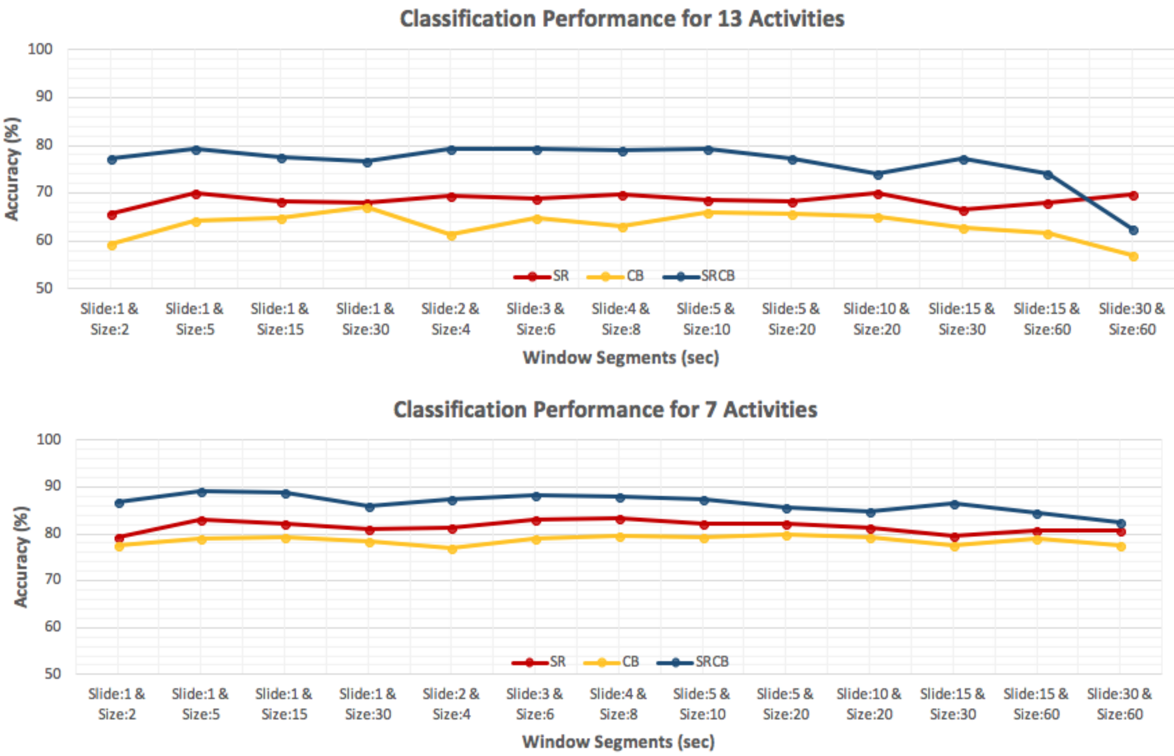


Figure 4.1: Thirteen window segments, with different window sizes and different overlap, are evaluated for the SR, CB and SRCB datasets. The top figure represents the classification performance for predicting 13 activities, while the bottom figure represents the classification performance for predicting 7 activities. The validation is based on LOSOCV for 15 subjects (detect 13 activities based on the Controlled Study). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

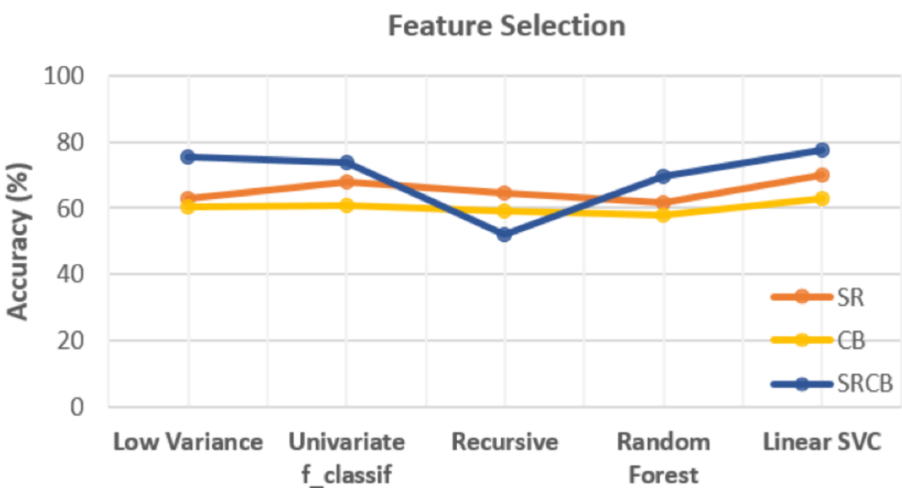


Figure 4.2: An overview of the methods used for the second phase of feature selection. These methods include removing features with low variance, removing features based on univariate statistical tests, removing features recursively and removing unimportant features. The validation is based on LOSOCV for 15 subjects (detect 13 activities based on the Controlled Study). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

After calculating and extracting the features based on the selected window segment, the phase of feature selection takes place. This step aims to filter the feature set and remove any redundant features. Specifically, highly correlated features are removed based on the Pearson's correlation coefficient, while features with low information gain are subsequently removed. For calculating low information gain, five methods are examined and presented in Figure 4.2. It can be seen through this figure that Linear SVC estimator gives the best classification accuracy, and thus, it is preferred for the second phase of feature selection. It is worth mentioning that linear SVC, similar to support vector machines (SVM), might perform better if the features have roughly the same magnitude. Thus, all the features were normalized per dataset for enhancing the classification performance.

Initially, 52 features are calculated for the SR and the CB dataset, while the SRCB dataset contains 104 features (including features from both the SR and CB dataset). After selecting the most relevant features, it is noticeable that the dimensionality reduction varies for the SR, CB and SRCB datasets. Specifically, 62% of the features are removed from the SR dataset, 54% of the features are removed from the CB dataset and 64% of the features are removed from the SRCB dataset. Figure 4.3 represents the number of returned features after each phase of feature selection for the three datasets. The selected features and their distribution is presented at Appendix A.1 Features Distribution in page 82.

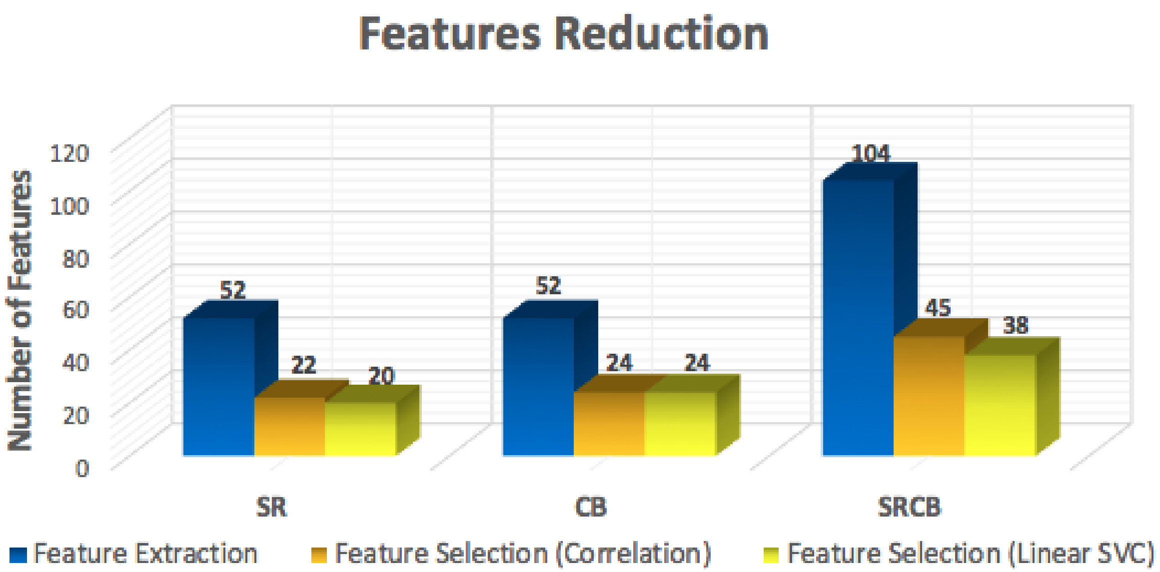


Figure 4.3: The final number of features per dataset is presented, after the process of feature selection. Feature extraction represents the number of initially calculated features, feature selection (correlation) represents the reduction of highly correlated features, and the feature selection (linear SVC) represents the reduction due to low information gain. The validation is based on LOSOCV for 15 subjects (detect 13 activities based on Controlled Study). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

Each dataset contains different features, and each feature has a different weight in the classification performance. Thus, the contribution of these features to the total classification performance varies. For instance, the mean distance of axes y and z is the most highly scored feature (SR_meandistance_y-z) for the SR and SRCB datasets, while the mean value of axis y (CB_mean_y) is the most highly scored feature for the CB dataset. The 15 most highly scored features for each dataset are depicted in the following figure (Figure 4.4). Regarding the SR dataset, 15 out of 20 features contribute to the total classification performance with an overall accuracy above 90%, while 15 out of 24 features contribute to the total classification performance with an overall score 85% for the CB dataset. On the other hand, 15 out of 98 features contribute to the total classification performance with an overall accuracy below 70% for the SRCB dataset. Hence, it is clear that the features on the SRCB dataset have different weights and all the selected features (38 in total) are required for better classification results.

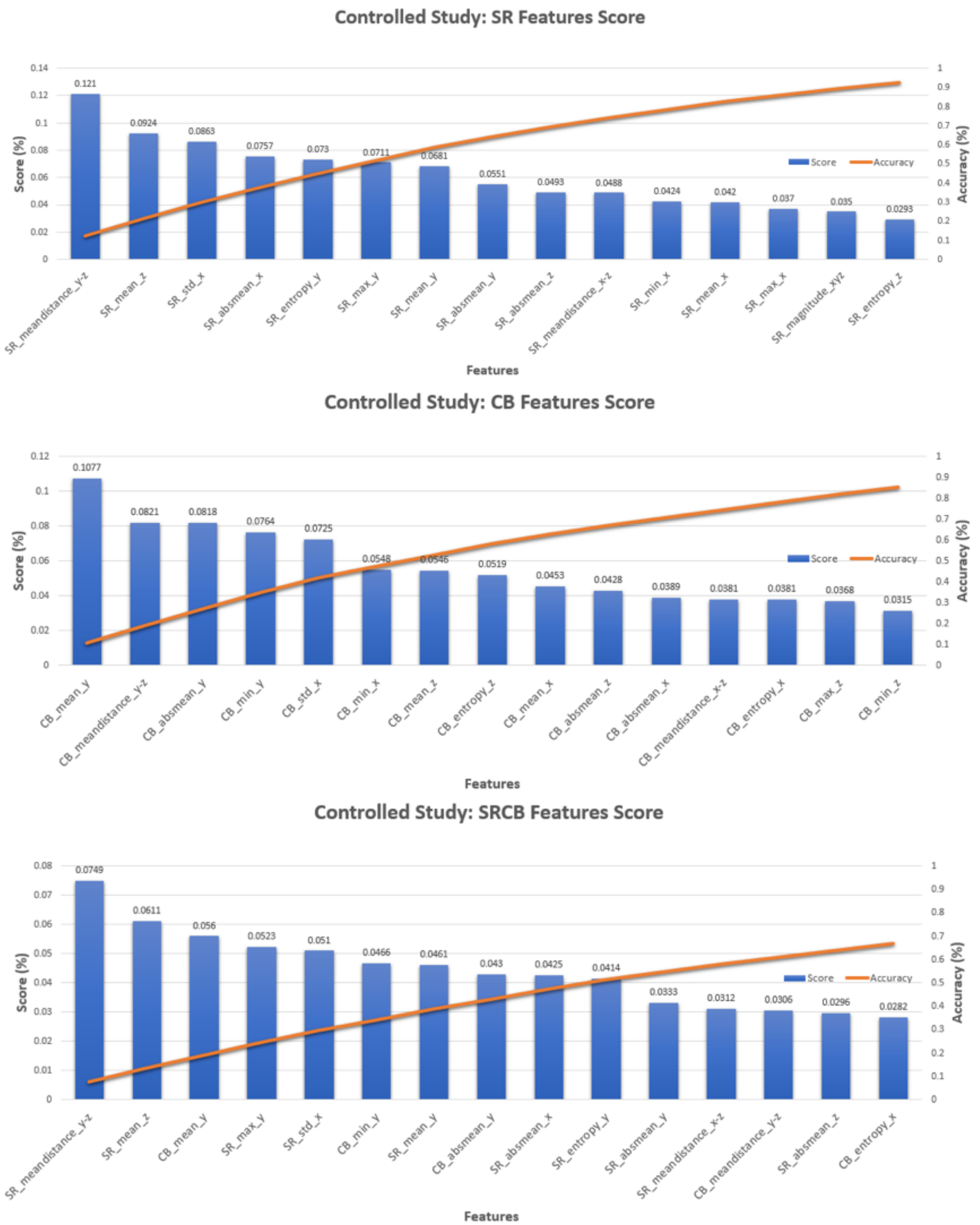


Figure 4.4: The 15 most highly scored features are depicted for each dataset in the Controlled Study. In addition to the score of each feature, the total contribution of the first 15 features to the classification performance is presented as well in the right axis (accuracy score). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

4.1.2. CLASSIFICATION

After data processing, the learning model (which consists of 15 out of 40 subjects from the Controlled Study) is divided into training and test set for classifying and predicting the activities. In order to split the dataset into training and test set, the method of Leave-One-Subject-Out Cross Validation (LOSOCV) is used. For this reason, the training set consists of 14 subjects and the test set consists of one single subject. This validation method is repeating 15 times, where the test set contains data from a unique subject every time, and the averaged classification performance is calculated.

The classification is examined based on four algorithms; kNN, naïve Bayes, SVC, and Random Forest. The algorithms are compared based on their performance in predicting the 13 activities on the Controlled Study and are presented in the following figure. As Figure 4.5 shows, the Random Forest classifier outperforms the other three classifiers. Specifically, Random Forest performs significantly better compared to kNN and naïve Bayes and slightly improved compared to SVC. Thus, the Random Forest algorithm will be used in this study for classifying the activities.

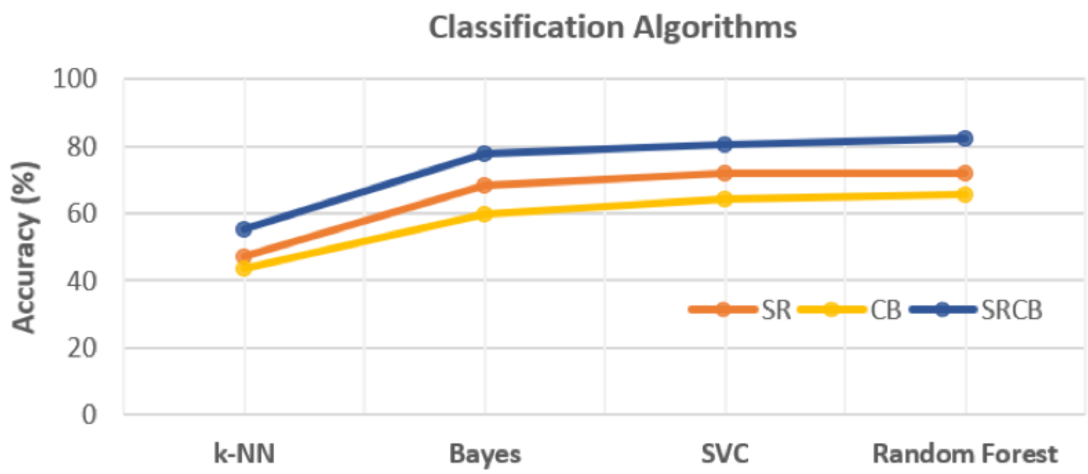


Figure 4.5: Four algorithms are compared based on their classification performance (accuracy score). The validation is based on LOSOCV for 15 subjects (detect 13 activities based on Controlled Study). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

The algorithms mentioned above have been compared based on their default parameters and hyper-parameters, and the Random Forest has been selected for the classification model. Additionally, the performance of this model is further enhanced though hyperparameter tuning using the Scikit-Learn tools in Python. The default Random Forest algorithm is compared with the optimized algorithm after the hyperparameter tuning in Figure 4.6.

Generally, Scikit-Learn implements a classification model with default hyperparameters without ensuring the optimization of the classification problem. Regarding Random Forest algorithms, hyperparameters include for instance the number of decision trees in the forest (n_estimators), the number of features considered by each tree when splitting a node (max_features), the number of levels in each decision tree (max_depth), while parameters include variables and thresholds used to split each node learned during training [81]. In order to define what are the best hyperparameters for the classification problem, a wide range of values is evaluated (for every hyperparameter) using the Scikit-Learn RandomizedSearchCV method [82]. Trying to avoid overfitting issues, the LOSOCV cross validation was used. Finally, the following hyperparameters were chosen for the classification model.

```
model = RandomForestClassifier (n_estimators=600, criterion='gini',
        max_depth=18, max_features = "auto", min_samples_leaf = 40,
        oob_score = True, n_jobs = -1)
```

As can be seen in Figure 4.6, the hyperparameter tuning has slightly enhanced the overall classification performance of the Random Forest algorithm. Specifically, the performance for the SR dataset has been increased by 1%, while the CB and SRCB classified datasets have been improved by 1.96% and 0.79%, respectively. Nevertheless, the Random Forest combined with the hyperparameters, as mentioned above, will be used as the classification algorithm for this activity recognition model.

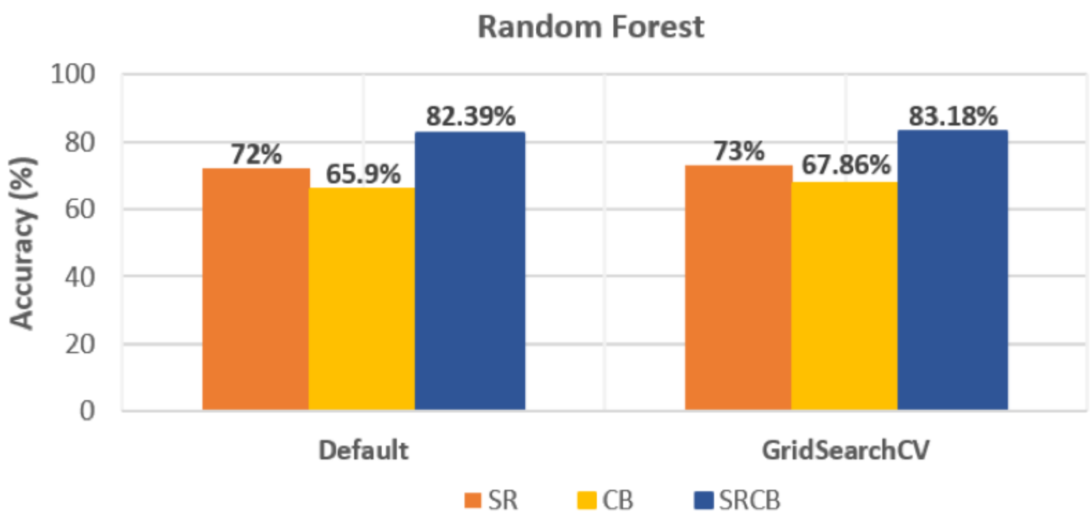
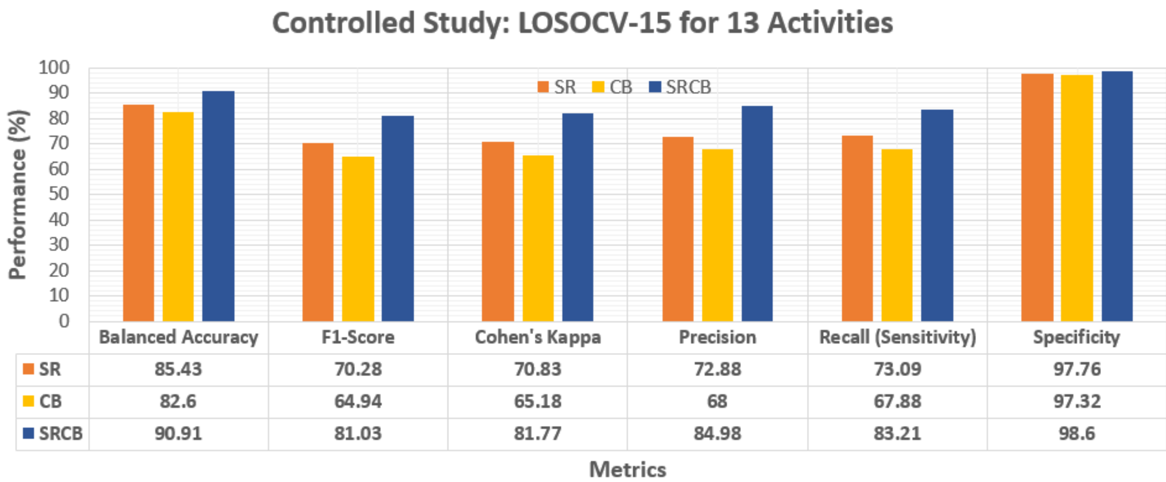


Figure 4.6: The performance of Random Forest classifier is compared by using either default or tuned hyperparameters (GridSearchCV). The validation is based on LOSOCV for 15 subjects (detect 13 activities based on Controlled Study). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

4.1.3. EVALUATION LOSOCV-15 FOR 13 ACTIVITIES

After splitting the learning dataset into training and test set, the evaluation of predicting activities on the test set takes place. In this section, the learning model is evaluated for predicting 13 activities based on LOSOCV for 15 subjects, for both SR, CB and SRCB datasets in the Controlled Study. The training set consists of data from 14 subjects, while the test set consists of 1 subject, repeating the validation for 15 times and calculating the average metric values. The evaluation is based on the confusion matrixes and the calculated metrics. The overall performance is presented in Figure 4.7.



	Balanced Accuracy	F1-Score	Cohen's Kappa	Precision	Recall (Sensitivity)	Specificity
SR	85.43	70.28	70.83	72.88	73.09	97.76
CB	82.6	64.94	65.18	68	67.88	97.32
SRCB	90.91	81.03	81.77	84.98	83.21	98.6

Figure 4.7: The learning model is evaluated for predicting 13 activities through different metrics. The validation is based on LOSOCV for 15 subjects from the Controlled Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

Concerning the stingray (SR) dataset, the overall classification performance for predicting 13 activities accurately is 70.28% (F1-score). Based on the confusion matrix (see Confusion Matrix at Appendix A.2 in page 84), it is clear that activities such as lying with face up, lying with face down, running and cycling are predicted accurately with a true positive score above 90%. On the other hand, activities such as sitting and dynamic sitting are misclassified (44% and 53%, respectively). Similarly, the activities standing and dynamic standing (50% and 77%, respectively) have been misclassified with each other but also with sitting and dynamic sitting. Furthermore, walking and walking on stairs have been misclassified as well (62% and 64%, respectively).

For the chillband (CB) dataset, the overall classification performance for predicting 13 activities is 64.94% (F1-score). Similar to SR dataset, activities such as running and cycling can be predicted accurately (96% and 94%, respectively), while the classification performance for sitting (49%), dynamic sitting (86%), standing (84%) and dynamic standing (87%) has been improved. However, the lying and walking activities have been misclassified compared to SR dataset. In particular, the average true positive score for predicting the four lying activities is 54.75%, the walking is 54% and the average score for walking on stairs is 75%.

By combining accelerometer data from the stingray and chillband sensors, the classification performance is enhanced. For the SRCB dataset, the overall classification performance for predicting 13 activities is 81.03% (F1-score). Even though the overall prediction score is better for this dataset, there are still some activities that are misclassified, including the activities sitting (70%), dynamic sitting (87%), lying to the left (74%), lying to the right (85%), walking upstairs (60%), walking downstairs (73%), and walking (61%). For the rest activities, the true positive score is significantly better: lying with face up (100%), lying with face down (92%), standing (93%), dynamic standing (94%), running (97%) and cycling (97%).

4.1.4. EVALUATION LOSOCV-15 FOR 7 ACTIVITIES

According to the evaluation for predicting 13 activities, some activities, such as sitting, dynamic sitting, lying to the left, lying to the right, walking upstairs and walking downstairs, cannot be predicted significantly. For this reason, we will also investigate the activity recognition system for predicting 7 main activities, after merging the similar ones. These 7 activities include sitting, lying, standing, walking on stairs, walking, running and cycling. Consequently, the learning model is also evaluated for predicting 7 activities based on LOSOCV for 15 subjects, for both SR, CB and SRCB datasets in the Controlled Study. The overall performance is presented in Figure 4.8 (see also Confusion Matrix at Appendix A.2 in page 84).

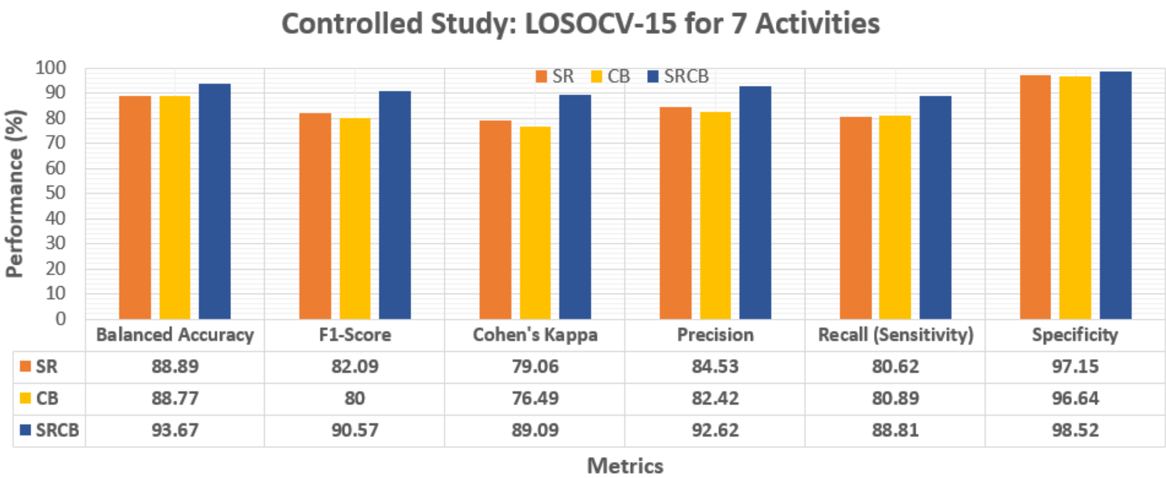


Figure 4.8: The learning model is evaluated for predicting 7 activities through different metrics. The validation is based on LOSOCV for 15 subjects from the Controlled Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

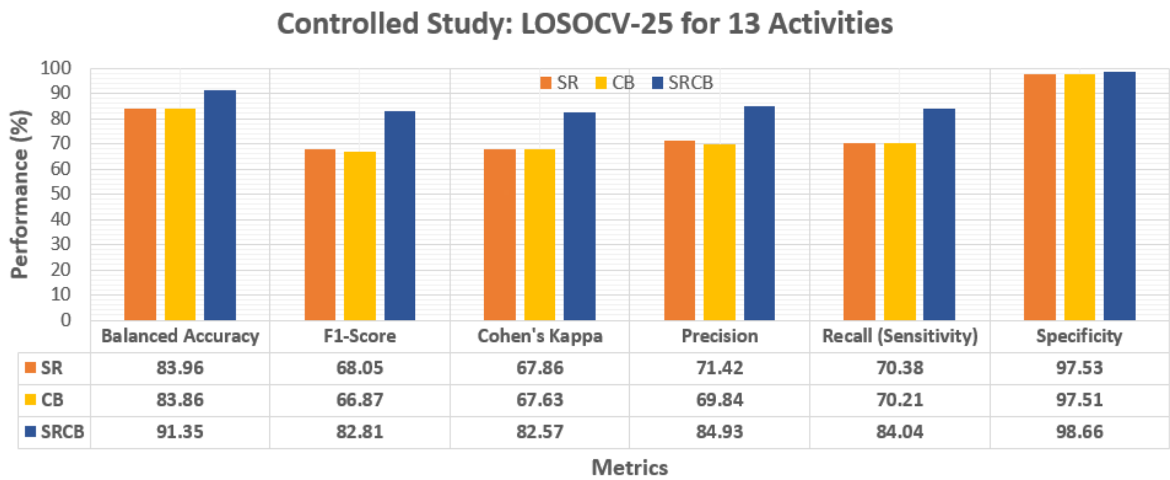
Regarding the stingray (SR) dataset, the overall classification performance for predicting 7 activities accurately is 82.09% (F1-score). Overall, the performance for lying, running and cycling is significant accurate (81%, 96% and 94%, respectively). However, the activity sitting is misclassified with standing (64% and 73%, respectively), and the activity walking on stairs is misclassified with walking (83% and 62%, respectively).

For the chillband (CB) dataset, the overall classification performance for predicting 7 activities is 80% (F1-score). Similar to SR dataset, activities such as running and cycling can be predicted accurately (96% and 94%, respectively), while the walking on stairs and walking are also misclassified (81% and 54%, respectively). In contrast to SR, the recognition of activities sitting and standing has been improved (75% and 86%, respectively); however, they are misclassified with the activity lying (81%).

Compared to SR and CB, the classification performance for the SRCB is significantly enhanced to 90.57% (F1-score) for recognizing 7 activities. Most of the activities can be detected accurately (sitting: 89%, lying: 98%, standing: 94%, running: 97% and cycling: 97%), despite the walking on stairs and walking which have a score 86% and 61%, respectively.

4.2. VALIDATION MODEL

The validation model for the Controlled Study consists of accelerometer data from the additional 25 subjects (15 out of 40 subjects have been used for the learning model). This model is divided into training and test set, based on LOSOCV for 25 subjects, in order to evaluate the system's performance for recognizing the unseen activities of the test set. Thus, the training set consists of data from 24 subjects, while the test set consists of 1 subject, repeating the validation for 25 times and calculating the average metrics. Furthermore, data from the learning model (15 subjects) are added to the training set in order to enhance the training (39 subjects in total). For the final validation of our system, two different approaches will be investigated for recognizing activities in a controlled environment. The first is based on predicting 13 activities and the second one is based on predicting 7 activities. The results for the validation of the system are presented in Figure 4.9. For further understanding, see also the actual versus predicted activities for the activity recognition model at Appendix B in page 91 (the plot is made based on the most highly scored feature for each sensor).



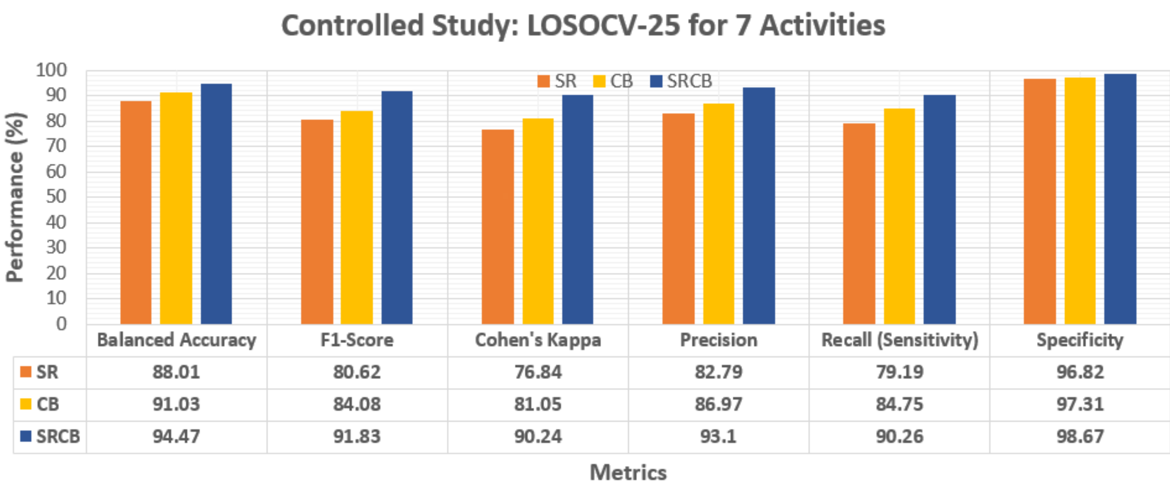


Figure 4.9: The validation model is evaluated for predicting 13 activities (top graph) and 7 activities (bottom graph) through different metrics. The validation is based on LOSOCV for 25 subjects from the Controlled Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

4.2.1. STINGRAY

Regarding the stingray (SR) sensor, the overall classification performance for predicting 13 activities is 68.05% (F1-score). Similar to the results from the learning model, the activities sitting and dynamic sitting have been misclassified with standing and dynamic standing. Furthermore, lying to the left is misclassified with lying to the right, while walking is misclassified with walking on stairs. Based on the confusion matrix (see Figure 4.10), the true positive score for the predicted activities is: sitting (36%), dynamic sitting (64%), lying with face up (94%), lying with face down (92%), lying to the left (67%), lying to the right (65%), standing (35%), dynamic standing (75%), walking upstairs (63%), walking downstairs (78%), walking (66%), running (94%) and cycling (85%).

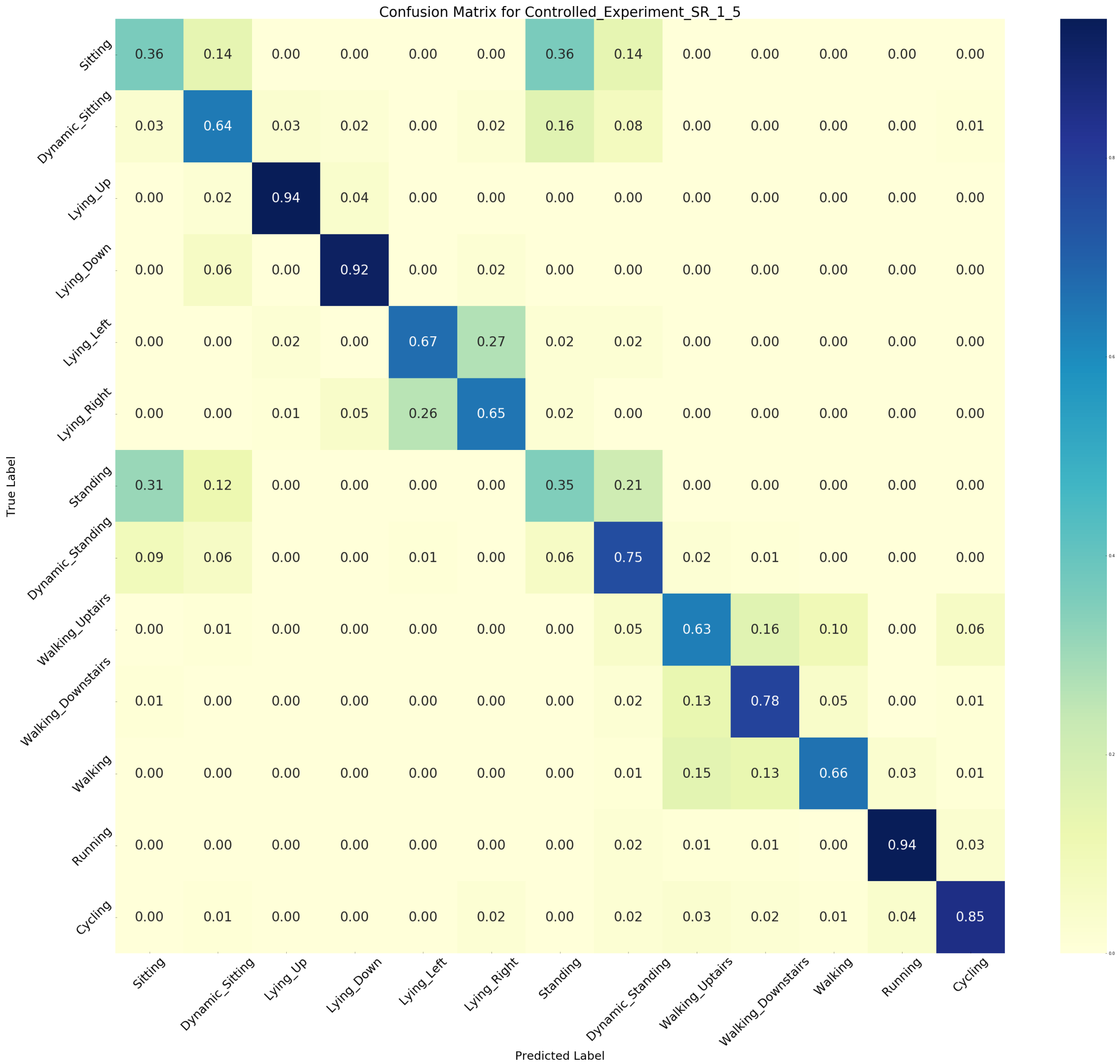


Figure 4.10: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-25 from the Controlled Study, by detecting 13 activities.

The overall classification performance for predicting 7 activities is 80.62% (F1-score). Based on the confusion matrix (see Figure 4.11), the true positive score for the predicted activities is: sitting (59%), lying (96%), standing (69%), walking on stairs (85%), walking (66%), running (94%) and cycling (85%).

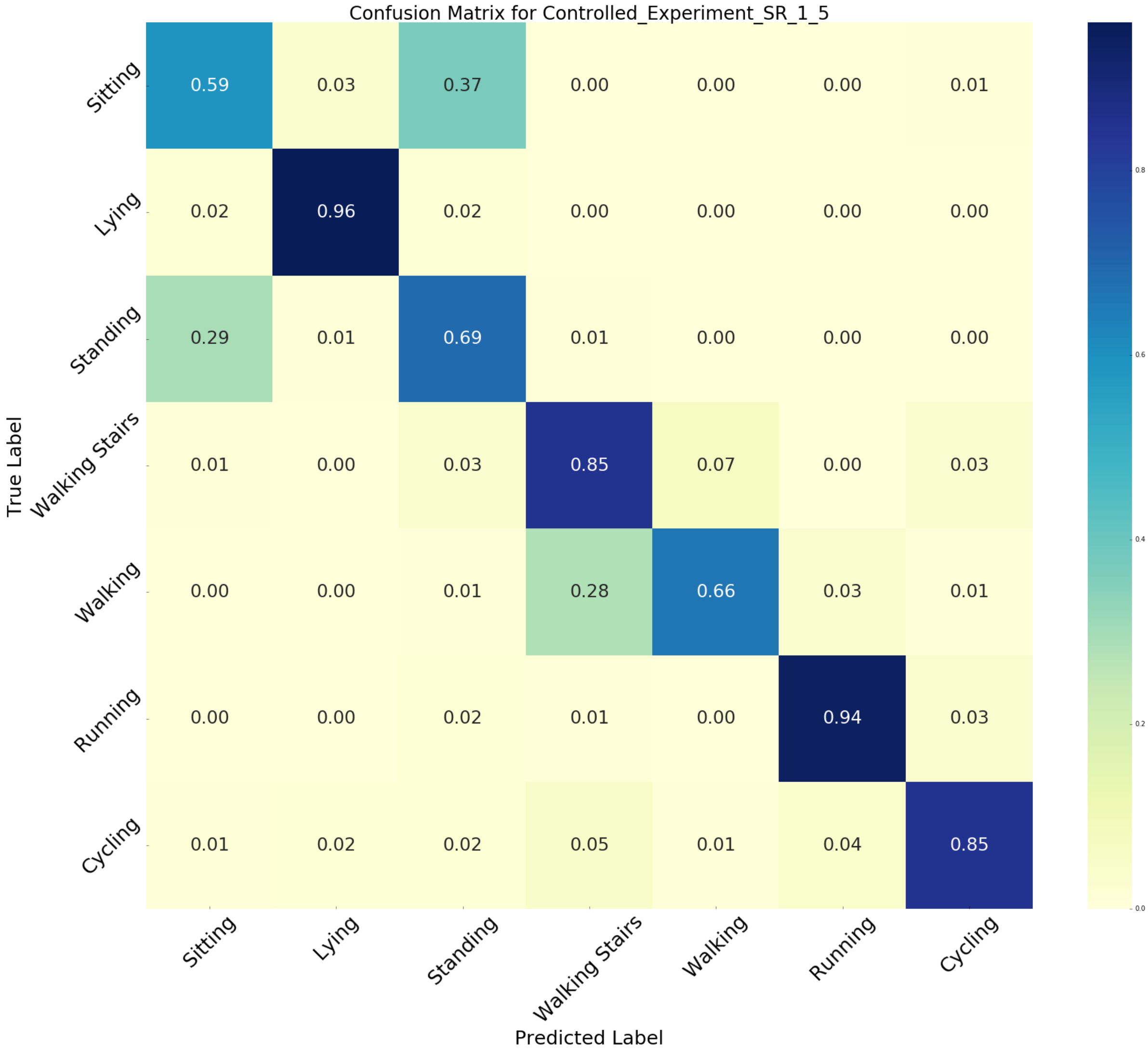


Figure 4.11: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-25 from the Controlled Study, by detecting 7 activities.

4.2.2. CHILLBAND

Regarding the chillband (CB) sensor, the overall classification performance for predicting 13 activities is 66.87% (F1-score). It can be seen from the confusion matrix (see Figure 4.12), that the activity sitting has been misclassified with lying down, the activity lying to the left has been misclassified with lying to the right, and the walking has been misclassified with walking on stairs. The true positive score for the predicted activities is: sitting (65%), dynamic sitting (92%), lying with face up (59%), lying with face down (55%), lying to the left (46%), lying to the right (47%), standing (94%), dynamic standing (91%), walking upstairs (61%), walking downstairs (50%), walking (64%), running (95%) and cycling (92%). Compared to the SR sensor, the CB performs significantly better on predicting the sitting, dynamic sitting, standing, dynamic standing, and cycling. On the other hand, CB performs worst for all the lying activities, as well as the walking downstairs activity.

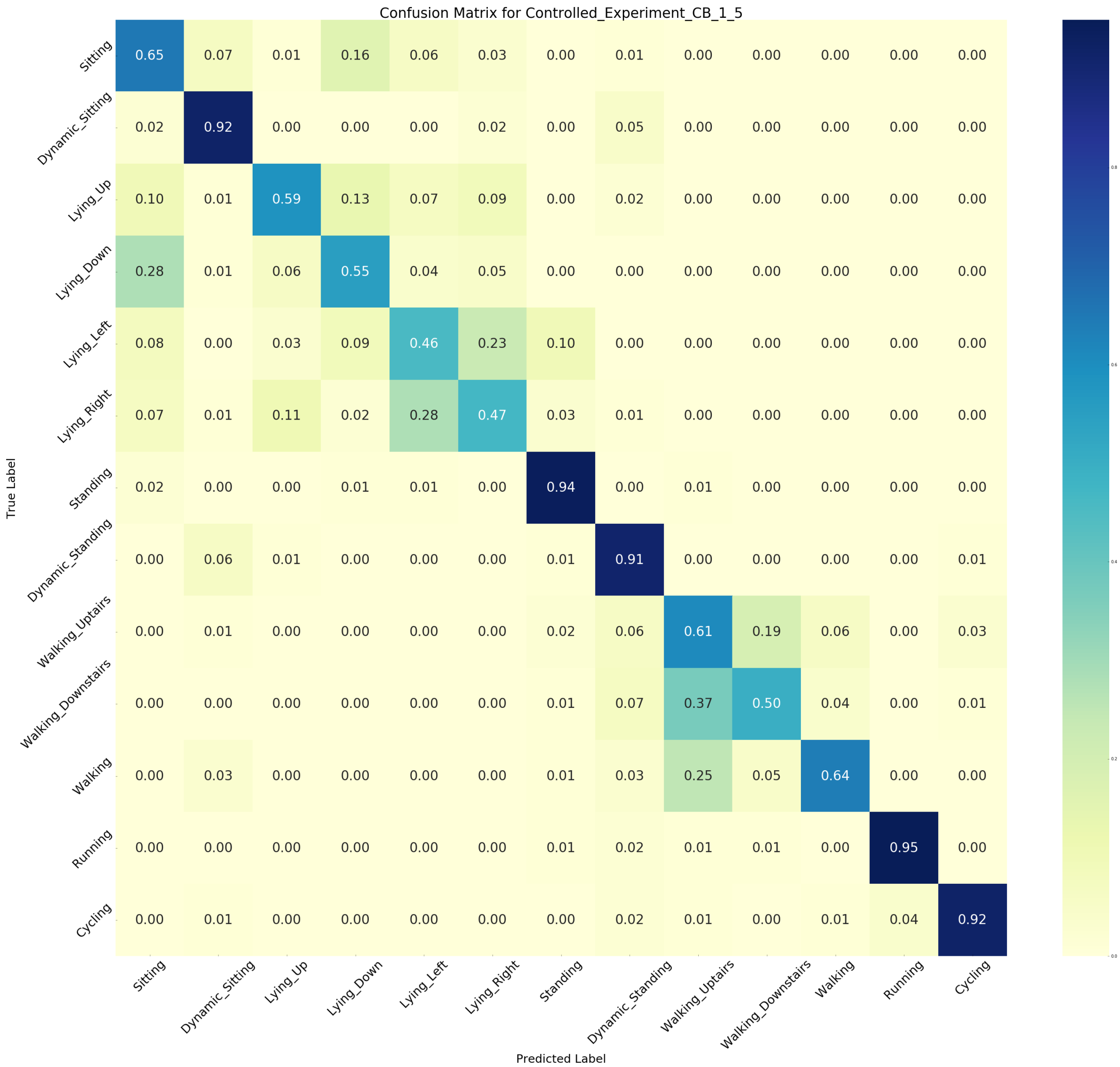


Figure 4.12: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-25 from the Controlled Study, by detecting 13 activities.

The overall classification performance for predicting 7 activities is 84.08% (F1-score). Based on the confusion matrix (see Figure 4.13), the score for the predicted activities is: sitting (83%), lying (81%), standing (93%), walking on stairs (84%), walking (64%), running (95%) and cycling (92%). It is noticeable that the activity recognition through CB for the activities sitting, standing and cycling has been significantly enhanced. Furthermore, it is worth mentioning that the CB outperforms the SR sensor for predicting the aforementioned 7 activities (compared to the learning model).

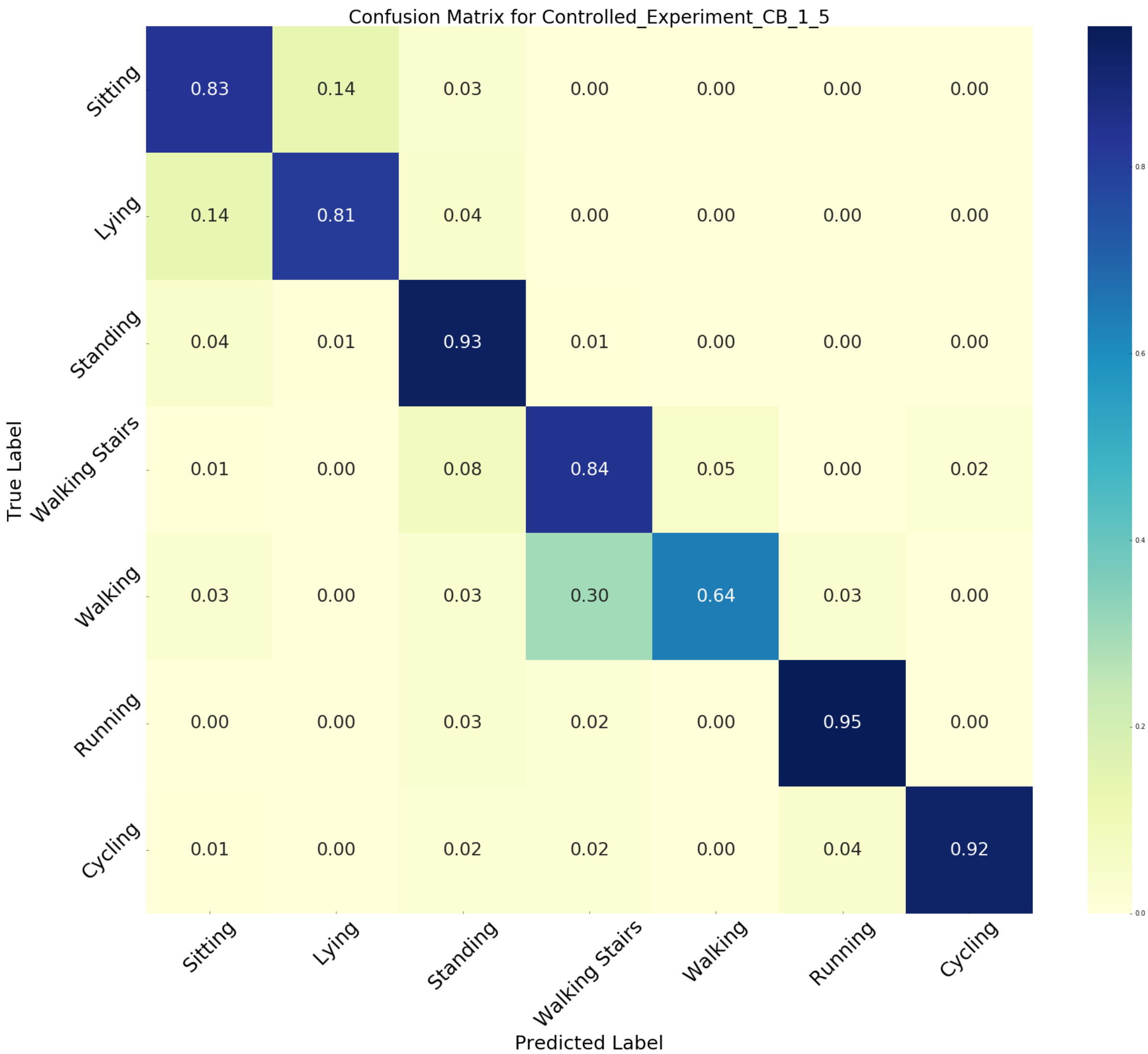


Figure 4.13: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-25 from the Controlled Study, by detecting 7 activities.

4.2.3. STINGRAY & CHILLBAND

The activity recognition based on the stingray-chillband (SRCB) sensor achieves the best classification performance for predicting 13 activities, which is 82.81% (F1-score). In particular, wearing both sensors instead of wearing accelerometers either on the wrist or on the chest, results to an increase of the classification performance by 15%. However, the activity lying to the left is misclassified with lying to the right, and the activity walking is misclassified with walking on stairs. Based on the confusion matrix (see Figure 4.14), the true positive score for the predicted activities is: sitting (93%), dynamic sitting (93%), lying with face up (96%), lying with face down (90%), lying to the left (66%), lying to the right (66%), standing (94%), dynamic standing (91%), walking upstairs (71%), walking downstairs (75%), walking (71%), running (92%) and cycling (92%).

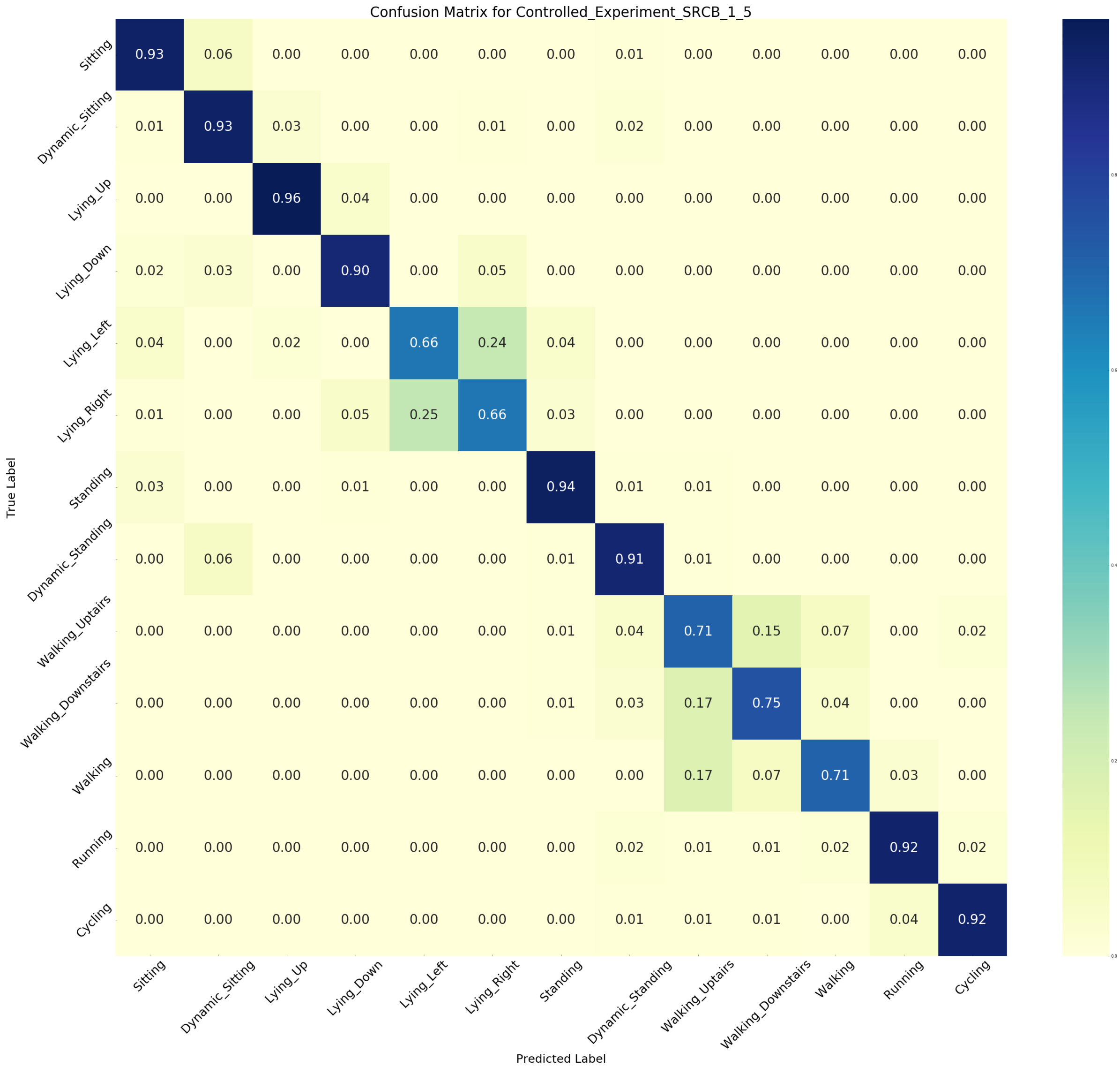


Figure 4.14: The confusion matrix for the stingray-chillband (SRCB) sensors is presented. The validation is based on LOSOCV-25 from the Controlled Study, by detecting 13 activities.

The overall classification performance for predicting 7 activities is 91.83% (F1-score). The SRCB accelerometer outperforms the SR and CB. Based on the confusion matrix (see Figure 4.15), the score for the predicted activities is: sitting (96%), lying (96%), standing (94%), walking on stairs (89%), walking (71%), running (92%) and cycling (92%). It is worth mentioning that our system for predicting the 7 activities has been significantly improved, by combining accelerometer data worn on the chest and on the wrist. Thus, the SRCB contributes sufficiently to recognizing most of the activities with a score above 90%, except the walking activity.

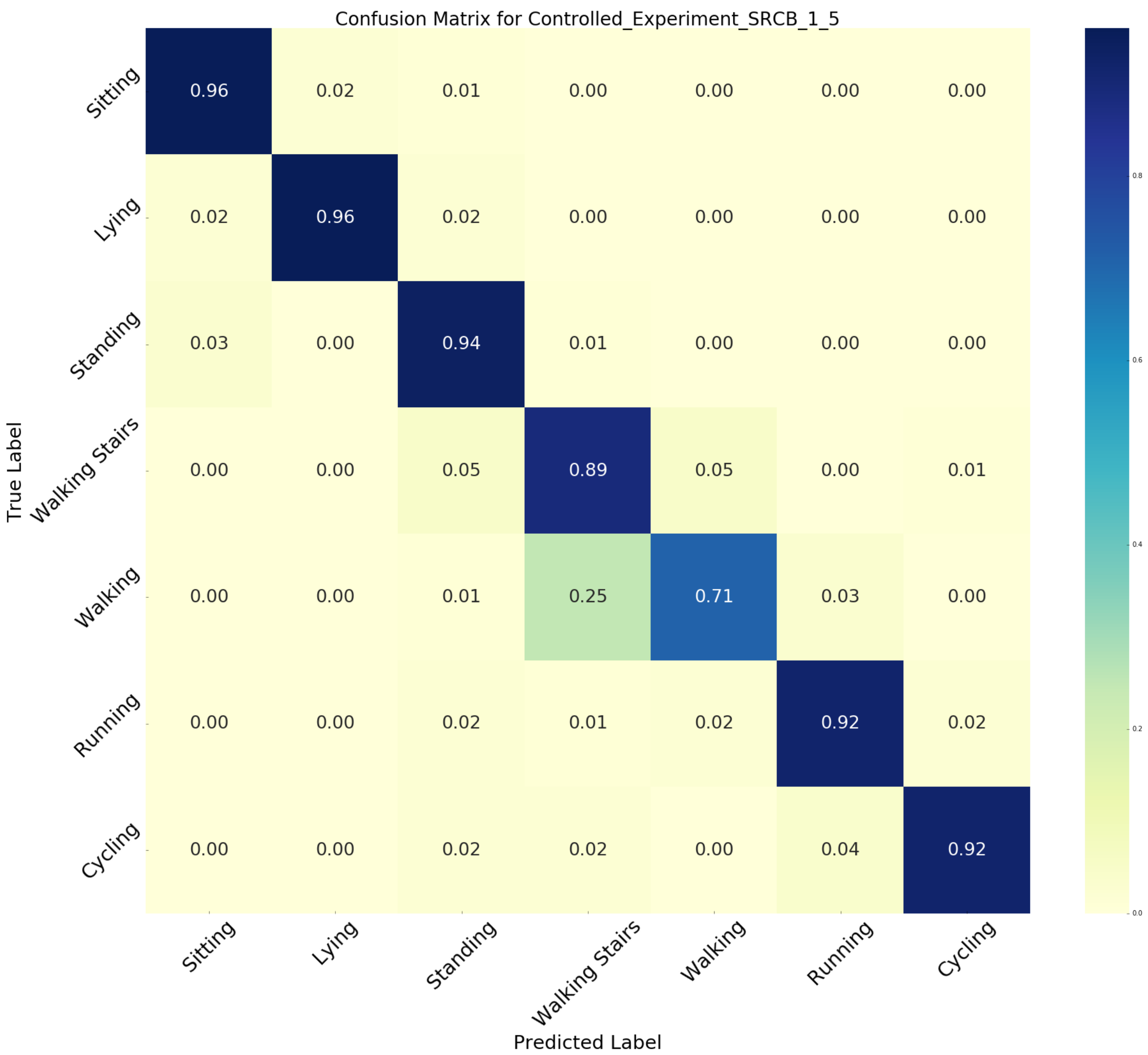


Figure 4.15: The confusion matrix for the stingray-chillband (SRCB) sensors is presented. The validation is based on LOSOCV-25 from the Controlled Study, by detecting 7 activities.

4.3. DISCUSSION

The Controlled Study includes data from 40 Subjects. In particular, data from 15 subjects were used for optimizing the learning model, while data from the other 25 subjects were used to validate the classification performance.

Based on the results from the learning model, we decided to implement the activity recognition system with the following parameters. The features extraction was performed based on 5 seconds window size and 4 seconds overlap. It is worth mentioning that we found some additional window segments that might have a similar classification performance (e.g., 6sec window size and 3sec overlap, and 10sec window size and 5sec overlap) with the one we selected to perform the data processing. However, these window segments need further investigation. Concerning the features selection, the calculated features were reduced to the ones that contribute the most to the classification performance. The most significant features are the mean, absolute mean, mean distance, standard deviation, max, min, and entropy. Nevertheless, each one of these features has a different weight in the classification, which varies per sensor. Additionally, we examined different classification algorithms and we concluded that the Random Forest performs significantly better in our classification problem, followed by the SVC.

Our classification model for activity recognition was evaluated based on data from the validation model (25 Subjects). One of the outcomes of this study is that combining chest-worn (SR) and wrist-worn (CB) sensors, instead of using standalone accelerometer sensors (either SR or CB), enhances the classification performance. Overall, the total classification performance for chest-worn sensors is approximately the same with wrist-worn sensors; however, the accuracy on detecting different activities varies per sensor.

For predicting 13 activities, the SR achieves a total 68.05% F1-score, while the CB achieves 66.87% F1-score. In particular, the SR performs better for detecting the lying (including lying with face up, lying with face down, lying to the left and lying to the right), and walking downstairs. On the other hand, the CB performs better for detecting the activities sitting, dynamic sitting, standing, dynamic standing, and cycling. The running activity can be predicted significantly well for both the wrist and chest worn sensors, while the prediction for the activities walking and walking on stairs is not always accurate. A possible reason for that could be that the walking activity recognition relies not only on accelerometers but also to other inertial sensors, such as gyroscopes. Compared to the activity recognition through standalone SR and CB sensors, the combination of these two sensors can significantly increase the model's prediction (82.81% F1-score). Consequently, our model can predict most of the activities of the Controlled Study accurately, except the lying to the left, lying to the right, walking and walking on stairs. A possible solution to this could be the use of additional inertial sensors or the placement of these sensors to other parts on the body.

To overcome the misclassification due to the similarity of some performed activities, we have also examined our activity recognition model for detecting 7 activities, after merging the similar ones (e.g., lying to the left, lying to the right, walking upstairs and walking downstairs). For predicting 7 activities, the SR receives a total 80.62% F1-score, the CB achieves 84.08% F1-score, while the SRCB outperforms the SR and CB, and achieves 91.83% F1-score. In particular, the SRCB can predict almost all the activities with a higher score, compared to SR and CB. However, it is worth mentioning that the running activity performs better with the accelerometer placed on the wrist (CB: 95%), followed by the chest-worn sensor (SR: 94%) and the combination of these two (SRCB: 92%).

Even though our activity recognition model is highly scored for predicting 7 activities through SRCB accelerometer data, we conclude that the accuracy of our system depends on the purpose of the application that will be used. For instance, if the aim is to recognize the running activity, a sensor placed on the wrist or the chest is sufficient. Otherwise, the system is significantly more accurate by recording accelerometer data, placed on both the wrist and the chest. Specifically, the SRCB contributes adequately to recognize most of the activities of the controlled study, except the walking activity, which is misclassified with the walking on stairs activity. However, if these activities are merged to walking, including both walking and walking on stairs, then the classification performance will be significantly improved.

5

Free-Living Study

In this chapter, we will analyze data from the Free-Living Study in order to implement a classification model for detecting activities, performed in an uncontrolled environment. The free-living experiment is also mentioned as uncontrolled or ambulatory experiment in this report. This dataset consists of raw accelerometer data performed from 37 subjects, in an uncontrolled environment. The acquired accelerometer dataset includes data from the stingray (SR) sensor, the chillband (CB) sensor and the combination of these two stingray-chillband sensors (SRCB), and will be investigated separately. It is worth mentioning that the total number of activities in this study is not balanced since subjects performed the activities without following any instructions. Compared to the Controlled Study, the activities lying with face up, lying with face down, lying to the left, lying to the right were not performed. Additionally, only one subject performed the activity running, but only for a few seconds, and thus, we decided to exclude it from the dataset. Consequently, 8 activities in total will be investigated, including sitting, dynamic sitting, standing, dynamic standing, walking upstairs, walking downstairs, walking and cycling. In addition to recognizing 8 activities, we will also evaluate the model for predicting 5 main activities, after merging the similar ones. These activities include sitting (both static and dynamic sitting), standing (both static and dynamic standing), walking on stairs (both ascending and descending stairs), walking, and cycling.

We evaluate the performance of our activity recognition system for uncontrolled settings using data from 22 subjects (test set) and based on the following three methods. Concerning the first method Validation A, data from 15 subjects of the Free-Living Study are used to train the classification model. For the Validation B, data from subjects of the Controlled Study are used for the learning model and for the Validation C, data from both the Controlled and Free-Living Study are used to train the model.

5.1. Learning Model

The learning model consists of data from the Free-Living Study, based on 15 subjects. In order to compare the results of this study with the Controlled Study, we decided to use the same parameters that we selected in the latter. Thus, we did not investigate further what window segment, features, and classification algorithm might perform better for this dataset. Briefly, we use 5 seconds window size with 4 seconds overlap, following the same procedure for the features extraction and selection, and classifying the model based on GridSearchCV Random Forest. Similar to the Controlled Study, the evaluation of the learning model is based on predicting 8 (all the activities) and 5 (after merging the similar ones) activities, subsequently.

5.1.1. Data Processing

After extracting the features based on 5 seconds window size and 4 seconds overlap, the phase of features selection takes place. Similar to the procedure that we followed on processing data in Controlled Study, highly correlated features are removed based on the Pearson's correlation coefficient, while features with low information gain are subsequently removed through the Linear SVC estimator.

Initially, 52 features are calculated for the SR and the CB dataset, while the SRCB dataset contains 104 features (including features from both the SR and CB dataset). After selecting the most relevant features, it is noticeable that the dimensionality reduction varies for the SR, CB and SRCB datasets. Specifically, 60% of the features are removed from the SR dataset, 58% of the features are removed from the CB dataset and 60% of the features are removed from the SRCB dataset. Figure 5.1 represents the number of returned features after each phase of feature selection for the three datasets. The selected features and their distribution is presented at Appendix C.1 Features Distribution in page 100.

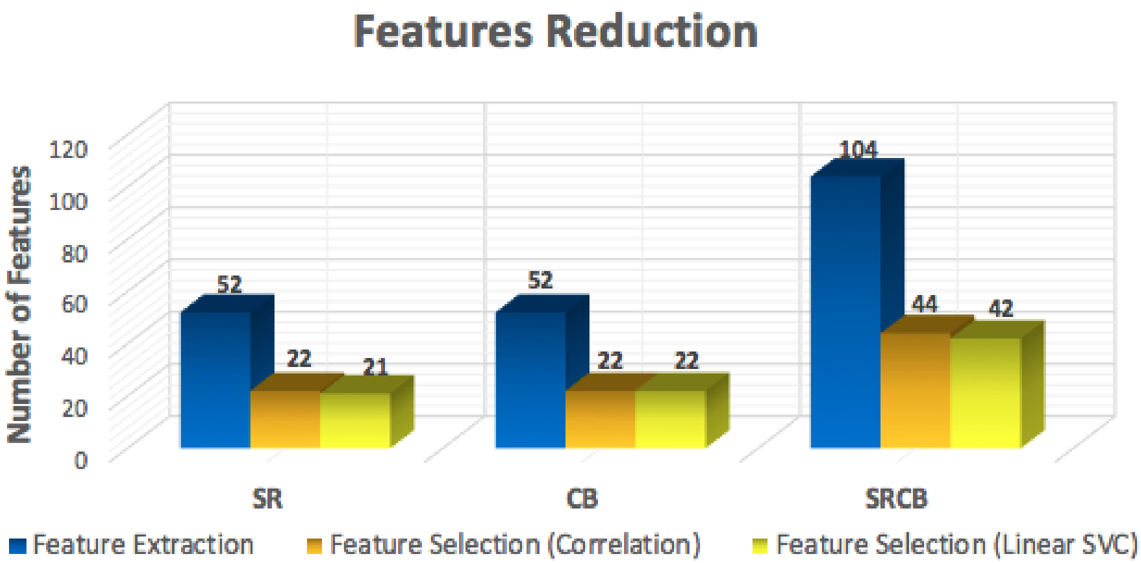


Figure 5.1: The final number of features per dataset is presented, after the process of feature selection. Feature extraction represents the number of initially calculated features, feature selection (correlation) represents the reduction of highly correlated features, and the feature selection (linear SVC) represents the reduction due to low information gain. The validation is based on LOSOCV for 15 subjects (detect 8 activities based on Free-Living Study). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

It is clear that each dataset contains different features, and each feature has a different weight in the classification performance. For instance, the magnitude of the axes x, y and z is the most highly scored feature (SR_magnitude_xyz) for the SR dataset. The score for this feature is 12.64%, while the feature SR_meandistance_y-z achieves the best score (12.1%) in the Controlled Study. Similar to the Controlled Study, the mean values of axis y (CB_mean_y) is the most highly scored feature for the CB, but also for the SRCB. This feature achieves the highest score 9.8% in the SRCB dataset for the Free-Living Study, while in the Controlled Study achieves 5.6%. Thus, the contribution of these features to the total classification performance varies and especially for the SR and SRCB datasets, which differentiates from the Controlled Study. The 15 most highly scored features for each dataset are depicted in the following figure (Figure 5.2).

Additionally, during the data processing phase, we examined three different types of annotation for the activities in the Free-Living Study. As mentioned in chapter 3, the activities are annotated every 30 seconds based on a taken picture through a wearable camera, and thus, an activity can be annotated only for two times per minute. Annotation type A refers to annotate activities for the last 15 seconds of every taken picture. Hence, we labeled the activity not only for the corresponding timestamp of the taken picture, but also for the previous 14 seconds, and we linked them with the timestamps of the SR, CB and SRCB dataset. Annotation type B refers to 30 seconds, including 14 seconds before and 14 seconds after a taken picture. Annotation type C is related to the exact second of every taken picture. Based on Figure 5.3, we can see that the type A outweighs the other types, and thus, will be used for this study.

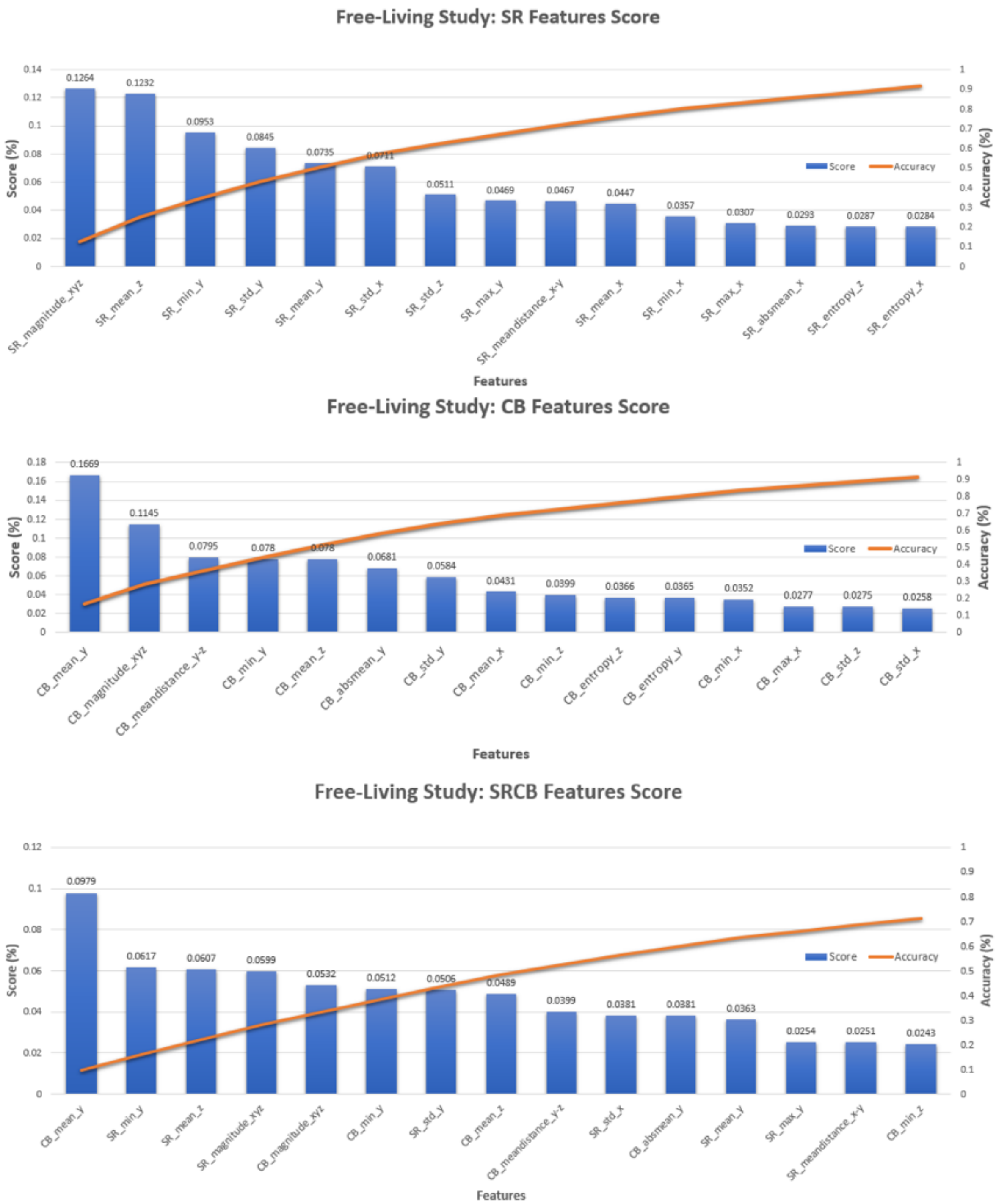


Figure 5.2: The 15 most highly scored features are depicted for each dataset in the Free-Living Study. In addition to the score of each feature, the total contribution of the first 15 features to the classification performance is presented as well in the right axis (accuracy score). SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

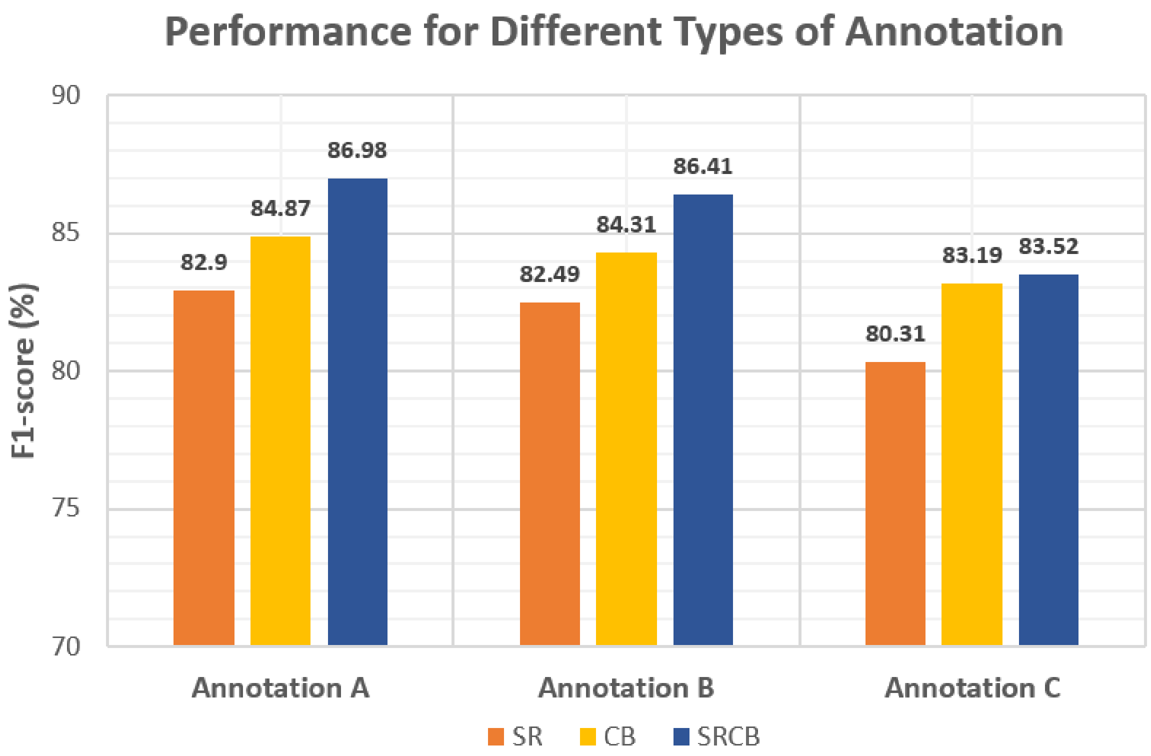


Figure 5.3: The performance of the three types of activities annotation for the Free-Living Study. F1-score is used to evaluate the performance for predicting 5 activities. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

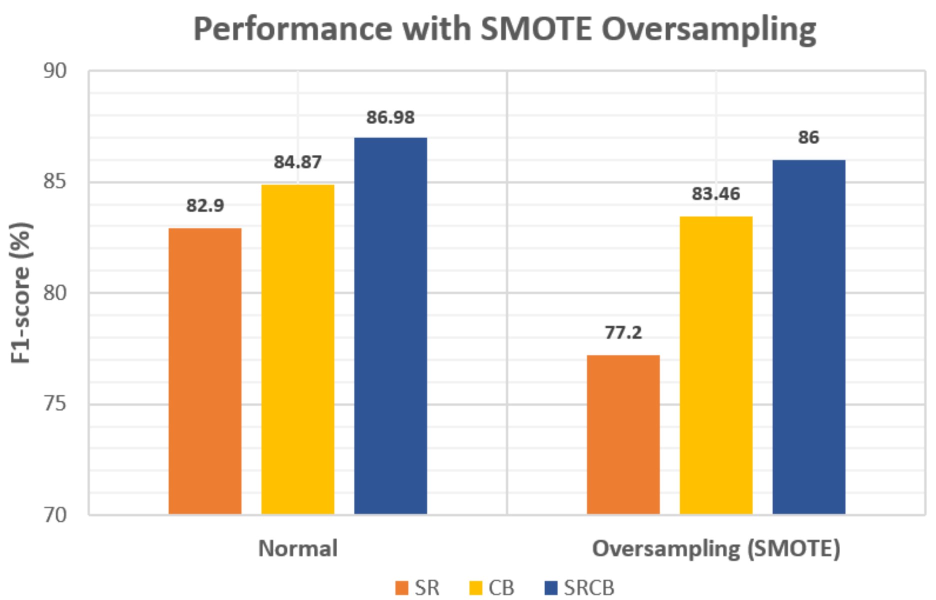


Figure 5.4: The performance of the learning model in the Free-Living Study with and without the oversampling method SMOTE. F1-score is used to evaluate the performance for predicting 5 activities. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

For the last step in data processing, we tested the oversampling method based on SMOTE (Synthetic Minority Oversampling Technique) algorithm, trying to address the imbalanced data. Overall, oversampling the minority class observations can significantly enhance the classification performance. SMOTE creates synthetic observations of the minority classes (e.g., walking upstairs and walking downstairs) by finding similar observations based on the k-nearest-neighbors for minority class observations [83]. Specifically, it randomly chooses one of the k-nearest neighbors and uses this to create a similar new observation (but randomly tweaked). The SMOTE performance for the learning model is presented in Figure 5.4. As it can be seen, the oversampling method does not improve the classification performance, and thus, is not preferred for this study. A problem for that could be that all the performed activities, except standing and dynamic standing, are considered as minority classes, leading to poor observations (see also Figure 3.3 in page 21). Additionally, it is worth mentioning that oversampling methods demands long time and computational complexity.

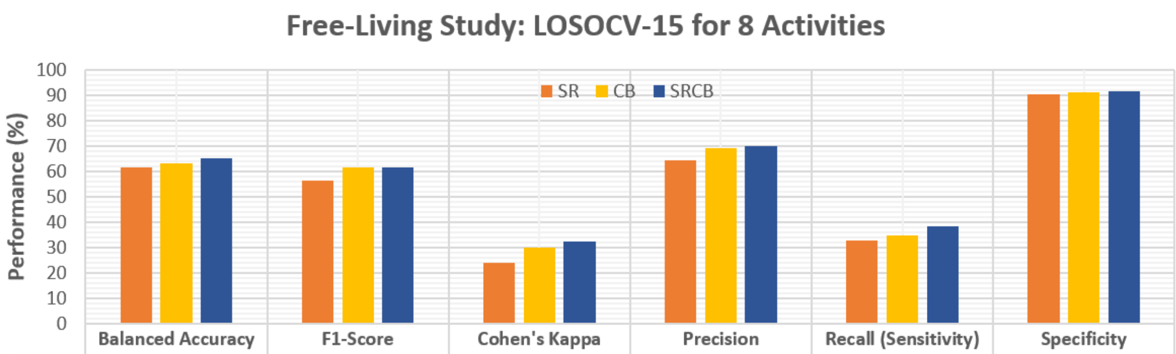
5.1.2. CLASSIFICATION

Regarding the classification algorithm, the GridSearchCV Random Forest was selected based on the results from the Controlled Study.

```
model = RandomForestClassifier (n_estimators=600, criterion='gini',
        max_depth=18, max_features = "auto", min_samples_leaf = 40,
        oob_score = True, n_jobs = -1)
```

5.1.3. EVALUATION LOSOCV-15 FOR 8 ACTIVITIES

After splitting the learning dataset into training and test set, the evaluation of predicting activities on the test set takes place. In this section, the learning model is evaluated for predicting 8 activities based on LOSOCV for 15 subjects, for both SR, CB and SRCB datasets in the Free-Living Study. Thus, the training set consists of data from 14 subjects, while the test set consists of 1 subject, repeating the validation for 15 times and calculating the average metric values. The evaluation is based on the confusion matrixes and the calculated metrics. The results for the validation of the system are presented in Figure 5.5.



	Balanced Accuracy	F1-Score	Cohen's Kappa	Precision	Recall (Sensitivity)	Specificity
SR	61.52	56.33	23.89	64.39	32.62	90.41
CB	63.11	61.63	30.16	69.03	34.83	91.39
SRCB	65.05	61.78	32.42	69.94	38.43	91.66

Figure 5.5: The learning model is evaluated for predicting 8 activities through different metrics. The validation is based on LOSOCV for 15 subjects from the Free-Living Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

Concerning the stingray (SR) dataset, the overall classification performance for predicting 8 activities is 56.33% (F1-score). Based on the confusion matrix (see Confusion Matrix at Appendix C.2 in page 102), we can see that many activities are misclassified. The dynamic sitting achieves the best prediction (85%) among all the activities. However, sitting (17%), standing (0%) and dynamic standing (11%) are misclassified with dynamic sitting. The reason for overfitting the dynamic sitting is due to the imbalanced dataset. Furthermore, walking (67%) is misclassified with walking upstairs and walking downstairs. Similarly, the insufficient number of data for the activities walking upstairs and downstairs affects the prediction model. Finally, cycling has a true positive score 71%.

For the chillband (CB) dataset, the overall classification performance for predicting 8 activities is 61.61% (F1-score). Similar to SR dataset, some activities are misclassified and due to imbalanced data the dynamic sitting and walking cause overfitting. However, among all the activities, the prediction model for the dynamic sitting performs significantly well with 90%. Based on the confusion matrix, the true positive score for sitting is 22%, standing 19%, dynamic standing 17%, walking upstairs 0%, walking downstairs 0%, walking 55%, and cycling 67%.

For the stingray-chillband (SRCB) dataset, the overall classification performance for predicting 8 activities is 61.78% (F1-score). Similar to SR and CB datasets, some activities are misclassified due to imbalanced data. However, the activities walking (68%) and cycling (73%) receive a better score compared to SR and CB datasets.

5.1.4. EVALUATION LOSOCV-15 FOR 5 ACTIVITIES

According to the evaluation for recognizing 8 activities, some activities, such as standing, dynamic standing, walking upstairs and walking downstairs, cannot be predicted accurately. For this reason, we will also investigate the activity recognition system for predicting 5 main activities, after merging the similar ones. These 5 activities consist of sitting, standing, walking on stairs, walking, and cycling. Consequently, the learning model is also evaluated for predicting 5 activities based on LOSOCV for 15 subjects, for both SR, CB and SRCB datasets in the Free-Living Study. The results for the validation of the system are presented in Figure 5.6. See also Confusion Matrix at Appendix C.2 in page 102).

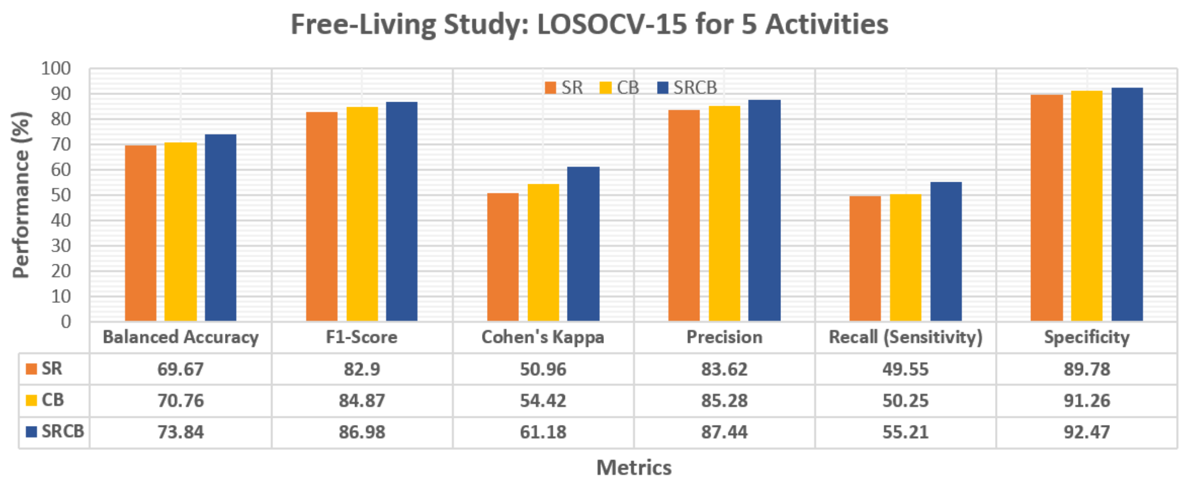


Figure 5.6: The learning model is evaluated for predicting 5 activities through different metrics. The validation is based on LOSOCV for 15 subjects from the Free-Living Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

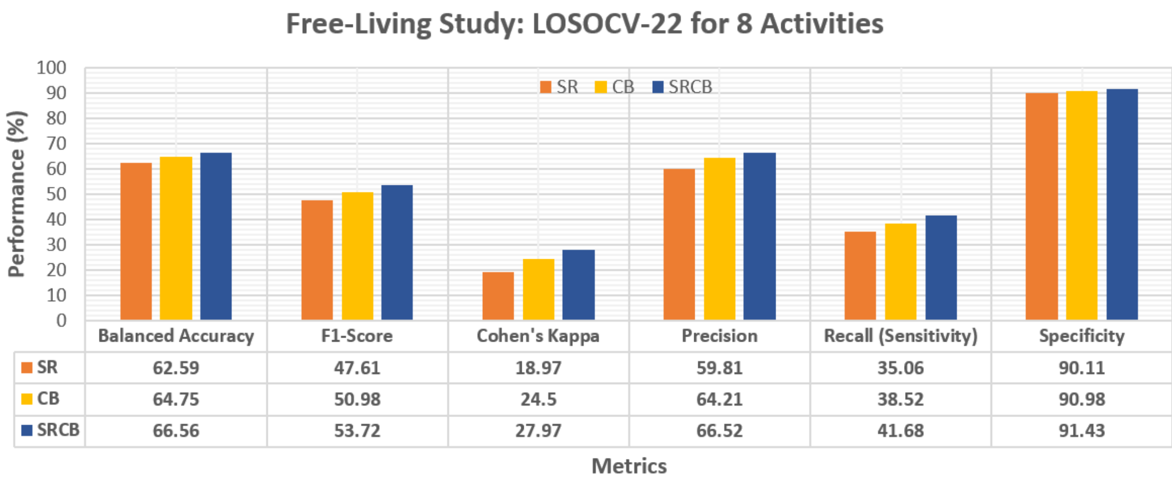
Regarding the stingray (SR) dataset, the overall classification performance for predicting 5 activities accurately is 82.9% (F1-score). The performance for sitting is significant accurate with 98% true positive score. However, the activity standing (11%) is misclassified with sitting, and the activity walking on stairs is misclassified with walking (0% and 67%, respectively). Finally, the score for cycling is 71%.

For the chillband (CB) dataset, the overall classification performance for predicting 5 activities is 84.87% (F1-score). Similar to SR dataset, the activity sitting can be predicted accurately with 97% score, while standing (29%) is misclassified to sitting. The activity walking on stairs is misclassified to walking (55%) and cycling receives a 67% score. In contrast to SR, the prediction score for walking and cycling is reduced.

Compared to SR and CB, the classification performance for the SRCB is enhanced to 86.98% (F1-score) for recognizing all the 5 activities. Similar to SR and CB, standing has been misclassified to sitting, and walking on stairs has been wrongly predicted as walking.

5.2. VALIDATION MODEL A

The validation model A for the Free-Living Study consists of data from the additional 22 subjects (15 out of 37 subjects have been used for the learning model). This model is divided into training and test set, based on LOSOCV for 22 subjects, in order to evaluate the system's performance for recognizing the unseen activities of the test set. Thus, the training set consists of data from 21 subjects, while the test set consists of 1 subject, repeating the validation for 22 times and calculating the average metrics. Furthermore, data from the learning model (15 subjects) are added to the training set in order to enhance the training (36 subjects in total). For the final validation of our system, two different approaches will be investigated for recognizing activities in an uncontrolled environment. The first is based on predicting 8 activities and the second one is based on predicting 5 activities. The results of the validation model A are presented in Figure 5.7. For further understanding, see also the actual versus predicted activities for the activity recognition model at Appendix D in page 109 (the plot is made based on the most highly scored feature for each sensor).



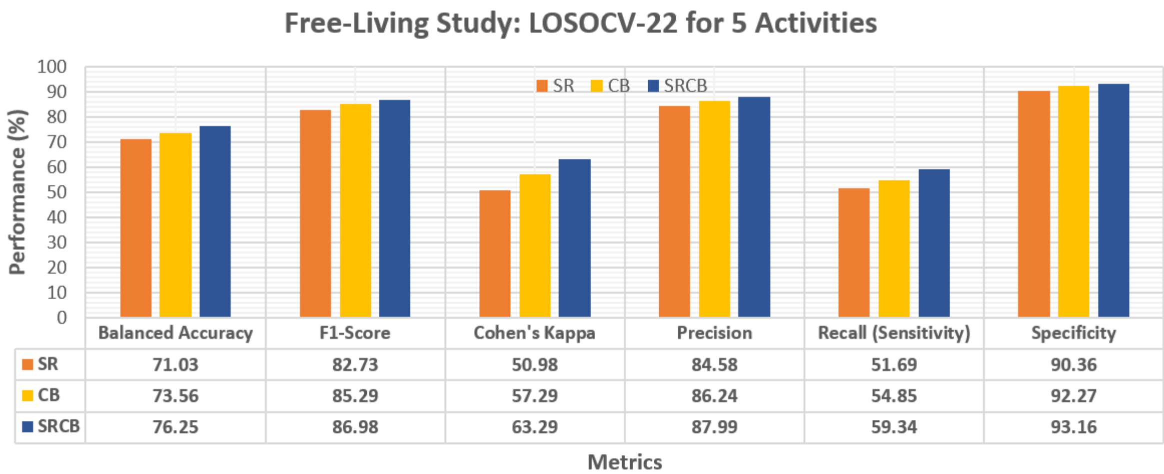


Figure 5.7: The validation model A is evaluated for predicting 8 activities (top graph) and 5 activities (bottom graph) through different metrics. The validation is based on LOSOCV for 22 subjects from the Free-Living Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

5.2.1. STINGRAY

Regarding the stingray (SR) sensor, the overall classification performance for predicting 8 activities is 47.61% (F1-score). Similar to the learning model, many activities are misclassified due to the imbalanced data. Based on the confusion matrix (see Figure 5.8), the three most highly predicted classes are the activity walking (74%), cycling (66%) and dynamic sitting (61%).

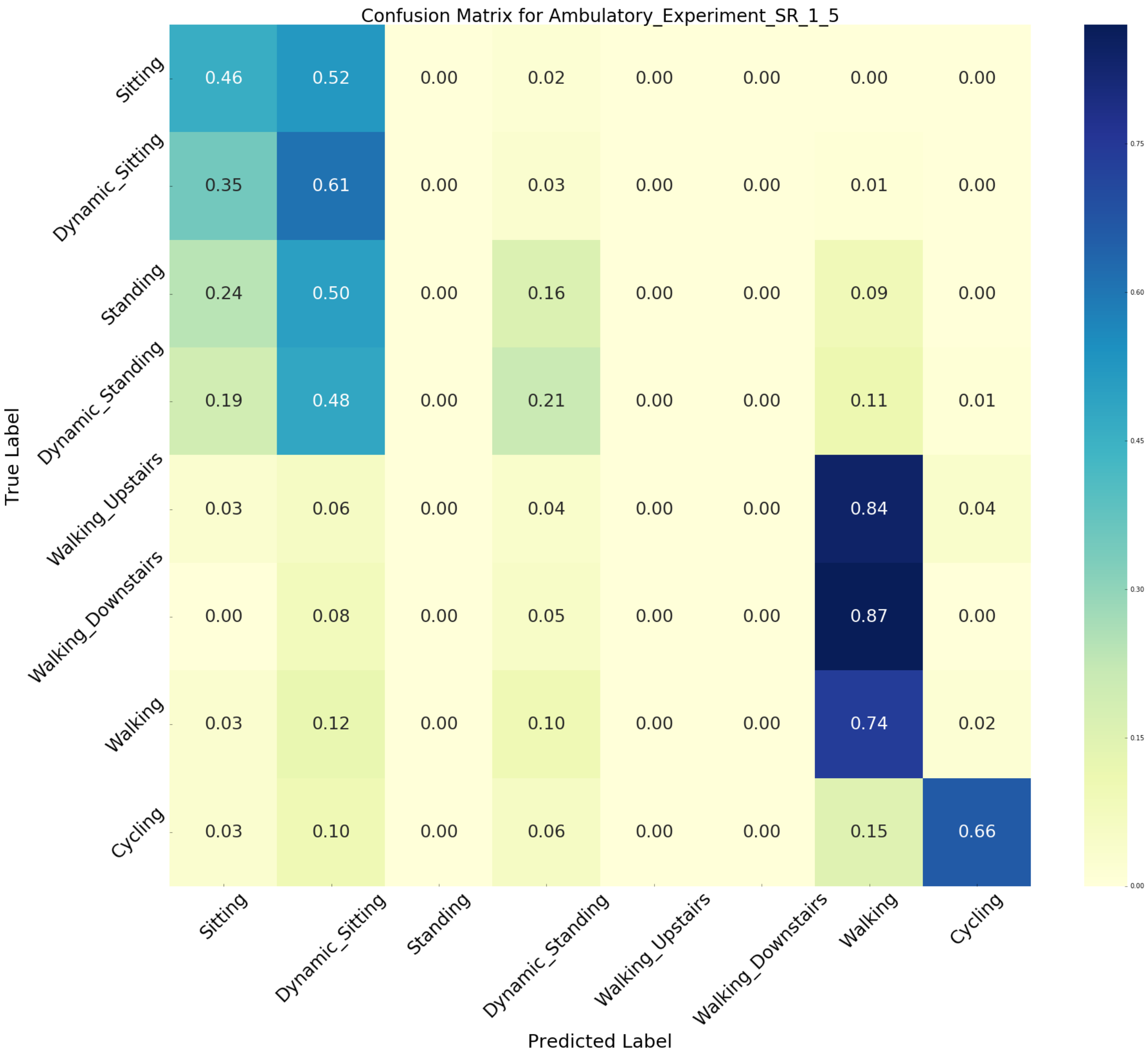


Figure 5.8: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 8 activities.

The overall classification performance for predicting 5 activities is 82.73% (F1-score). After merging similar activities, the classification performance has been significantly improved. The class sitting receives the most accurate prediction. However, standing is still misclassified with sitting, while walking on stairs is totally misclassified with walking. Based on the confusion matrix (see Figure 5.9), the true positive score for sitting is 97%, standing 20%, walking on stairs 0%, walking 74%, and cycling 66%.

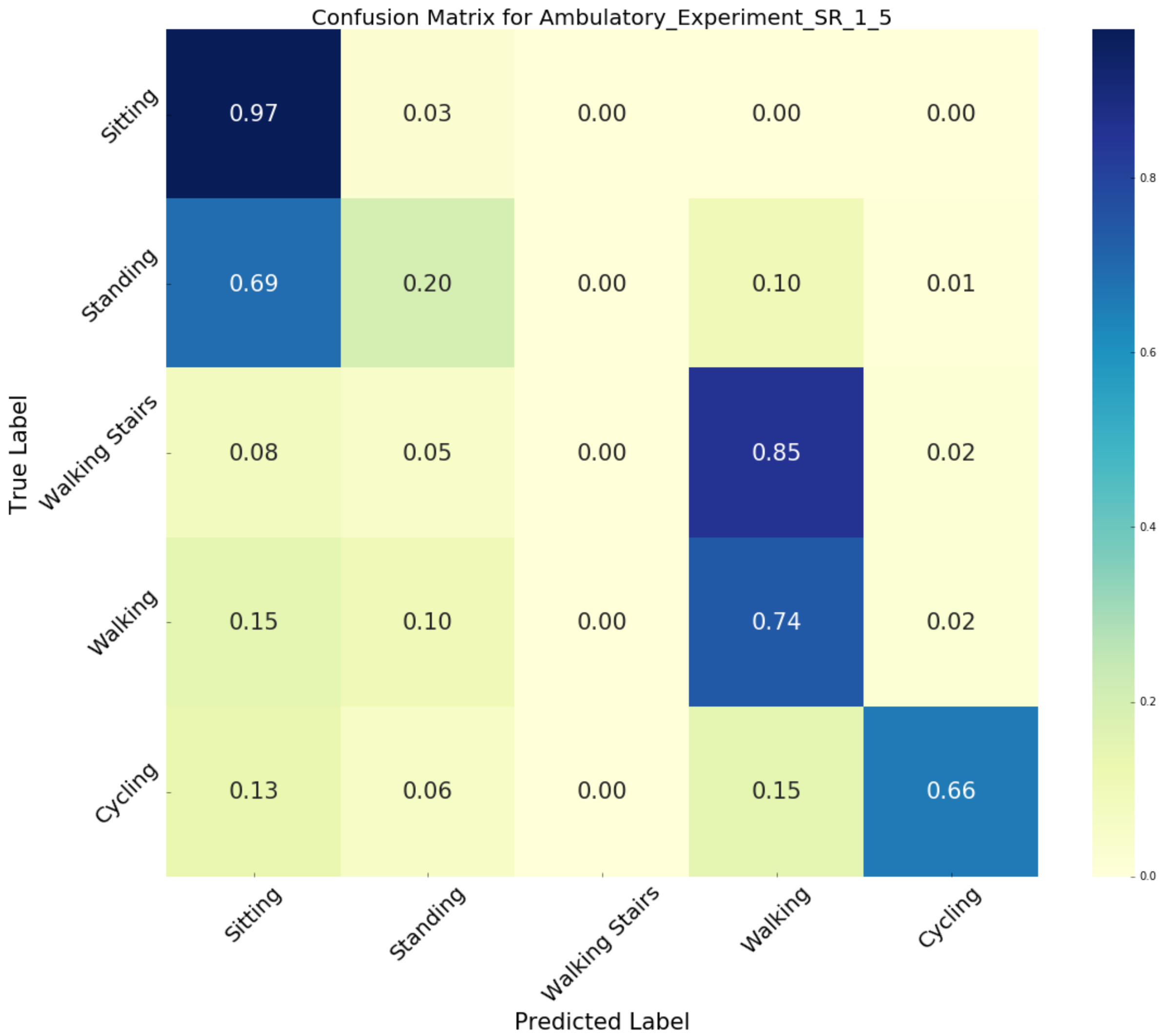


Figure 5.9: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 5 activities.

5.2.2. CHILLBAND

Concerning the chillband (CB) sensor, the overall classification performance for predicting 8 activities is 50.98% (F1-score). Compared to SR, the wrist-worn sensor performs slightly better. However, many activities are still misclassified due to the imbalanced data. Based on the confusion matrix (see Figure 5.10), the three most highly predicted classes are the activity dynamic sitting (67%), cycling (66%) and walking (62%).

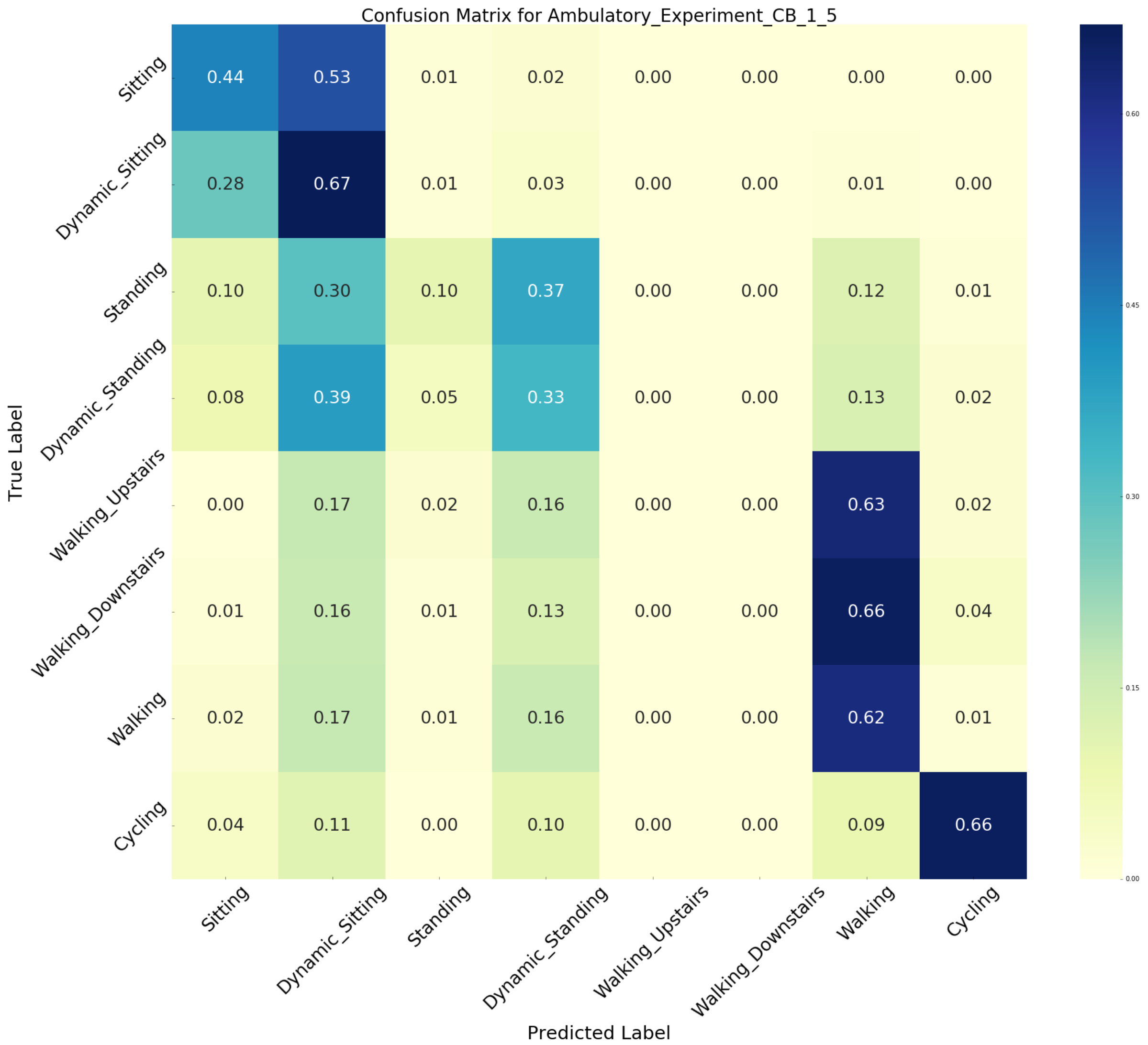


Figure 5.10: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 8 activities.

The overall classification performance for predicting 5 activities is 85.29% (F1-score). After merging similar activities, the classification performance has been significantly improved. The class sitting receives the most accurate prediction. However, standing is still misclassified with sitting, while walking on stairs is completely misclassified with walking. Based on the confusion matrix (see Figure 5.11), the true positive score for sitting is 96%, standing 40%, walking on stairs 0%, walking 62%, and cycling 66%.

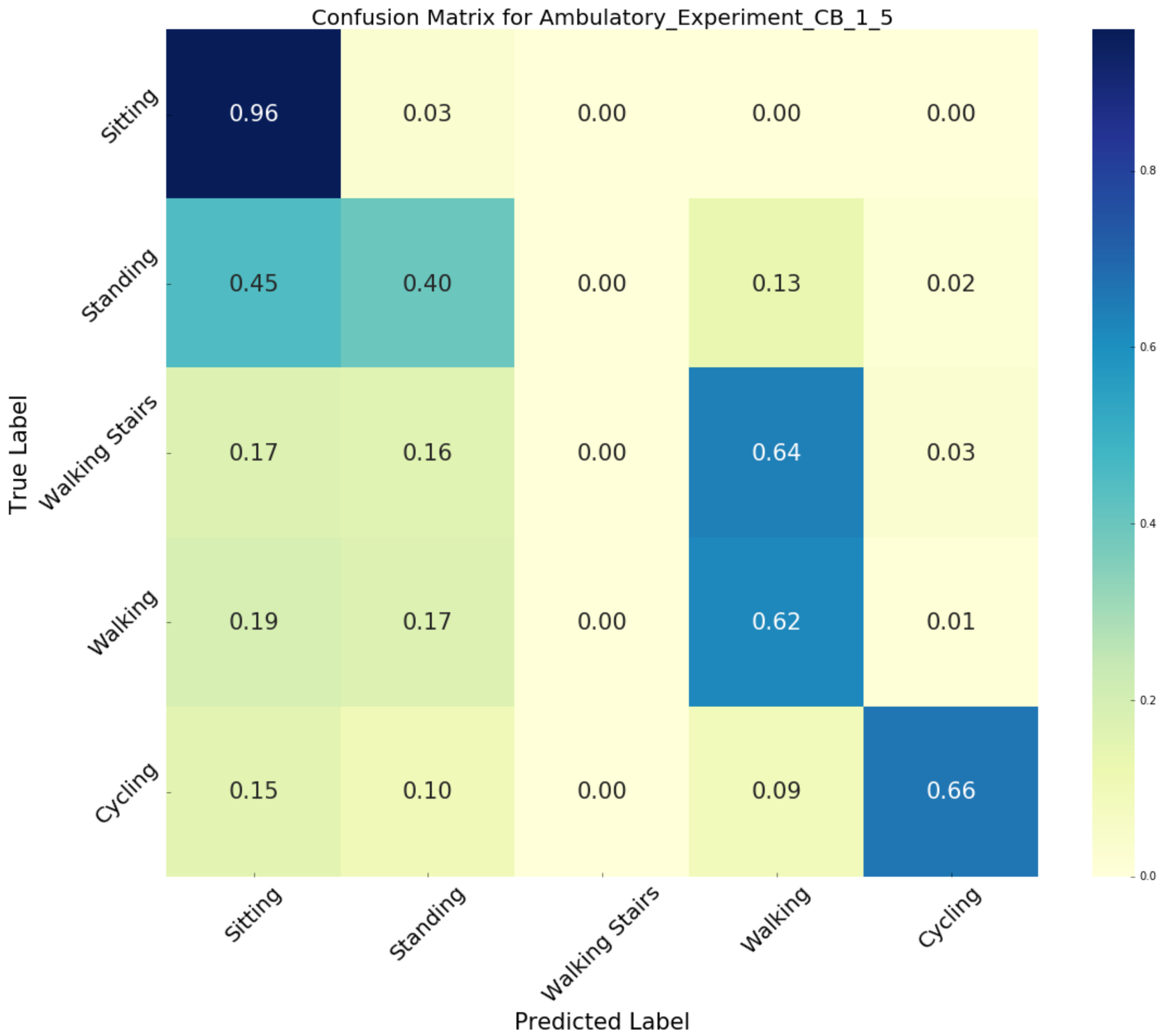


Figure 5.11: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 5 activities.

5.2.3. STINGRAY & CHILLBAND

The activity recognition based on the stingray-chillband (SRCB) sensor achieves the best classification performance for predicting 8 activities, which is 53.72% (F-score). Compared to SR and CB, the prediction for dynamic sitting (71%) and cycling (74%) has been improved. However, many activities are still misclassified due to the imbalanced data. Based on the confusion matrix (see Figure 5.12), the true positive score for sitting is 46%, standing 10%, dynamic standing 36%, walking on stairs 0%, and walking 74%.

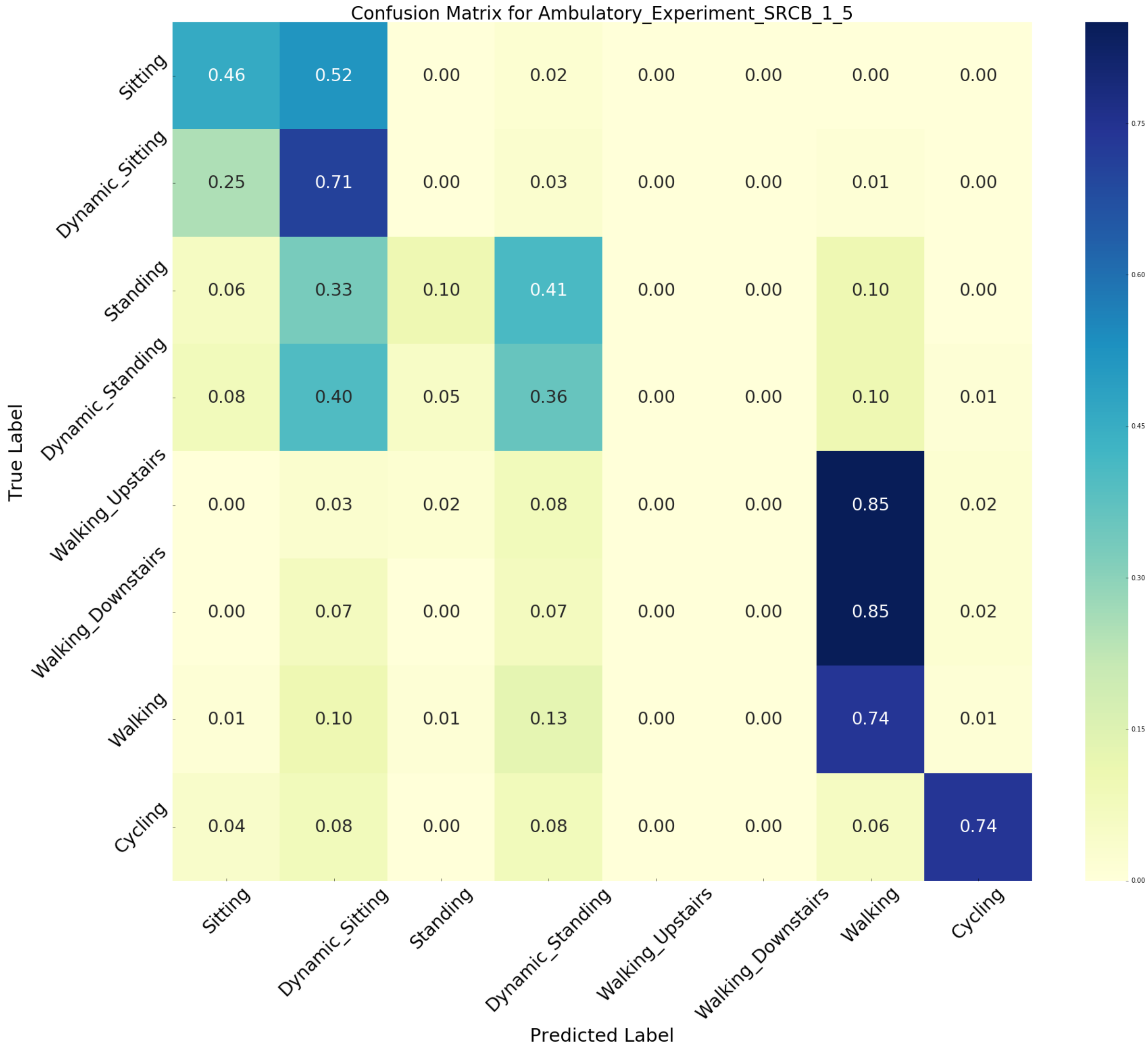


Figure 5.12: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 8 activities.

After merging the similar activities, the overall classification performance for predicting 5 activities is 86.98% (F-score). Thus, SRCB accelerometer outperforms the SR, followed by the CB. Based on the confusion matrix (see Figure 5.13), the score for the predicted activities is: sitting 97%, standing 43%, walking on stairs 0%, walking 74%, and cycling 74%.

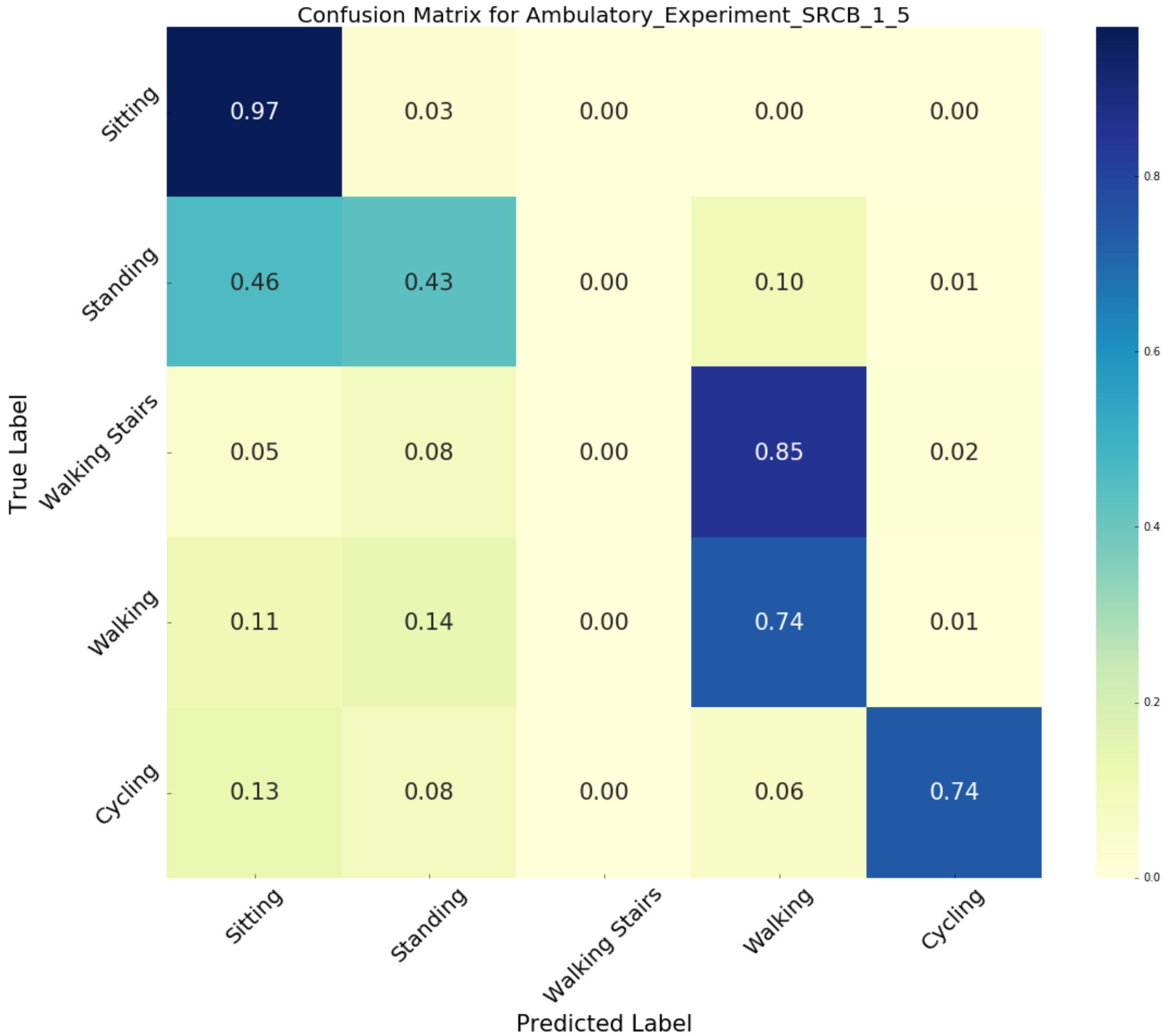


Figure 5.13: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 5 activities.

5.3. VALIDATION MODEL B

The validation model B will assess the classification performance of the activities performed in the Free-Living Study from 22 subjects, using data from the Controlled Study for the training model. The training set consists of data from 40 subjects of the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. For the final validation of our system, two different approaches will be investigated for recognizing activities in an uncontrolled environment. The first is based on predicting 8 activities and the second one is based on predicting 5 activities. The results for the validation model B are presented in Figure 5.14.

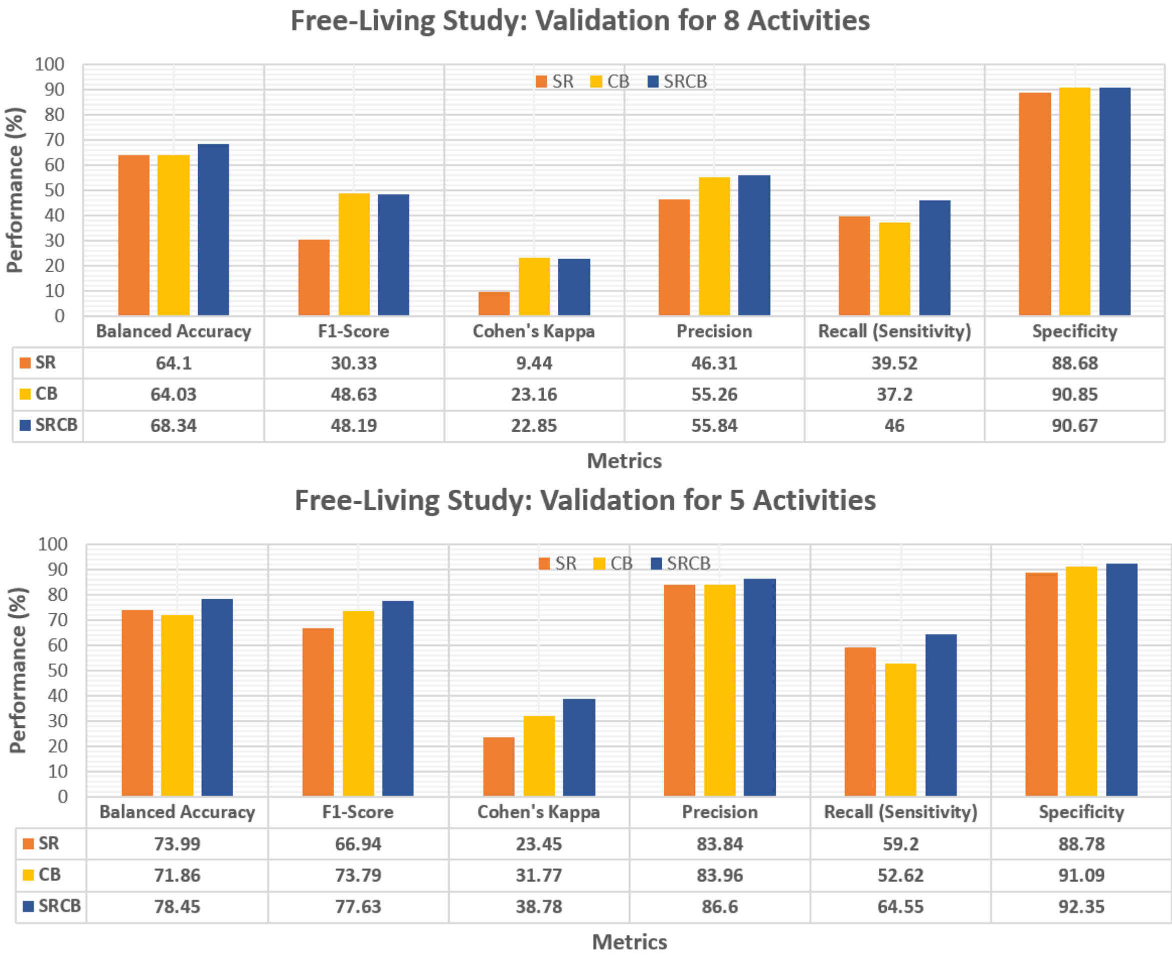


Figure 5.14: The validation model B is evaluated for predicting 8 activities (top graph) and 5 activities (bottom graph) through different metrics. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

5.3.1. STINGRAY

Regarding the stingray (SR) sensor, the overall classification performance for predicting 8 activities is 30.33% (F1-score). Compared to validation model A, some classes perform better, such as dynamic standing (40%), walking downstairs (62%), and cycling 76%, but in total this model performs worse. Despite the balanced dataset from the Controlled Study that is used for the training set, the performed activities differ with the ones from the Free-Living Study, resulting in activities misclassification. Based on the confusion matrix (see Figure 5.15), the prediction score is: sitting (16%), dynamic sitting (43%), standing (17%), dynamic standing (40%), walking upstairs (24%), walking downstairs (62%), walking (38%) and cycling (76%).

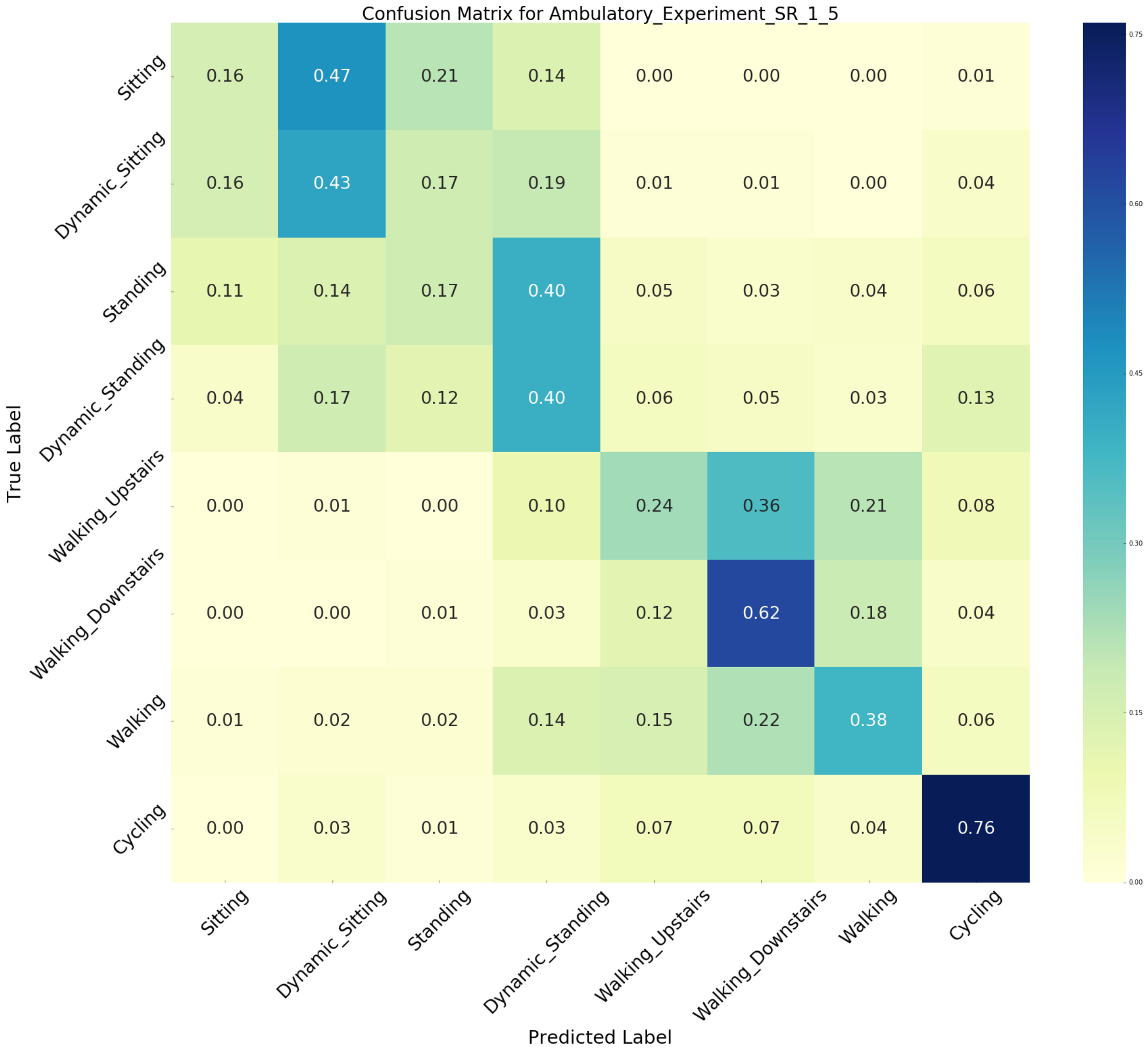


Figure 5.15: The confusion matrix for the stingray (SR) sensor is presented. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. The validation is for the Free-Living Study, by detecting 8 activities.

The overall classification performance for predicting 5 activities is 66.94% (F1-score). After merging similar activities, the classification performance has been significantly improved. Overall, model B did not outperform the model A. However, the activities standing, walking on stairs, and cycling are predicted more accurately, compared to model A. Based on the confusion matrix (see Figure 5.16), the prediction score is: sitting (61%), standing (53%), walking on stairs (67%), walking (38%) and cycling (76%).

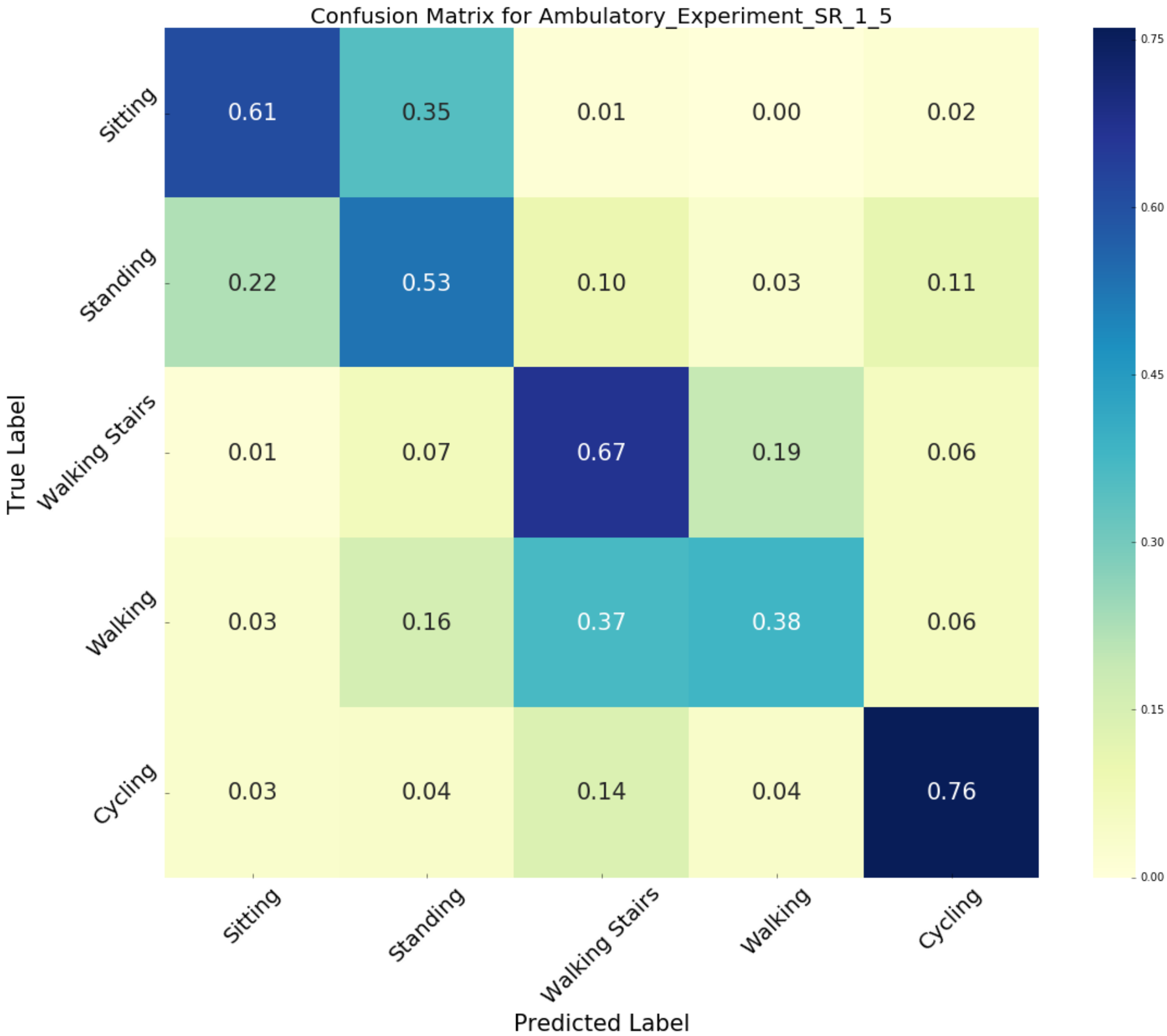


Figure 5.16: The confusion matrix for the stingray (SR) sensor is presented. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. The validation is for the Free-Living Study, by detecting 5 activities.

5.3.2. CHILLBAND

For the chillband (CB) sensor, the overall classification performance for predicting 8 activities is 48.63% (F1-score). CB performs better than SR. However, compared to validation model A, this model performs slightly worse (except the activity sitting). Based on the confusion matrix (see Figure 5.17), the prediction score is: sitting (55%), dynamic sitting (34%), standing (25%), dynamic standing (30%), walking upstairs (14%), walking downstairs (21%), walking (49%) and cycling (70%).

Figure 5.17: The confusion matrix for the chillband (CB) sensor is presented. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. The validation is for the Free-Living Study, by detecting 8 activities.

The overall classification performance for predicting 5 activities is 73.79% (F1-score). After merging similar activities, the classification performance has been significantly improved. Based on the confusion matrix (see Figure 5.18), the true positive score for sitting is 73%, standing 43%, walking on stairs 28%, walking 49%, and cycling 70%.

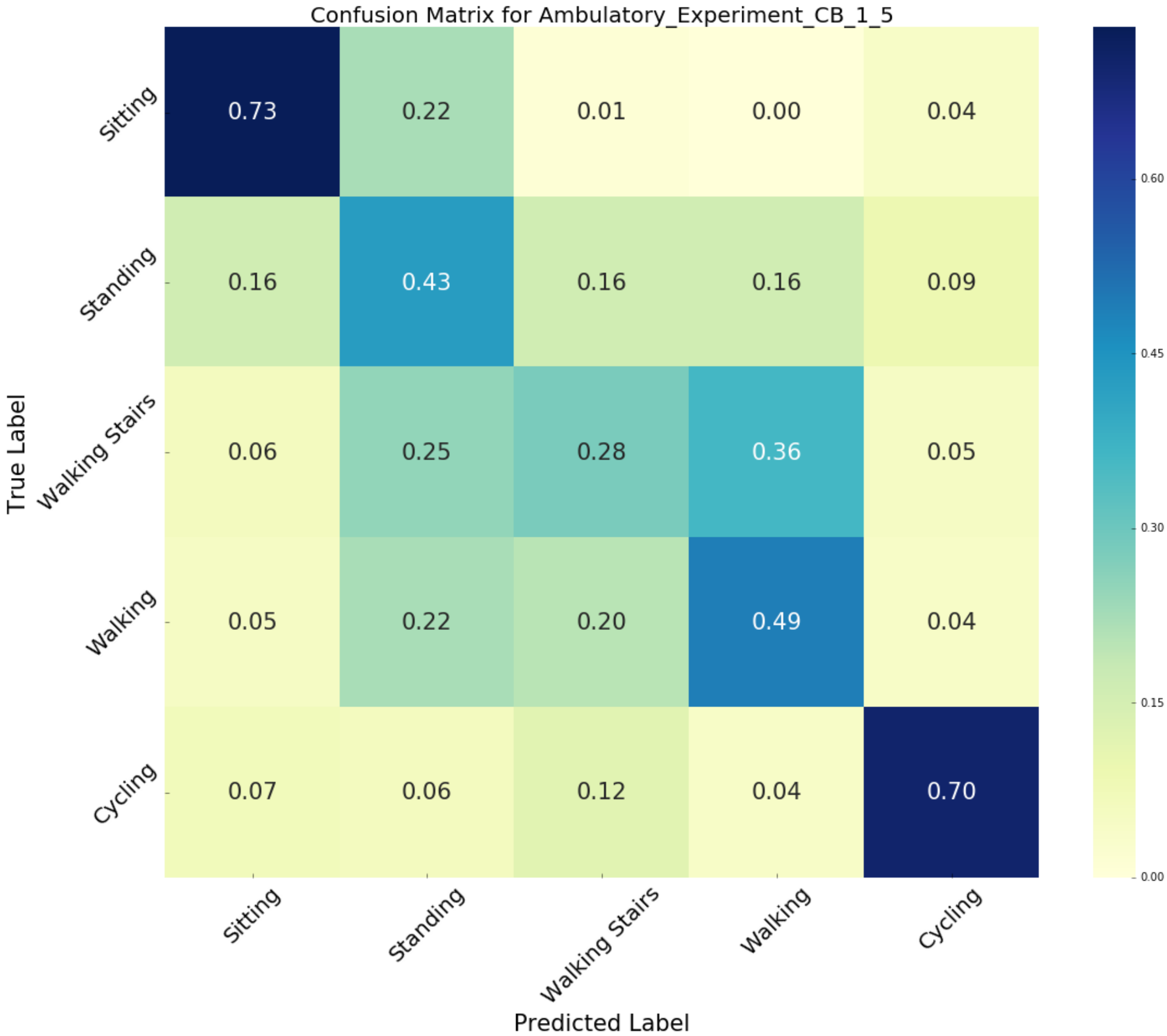


Figure 5.18: The confusion matrix for the chillband (CB) sensor is presented. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. The validation is for the Free-Living Study, by detecting 5 activities.

5.3.3. STINGRAY & CHILLBAND

The activity recognition based on the stingray-chillband (SRCB) sensor achieves 48.19% (F-score), which is better compared to SR and slightly worse to CB. That means that accelerometer data from the stingray sensor does not improve the classification performance, while the wrist-worn sensor, as a standalone device, performs better in predicting activities in the Free-Living environment. Based on the confusion matrix (see Figure 5.19), the prediction score is: sitting (48%), dynamic sitting (42%), standing (29%), dynamic standing (37%), walking upstairs (33%), walking downstairs (53%), walking (45%) and cycling (83%).

Figure 5.19: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. The validation is for the Free-Living Study, by detecting 8 activities.

The overall classification performance for predicting 5 activities is 77.63% (F-score). Thus, SRCB accelerometer outperforms the SR and CB, however performs worse than model A. Based on the confusion matrix (see Figure 5.20), the score for the predicted activities is: sitting 77%, standing 53%, walking on stairs 65%, walking 45%, and cycling 83%.

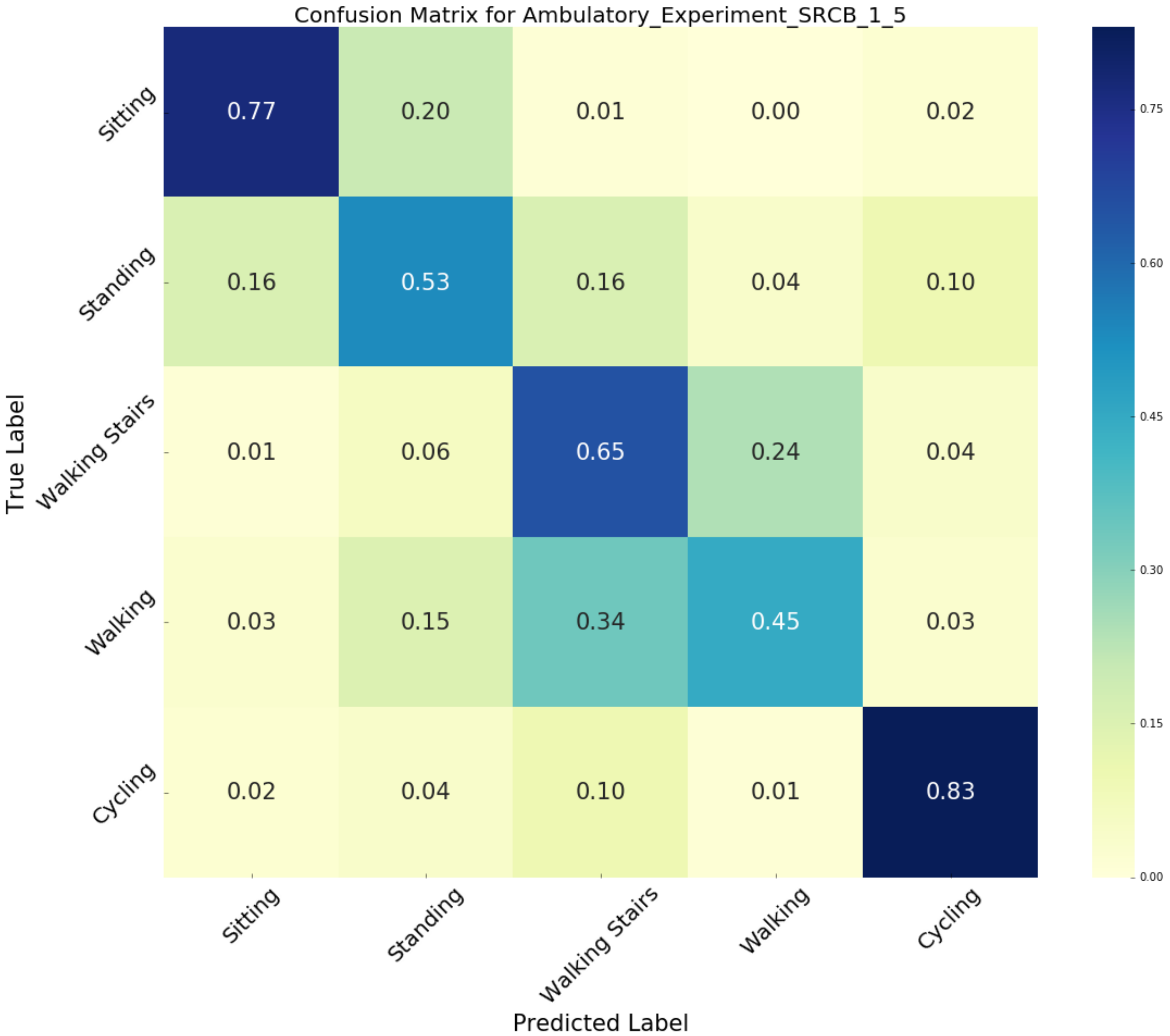


Figure 5.20: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The training set consists of 40 subjects from the Controlled Study, while the test set consists of 22 subjects from the Free-Living Study. The validation is for the Free-Living Study, by detecting 5 activities.

5.4. VALIDATION MODEL C

The validation model C is a combination of the previous models A and B and uses data from both the Controlled and Free-Living Study for the training the classifier in order to validate the prediction for the activities performed in the uncontrolled environment. Overall, model C performs better compared to model B, but slightly worse compared to model A. This model is divided into training and test set, based on LOSOCV for 22 subjects, in order to evaluate the system's performance for recognizing the unseen activities on the test set. Thus, the training set consists of data from 21 subjects (Free-Living Study), while the test set consists of 1 subject, repeating the validation for 22 times and calculating the average metrics. Additionally, data from the learning model (15 subjects from the Free-Living Study) but also data from 40 subjects of the Controlled Study are added to the training set in order to enhance the training. It is worth mentioning that for every subject of the test set (22 subjects from the Free-Living Study in total), the similar subject from the training set that belongs to the Controlled Study is excluded in order to avoid any correlation issues that may lead to overfitting. The reason for that is that a subject might have performed the activities in the Controlled and Free-Living Study in a similar way. For the final validation of our system, two different approaches will be investigated for recognizing activities in an uncontrolled environment. The first is based on predicting 8 activities and the second one is based on predicting 5 activities. The results for the validation model C are presented in Figure 5.21.

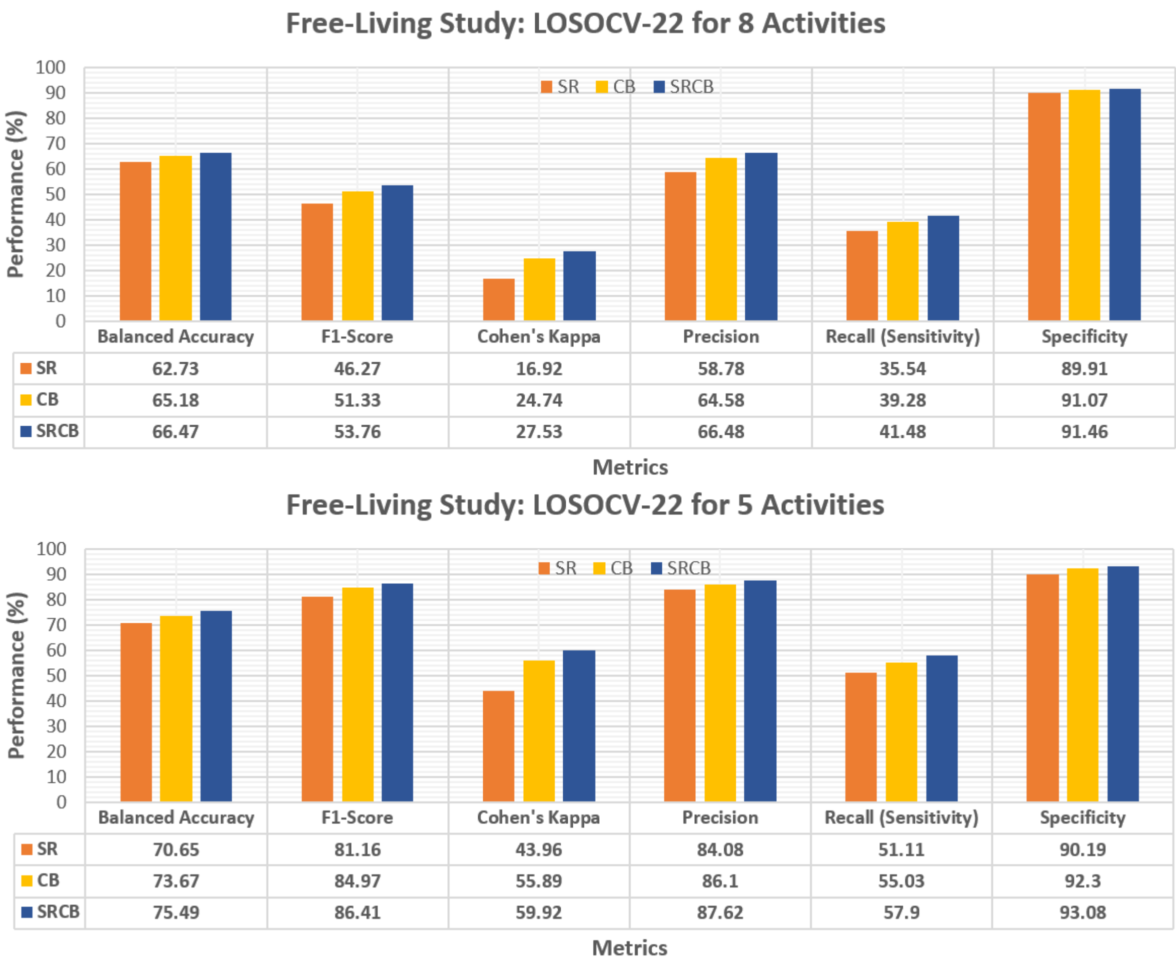


Figure 5.21: The validation model C is evaluated for predicting 8 activities (top graph) and 5 activities (bottom graph) through different metrics. The validation is based on LOSOCV for 22 subjects from the Free-Living Study. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study. SR dataset represents the stingray sensor, CB dataset represents the chillband sensor, and SRCB dataset represents the combination of stingray and chillband sensors.

5.4.1. STINGRAY

Regarding the stingray (SR) sensor, the overall classification performance for predicting 8 activities is 46.27% (F1-score). Similar to model A, some activities are misclassified. Based on the confusion matrix (see Figure 5.22), the prediction score is: sitting (46%), dynamic sitting (54%), standing (1%), dynamic standing (24%), walking upstairs (6%), walking downstairs (18%), walking (62%) and cycling (60%).

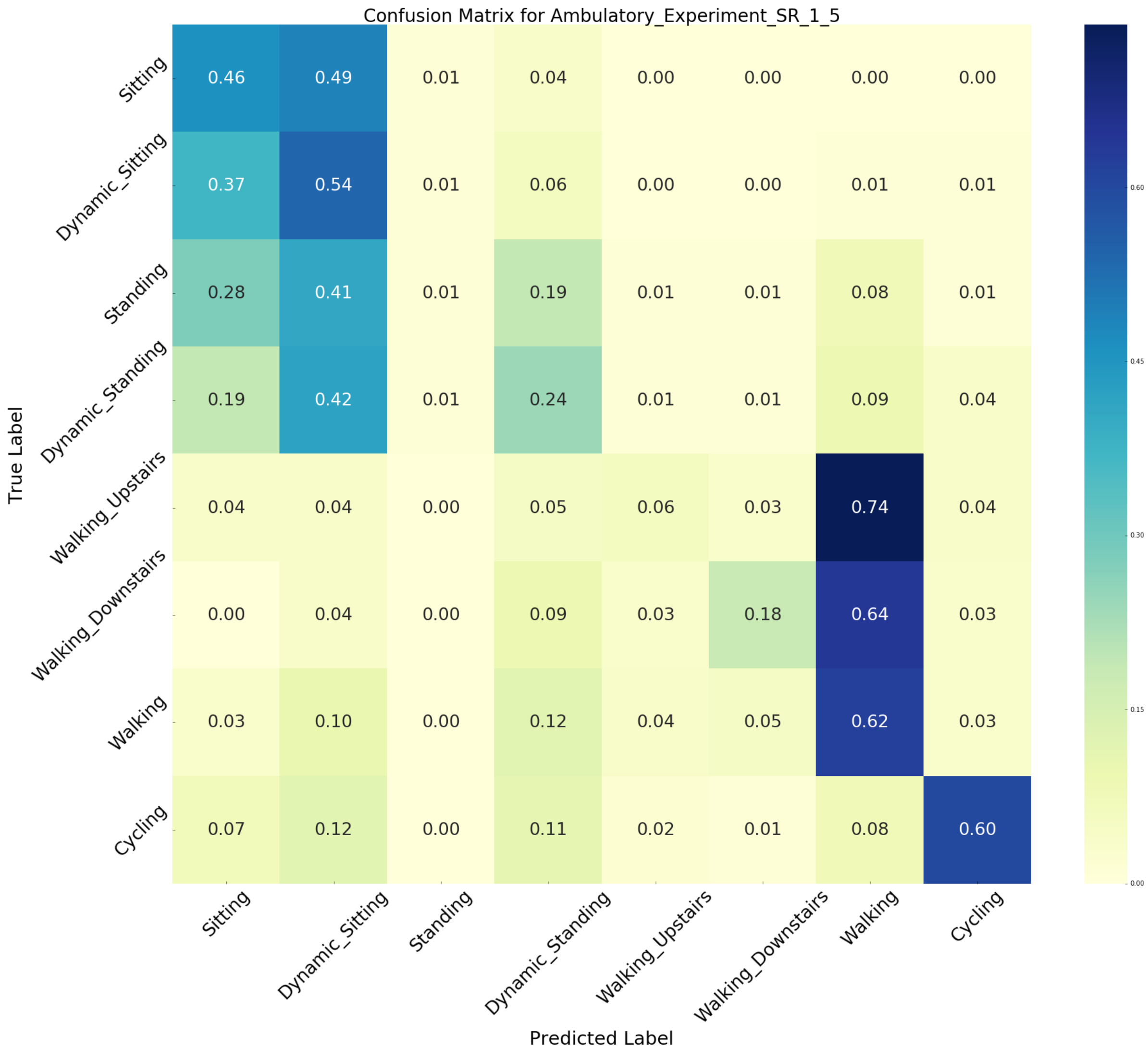


Figure 5.22: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 8 activities. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study.

The overall classification performance for predicting 5 activities is 81.16% (F1-score). After merging similar activities, the classification performance has been significantly improved. The class sitting receives the most accurate prediction (93%). However, standing is still misclassified with sitting, while walking on stairs is misclassified with walking. Based on the confusion matrix (see Figure 5.23), the true positive score for standing 23%, walking on stairs 15%, walking 62%, and cycling 60%.

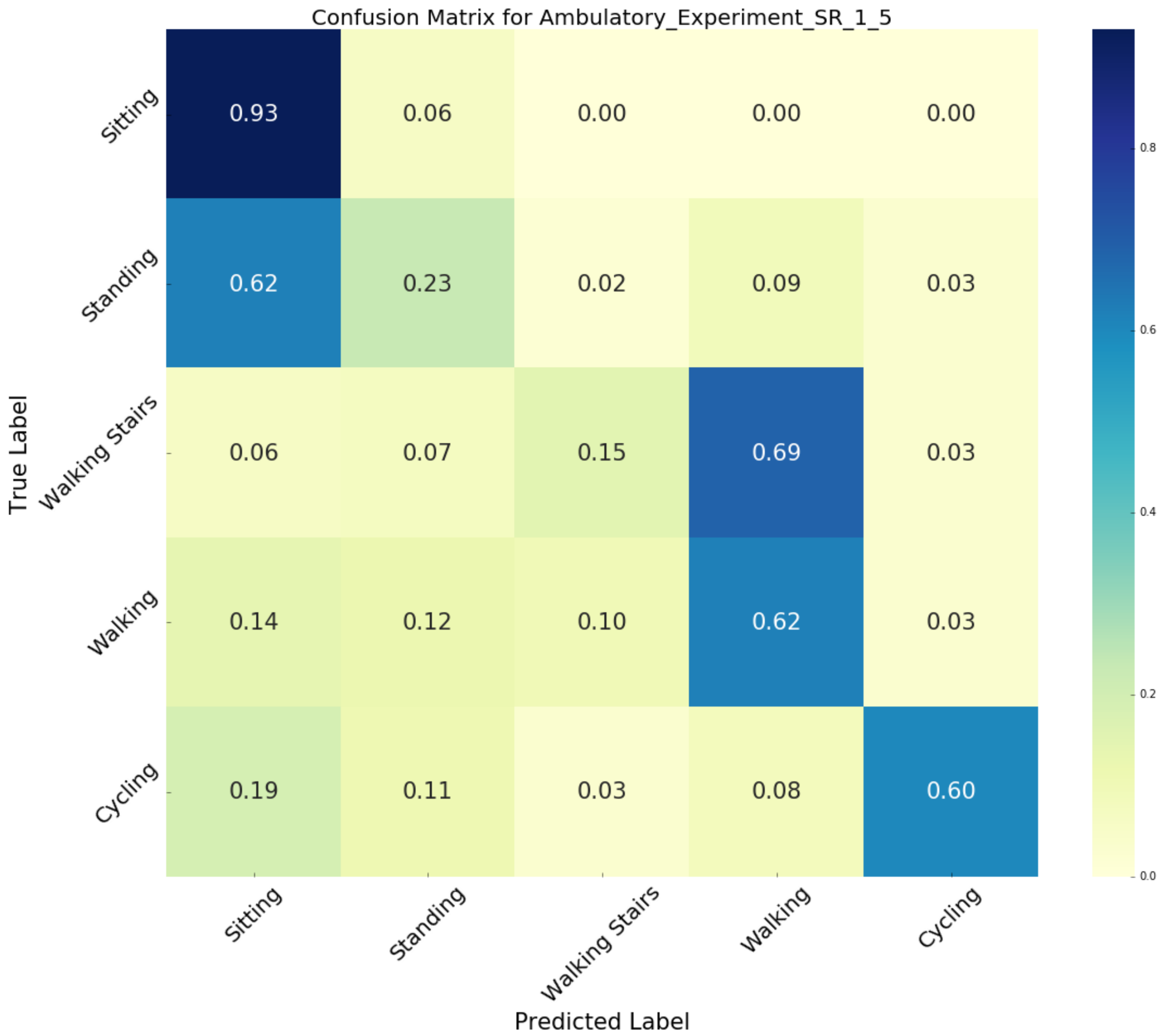


Figure 5.23: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 5 activities. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study.

5.4.2. CHILLBAND

Concerning the chillband (CB) sensor, the overall classification performance for predicting 8 activities is 51.33% (F1-score). For this sensor, model C performs slightly better than model A. In particular, the wrist-worn sensor performs better than the SR. However, some activities are still misclassified. Based on the confusion matrix (see Figure 5.24), the prediction score is: sitting (44%), dynamic sitting (67%), standing (20%), dynamic standing (31%), walking upstairs (3%), walking downstairs (10%), walking (56%) and cycling (68%).

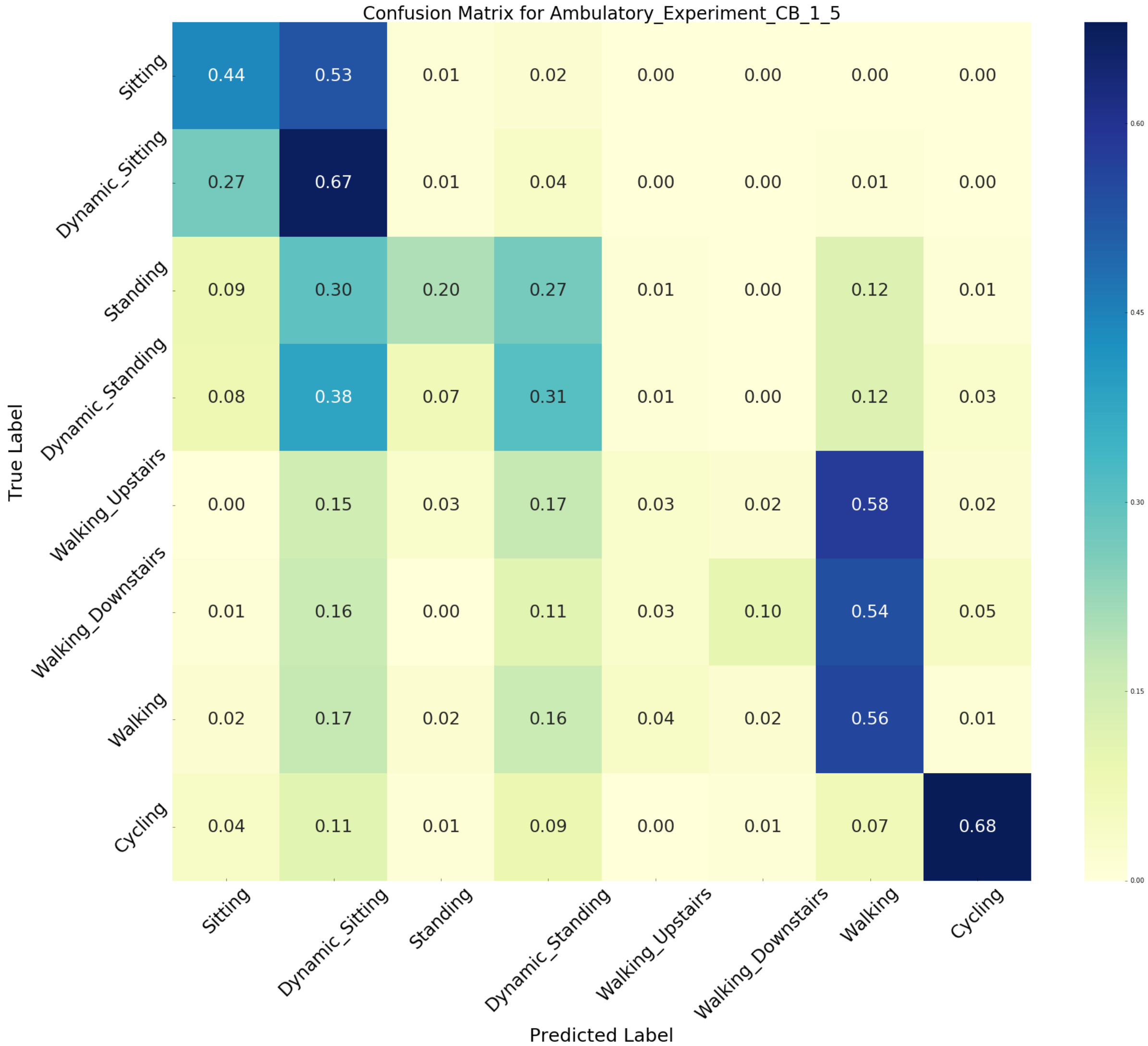


Figure 5.24: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 8 activities. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study.

The overall classification performance for predicting 5 activities is 84.97% (F1-score). After merging similar activities, the classification performance has been significantly improved. The class sitting receives the most accurate prediction (96%). However, standing is still misclassified with sitting, while walking on stairs is totally misclassified with walking. Based on the confusion matrix (see Figure 5.25), the true positive score for standing 40%, walking on stairs 9%, walking 56%, and cycling 68%.

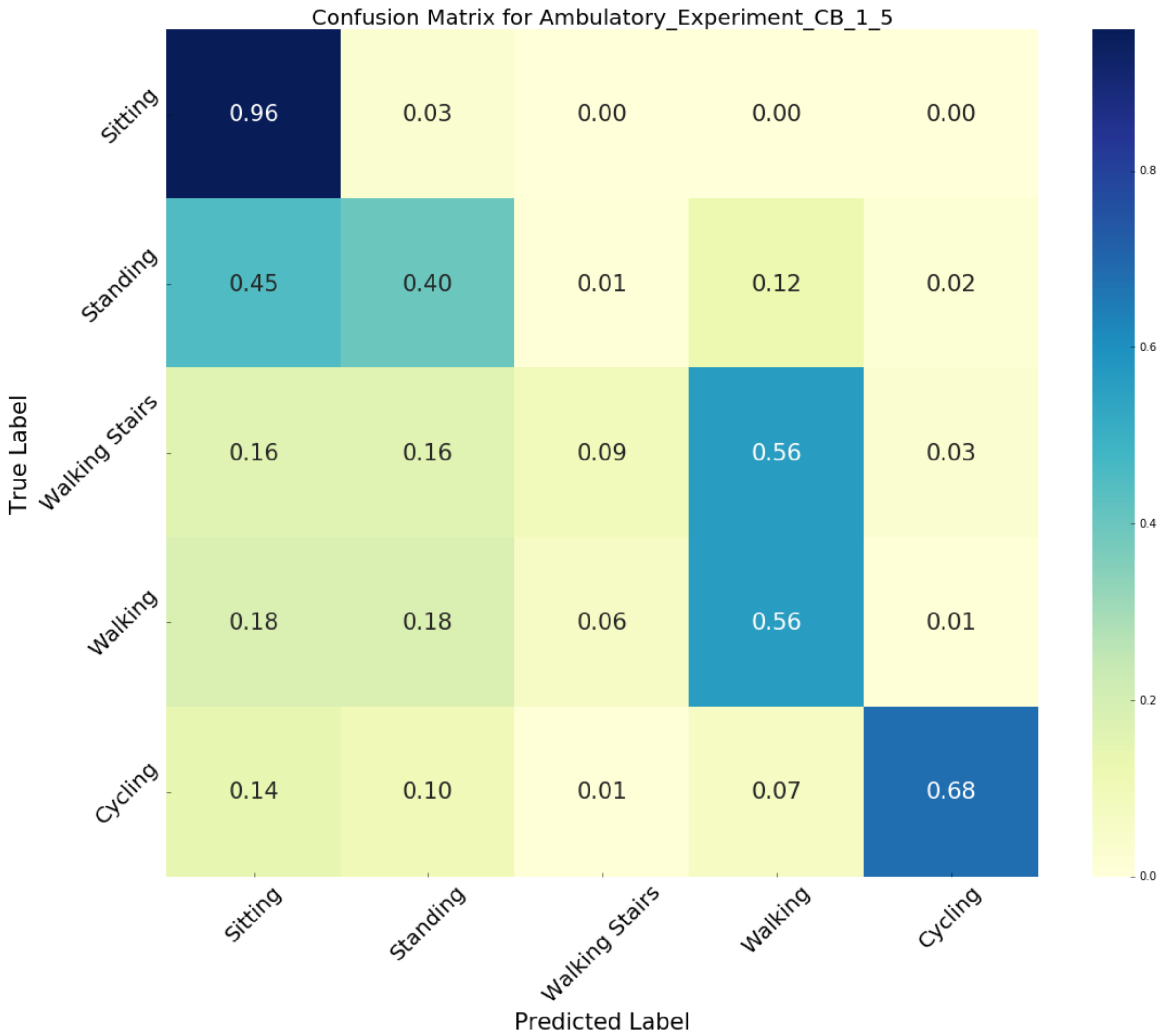


Figure 5.25: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 5 activities. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study.

5.4.3. STINGRAY & CHILLBAND

The activity recognition based on the stingray-chillband (SRCB) sensor achieves the best classification performance for predicting 8 activities, which is 53.76% (F-score). Compared to SR and CB, the prediction for dynamic sitting (71%) and walking (65%) has been improved. Compared to model A, the activities walking and cycling perform worse. Based on the confusion matrix (see Figure 5.26), the prediction score is: sitting (46%), dynamic sitting (71%), standing (21%), dynamic standing (35%), walking upstairs (4%), walking downstairs (8%), walking (65%) and cycling (67%).

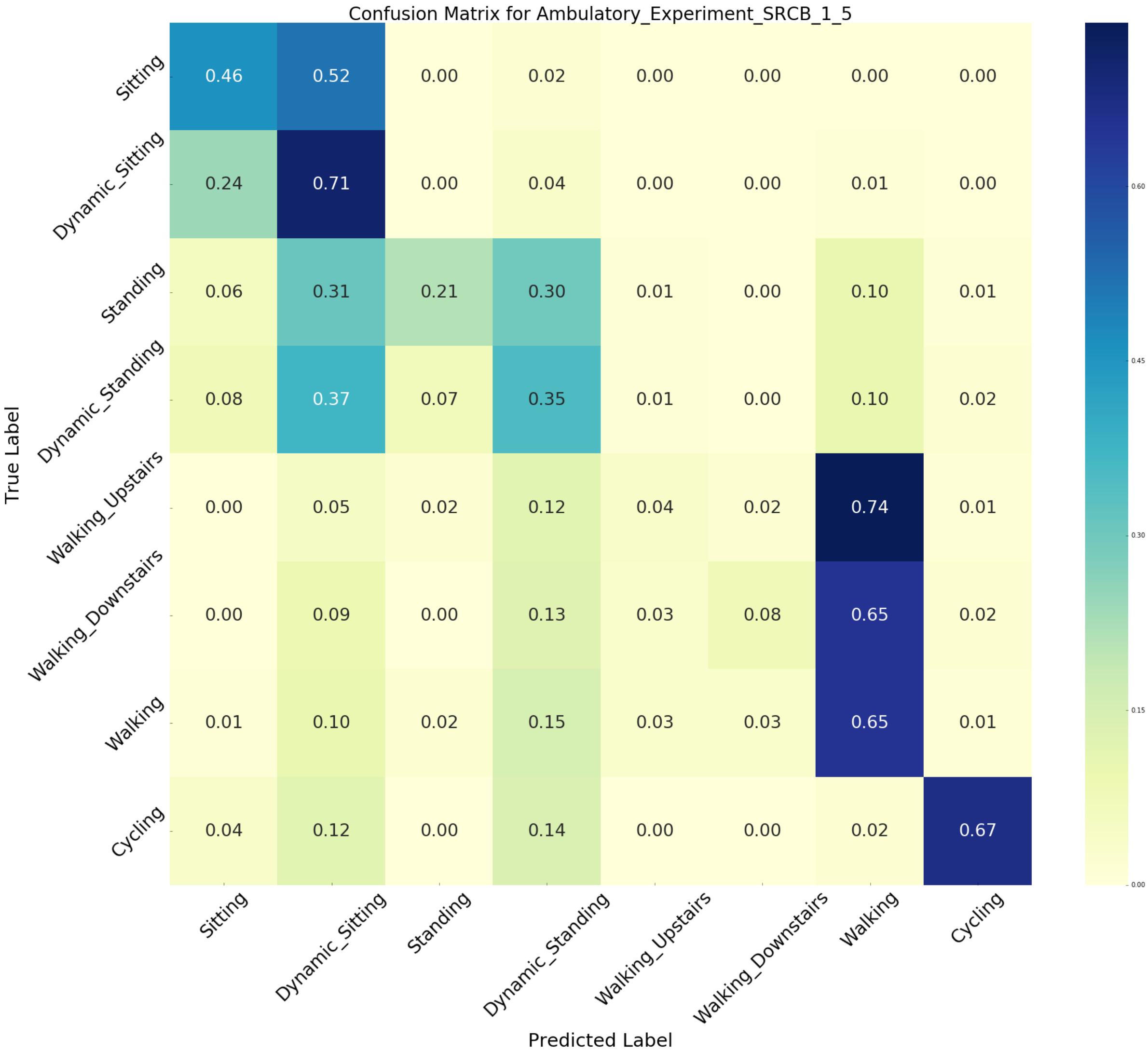


Figure 5.26: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 8 activities. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study.

The overall classification performance for predicting 5 activities is 86.41% (F-score). Thus, SRCB accelerometer outperforms the SR and CB. Based on the confusion matrix (see Figure 5.27), the score for the predicted activities is: sitting 96%, standing 44%, walking on stairs 8%, walking 65%, and cycling 67%.

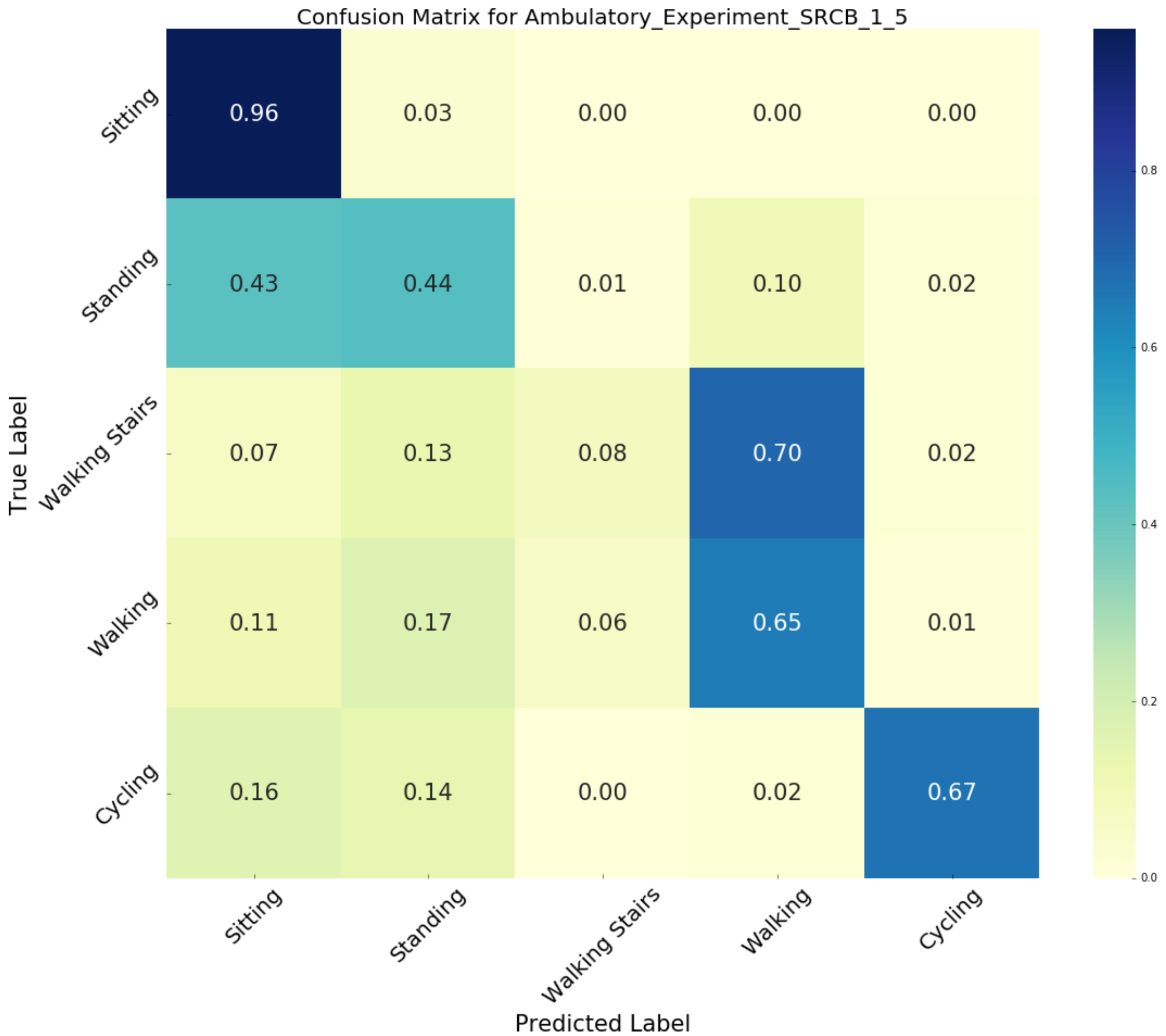


Figure 5.27: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-22 for the Free-Living Study, by detecting 5 activities. The training set consists of 40 subjects from the Controlled Study and 15 subjects from the Free-Living Study, while the test set consists of 22 subjects from the Free-Living Study.

5.5. DISCUSSION

The Free-Living Study includes data from 37 Subjects. Data from 15 subjects were used to train the learning model. Concerning the data processing, we used the same parameters with the ones from the Controlled Study. The reason for selecting the same parameters concerning the window segment, feature selection, and classification algorithm is to make a valid comparison between the two studies. Otherwise, it would be impossible to investigate if there is any correlation between the activities performed in controlled and uncontrolled environments. However, a further examination of these parameters could enhance the prediction of activities performed in the Free-Living Study. For instance, the performance might be improved by selecting another window segment or another classification algorithm. Furthermore, we investigated three different types of annotation for the performed activities in the Free-Living Study, and we evaluated the oversampling method SMOTE. We concluded that activity annotation type A (annotating 15 seconds for every performed activity based on the taken picture) performs best, while the SMOTE method does not contribute to an increase in the performance.

A limitation of this study is that the number of performed activities is not balanced. According to the scope of this experiment, the subjects did not receive any instructions on what types of activities to perform and for how long. As a result, most of the subjects spent the 8 hours of the experiment sitting at their desk. Compared to the Controlled Study, the activities lying with face up, face down, to the left and to the right were not performed, while the activity running was performed only from one subject for around 1 minute. Thus, the activity running was excluded from the dataset of this study. Regarding the distribution of the performed activities for the Free-Living Study, 38.39% of the data are for sitting, 42.5% for dynamic sitting, 3.03% for standing, 7.88% for dynamic standing, 0.19% for walking upstairs, 0.18% for walking downstairs, 5.89% for walking, and 1.92% for cycling. Based on the frequency of the performed activities which could bias the positive predicted activities, we can estimate that there will be an overfitting for some classes. A solution to this classification problem is to apply an oversampling method. However, we examined the SMOTE method and we found that this does not improve the performance. Consequently, the learning model is affected with overfitting for the activities dynamic sitting, dynamic standing, and walking. For instance, due to the small amount of labeled data for the activities walking upstairs and walking downstairs, these were misclassified with the activity walking.

Another limitation of this study concerns the activity annotation, which is of major importance in the models for uncontrolled experiments. Even though we annotated sufficiently the real activities based on pictures taken every 30 seconds (each activity was annotated for the last 15 seconds based on annotation type A), this could be a misleading factor for the accurate annotation of all the performed activities. For instance, there is a possibility that during 15 seconds, a subject might have performed more than one activity, such as sitting on a chair, standing up, walking a bit and again sitting on a chair. During this time only one activity can be annotated through the taken picture. This could have a great impact on annotating the activities that might be performed in a similar way, such as sitting and dynamic sitting, but also standing and dynamic standing. For this reason, we also investigate the activity recognition for classifying 5 activities, after merging the similar ones.

For validating the prediction on activities performed in uncontrolled settings, we used 22 subjects from the Free-Living Study for the test set. In particular, we used three validation models with different training sets. Validation model A trains the classifier based on data from the Free-Living Study, validation model B trains the model based on data from the Controlled Study, while validation model C trains the classifier using data from both studies. Based on the evaluation metrics F1-score and Cohen's kappa, which are commonly suggested for evaluating imbalanced data, the validation model A performs better in predicting both 8 and 5 activities of the Free-Living Study, followed by the model C. Regarding the model B, even though the overall performance does not outperform the other two models, it achieves a high score for the balanced accuracy metric.

For predicting 8 activities, the averaged performance of model A outweighs the other two models. For the model A, the F1-score for the SR sensor is 47.61%, for the CB is 50.98%, and for the SRCB is 53.72%. The performance is decreased due to data imbalanced, which lead to activities misclassification. Based on the SR confusion matrix, dynamic sitting receives 61% and is misclassified with sitting, standing and dynamic standing. Similarly, walking is misclassified with walking upstairs and downstairs for the SR, but also for the

CB and SRCB sensors. In contrast to SR, CB sensor performs better for almost all the activities except walking. On the other hand, SRCB performs slightly better than the CB. That means that the combination of placing accelerometers on the wrist and the chest can enhance the prediction for all the activities.

Additionally, the model A outweighs the models B and C for predicting 5 activities, after merging sitting with dynamic sitting, standing with dynamic standing and walking upstairs with walking downstairs. In particular, the F1-score for the SR is 82.73%, for CB is 85.29% and for SRCB is 86.98%. The activity sitting can be detected accurately through all the sensors with a score 97%. However, this class label suffers from overfitting and is misclassified with standing. Similarly, walking is scored with 74% (SR and SRCB) but is misclassified with walking on stairs. Finally, cycling is improved for the SRCB (74%), due to the combination of significant features from both SR and CB.

The SRCB-based classification model A performs sufficiently well for predicting the activities performed in the Free-Living Study. Even though some activities such as standing and walking on stairs cannot be predicted accurately, it is clear that the activity recognition in uncontrolled environments depends on training the classifier with data from the free-living environment. Similar to the results of the Controlled Study, the accuracy of our activity recognition model strictly depends on the purpose of the application. If the aim is to predict some specific activities, the prediction score varies. For instance, a classifier for the cycling activity recognition can be trained based on data from the Controlled Study. In particular, cycling receives a true positive score of 83% through the SRCB sensors for the validation model B. The reason is that cycling is the only activity performed in a similar way for the controlled and uncontrolled environments, and thus, data from the Controlled Study can be used to predict the specific activity in an uncontrolled environment. If the aim of the application is to recognize activities using a single sensor we would suggest the CB for data recording. Otherwise, the SRCB sensors can provide in general the most accurate activity recognition.

It is worth mentioning that the imbalanced data have significantly biased the prediction. The training model of this study can be used to detect the activity sitting (including dynamic sitting) accurately. Otherwise, more experiments have to be conducted in order to enhance the training set, by collecting data for more activities (such as running and lying) and for a more significant time, for predicting all the activities precisely.

6

Conclusions and Future Recommendations

The thesis aims to answer the research question: *"How accurately can we classify physical activity based on wearable accelerometers placed on the wrist and chest in a controlled and in a free-living environment?"*. Therefore, we conducted two experiments and we evaluated our model for activity recognition in a controlled and uncontrolled environment. We recruited forty subjects and we asked them to wear two imec wearables sensors; the SR (chest-worn) and the CB (wrist-worn) sensors. Both sensors record 3-axial acceleration. The performance of our model was evaluated using accelerometer data from the SR, the CB and the SRCB (combination of SR and CB) sensors.

Concerning the Controlled Study, we asked the subjects to perform 13 consecutive activities, including the everyday activities: sitting, dynamic sitting, lying with face up and down, lying to the left and right, standing, dynamic standing, walking upstairs and downstairs, walking, running and cycling. For the Free-Living Study, 37 subjects participated and performed the activities in an uncontrolled environment, without receiving any supervision for what activities to perform. In total 8 activities were performed, including sitting, dynamic sitting, standing, dynamic standing, walking upstairs and downstairs, walking and cycling, during working hours.

After the data acquisition phase, we performed the data processing in order to extract valuable features from raw accelerometer. In particular, we used a window segment with 5 seconds window length and 4 seconds overlap and we extracted 52 features based on frequency and time domain. Additionally, we used the Pearson's correlation coefficient to remove highly correlated features, and the LinearSVC model to remove features with low information gain. It is worth mentioning that each feature has a different weight in the classification and varies per sensor, but also per study. In the Controlled Study, for instance, the calculated feature for the mean values of y-axis contribute significantly in the performance of the CB, while the mean distance between the y- and z- axes is the feature that contributes the most for the SR and SRCB sensors. On the other hand, in the Free-Living Study, the magnitude among the three axes is the most significant feature for the SR sensor. Overall, the most significant features are the mean, mean distance, magnitude, absolute mean, standard deviation, max, min, and entropy. After the data processing, we performed the analysis using the Random Forest classifier. For the validation of our model, we used the Leave-One-Out-Subject Cross Validation and we tested the prediction on activities before and after merging the similar ones.

In order to answer the research question sufficiently, we have examined the following four sub-questions. For the first one *"Can 3-axial accelerometer data, from a single wrist-worn or chest-worn sensor, be used to detect simple everyday activities?"*, we proved that a wrist-worn sensor can detect most of the everyday activities accurately. For the Controlled Study, the activities sitting, standing, walking on stairs, running, and cycling were detected accurately with 84.08% F1-score. Similarly, the activities sitting, standing and cycling were detected accurately with 85.29% F1-score in the Free-Living Study. However, a sufficient answer to this question also depends on the purpose of the activity recognition and what activities have to be detected. The activities walking upstairs and walking downstairs were not always predicted accurately and were misclassified with the activity walking. A solution to this problem could be either the use of an additional sensor placed on a different location on the body (e.g., thigh), or the recording of additional types of data (e.g., gyroscope). On the other hand, a chest-worn sensor is recommended if the purpose of the application is to detect the lying activities.

Regarding the question *"How does combining accelerometer sensors, placed on the wrist and chest, affect the accuracy of activity recognition? Does the classification performance improve for predicting certain types of activity?"*, we proved that accelerometer data recorded by two devices, one placed on the wrist and one on the chest, can significantly enhance the activity recognition system for both studies. In particular, the SRCB F1-score for predicting seven main activities in the Controlled Study is 91.83%, which outperforms the SR (80.62%) and CB (84.08%). It is worth mentioning that the use of the SRCB sensor can significantly improve the performance of the prediction model for the activity sitting. Additionally, the prediction for the activities standing and cycling has also been improved in the Free-Living Study through SRCB sensors.

An appropriate answer to the question *"How well does a sensor contribute to detecting both static and dynamic activities, such as static sitting and dynamic sitting, by recording accelerometer data from a single sensor?"* depends on the task of monitoring. Regarding the Controlled Study, we showed that a wrist-worn sensor could accurately detect the activities sitting and dynamic sitting, and especially, standing and dynamic standing. However, in the Free-Living Study the prediction model is prone to misclassifying the activities performed on a similar way, such as standing and dynamic standing, and thus, the SRCB is recommended for enhancing the performance. It is worth mentioning that by merging the similar activities, the classification performance can be significantly improved for both SR, CB, and SRCB sensors. Consequently, accelerometer data from a single sensor are not sufficient enough to detect these types of activity in a real-life environment.

Another important outcome of the current work is our approach to investigate any possible correlation in the activities performed in both controlled and uncontrolled environments. A clear answer to the question *"Is it possible to detect activities performed in uncontrolled settings, through a classification model that was trained with data from a controlled environment?"* is that most of the investigated activities are performed on a different way and cannot always be detected. However, the activity cycling is an exception. Based on the validation model B, cycling was accurately detected through SRCB using a classifier that was trained with data from the Controlled Study. Thus, a valid answer for the above question depends also on the task of the monitoring and what activities should be detected. However, an extra study is recommended to investigate this question. A possible option could be to apply a filter on the raw accelerometer data from the Free-Living Study, in order to remove any possible noise, and then use data from the Controlled Study to train the classifier.

A limitation of the current work could be the activities performed in the Free-Living Study. Due to imbalanced data, it is clear that a further experiment has to be conducted in order to collect data for more activities, recording data for more than 8 hours or even for multiple days. For instance, the activities performed during the office hours can significantly differentiate from the activities performed during a weekend. Another limitation is the activity annotation for the Free-Living Study. Even though the labeled activities were annotated sufficiently based on the taken pictures, we decided to label each activity for 15 seconds (according to the taken picture). Nevertheless, this approach might not be always accurate. A solution to this problem could be the use of video recording for annotating all the performed activities, continuously. However, this will increase the complexity of interpreting the activities.

Data from the Controlled Study were used in order to select the attributes for the data processing. However, we concluded that data from the Free-Living Study differentiate significantly, and thus, further research on optimizing the data processing techniques is suggested. Hence, future work and optimization on the Free-Living Study could potentially enhance the activity recognition even further. For instance, a window size, more prominent than 5 seconds, could be preferred for extracting features in uncontrolled environments due to the complexity of the performed activities. Additionally, investigating another classifier might enhance the prediction model in uncontrolled settings.

Overall, wearable accelerometers, placed on the wrist and chest, can be significant accurate for predicting everyday activities. For most of the performed activities, a combination of the sensors mentioned above is recommended in order to enhance the classification performance. It is worth mentioning that the wrist-worn sensor, as a standalone device, should be highly recommended for any system in activity recognition. However, this also depends on the task of monitoring and the activities that must be detected. For instance, the chest-worn sensor should be considered for detecting the activities lying (including lying with face up-/down and lying to the left/right). Furthermore, the classification performance for predicting the activities walking and walking on stairs might be increased by placing an additional sensor on the thigh or by combining accelerometer with gyroscope data. Finally, we conclude that future research should extensively focus on activity recognition in uncontrolled environments for detecting real-life activities, performed in the wild.

A

Controlled Study: Learning Model

A.1. FEATURES DISTRIBUTION

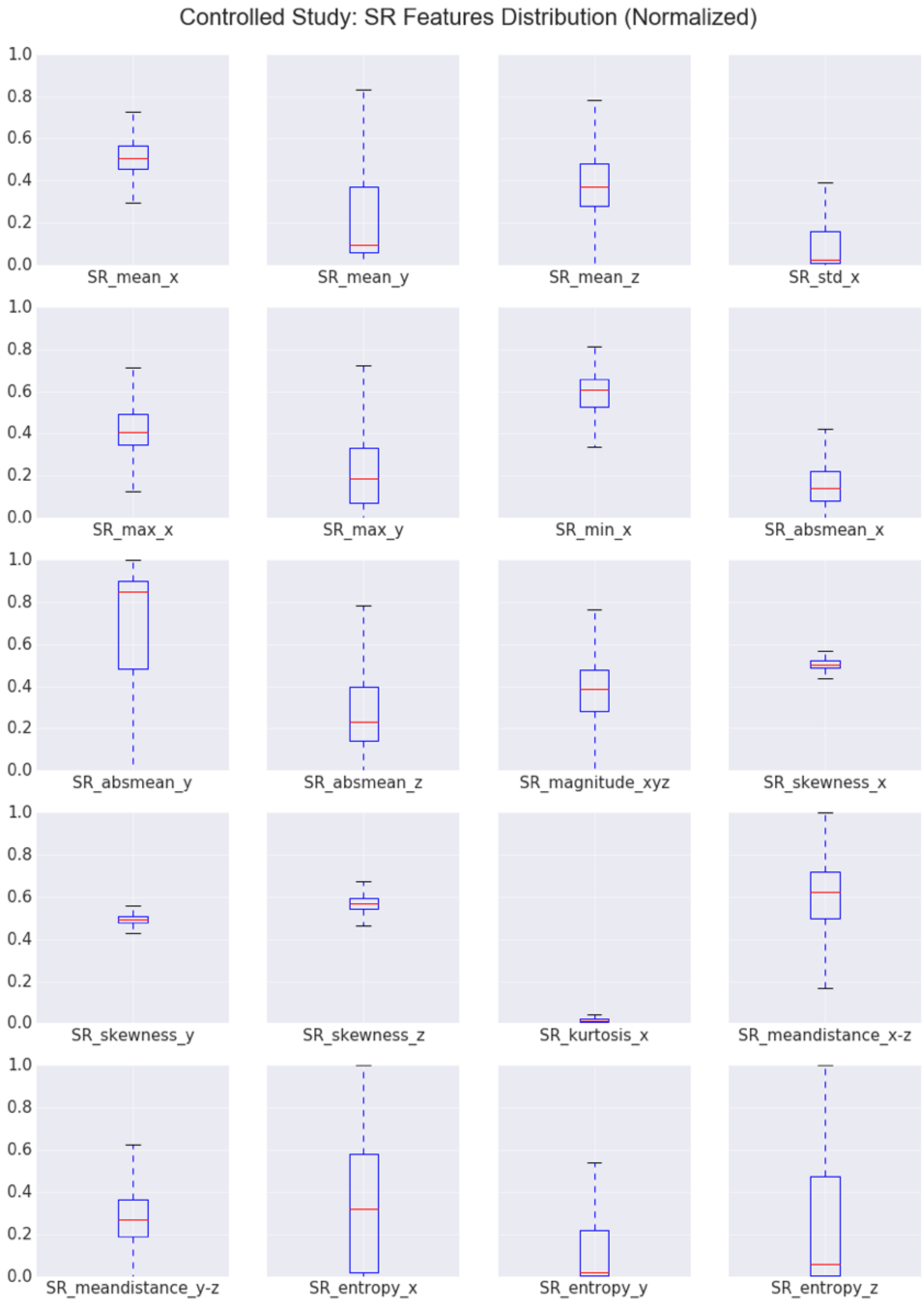


Figure A.1: SR features distribution for the learning model (Controlled Study).

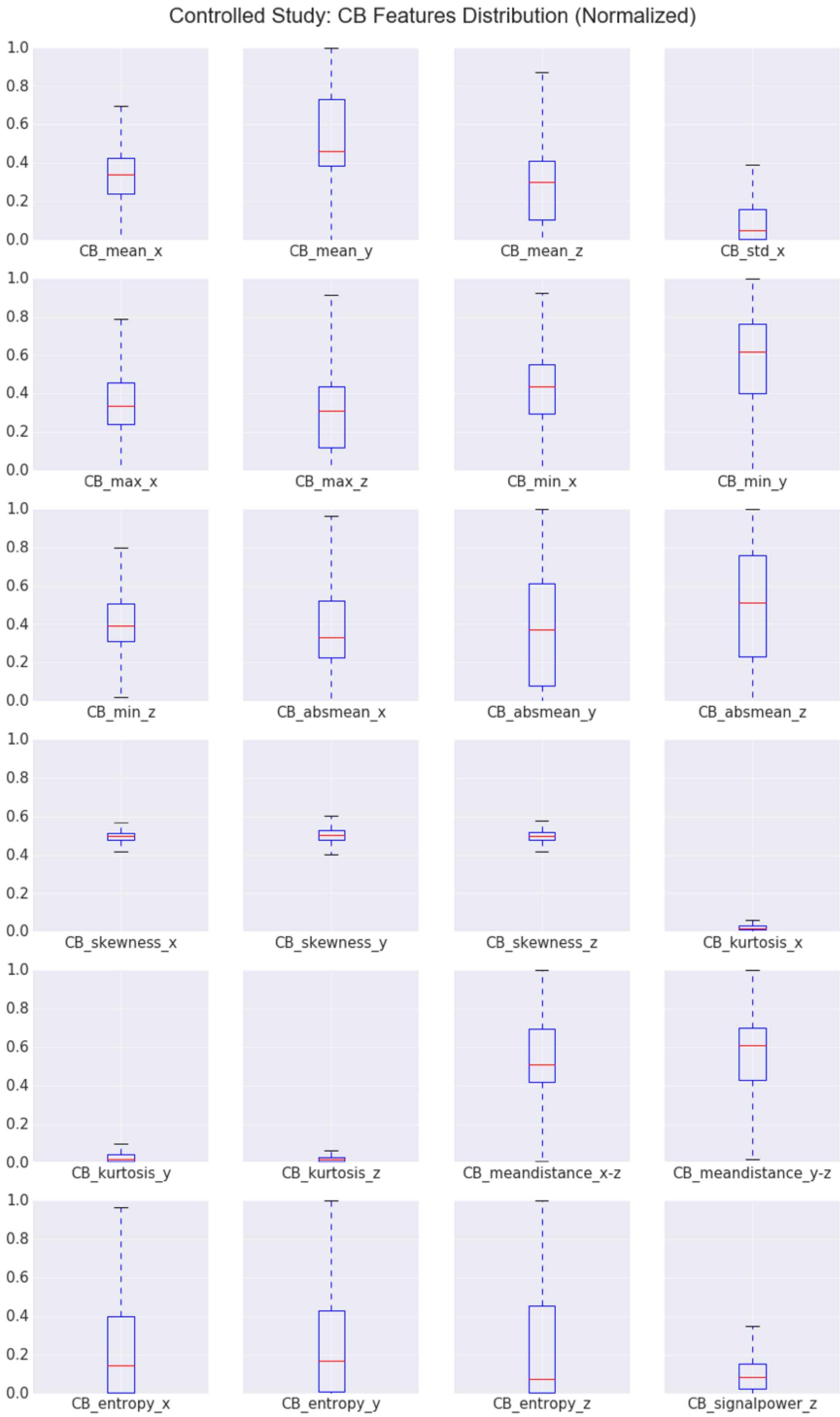


Figure A.2: CB features distribution for the learning model (Controlled Study).

A.2. CONFUSION MATRIX

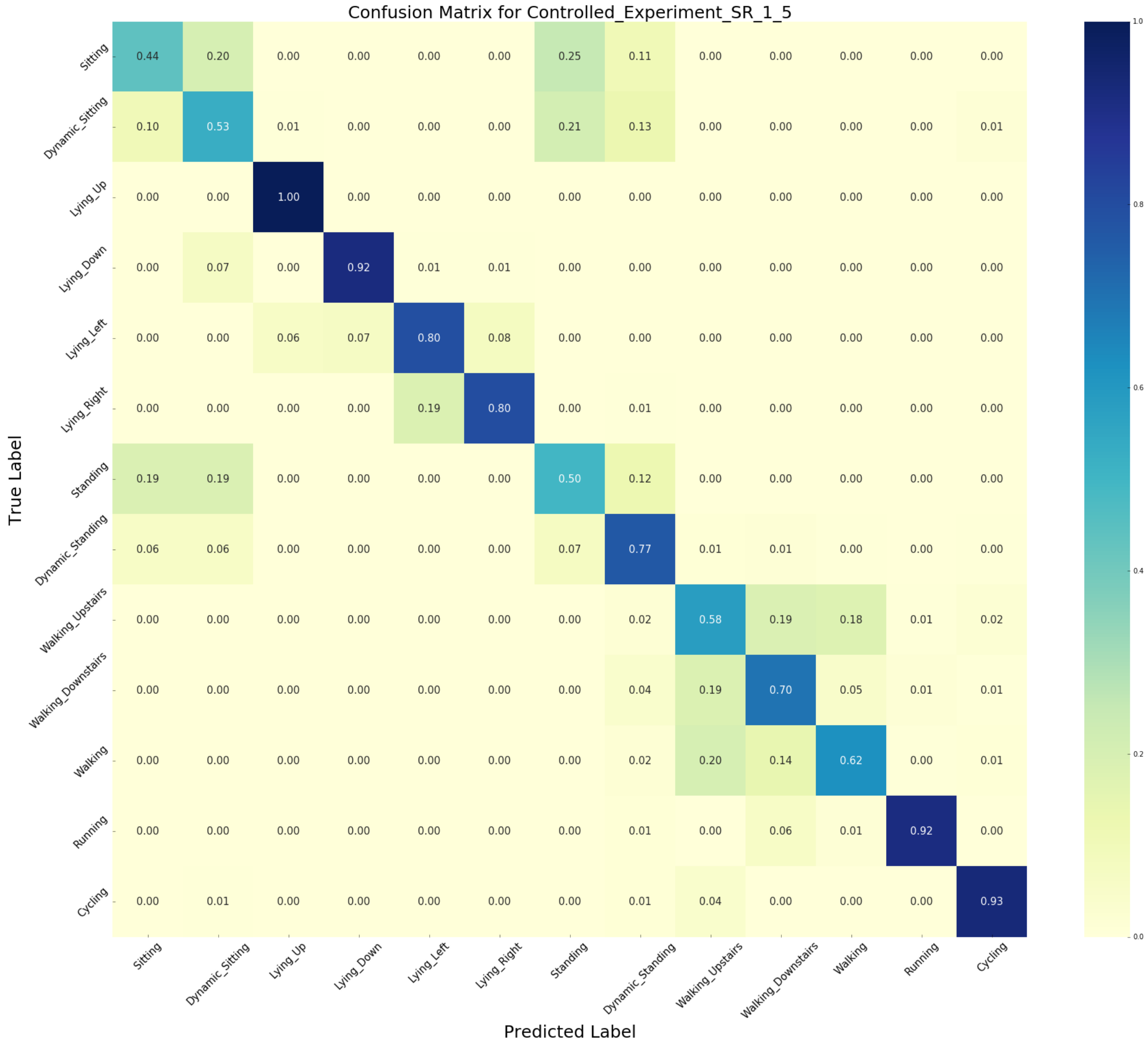


Figure A.3: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-15 for Controlled Study, by detecting 13 activities.

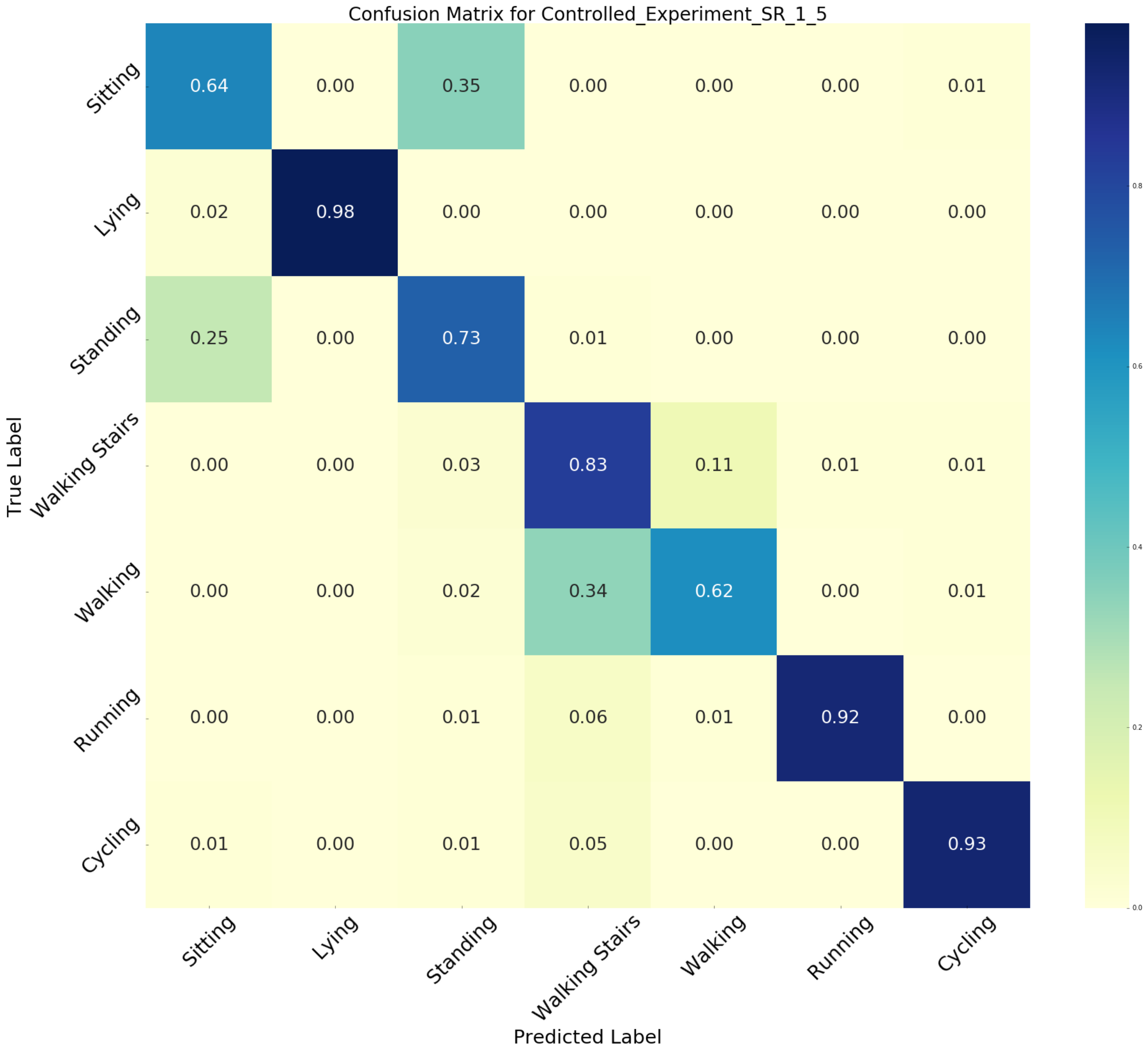


Figure A.4: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-15 for Controlled Study, by detecting 7 activities.

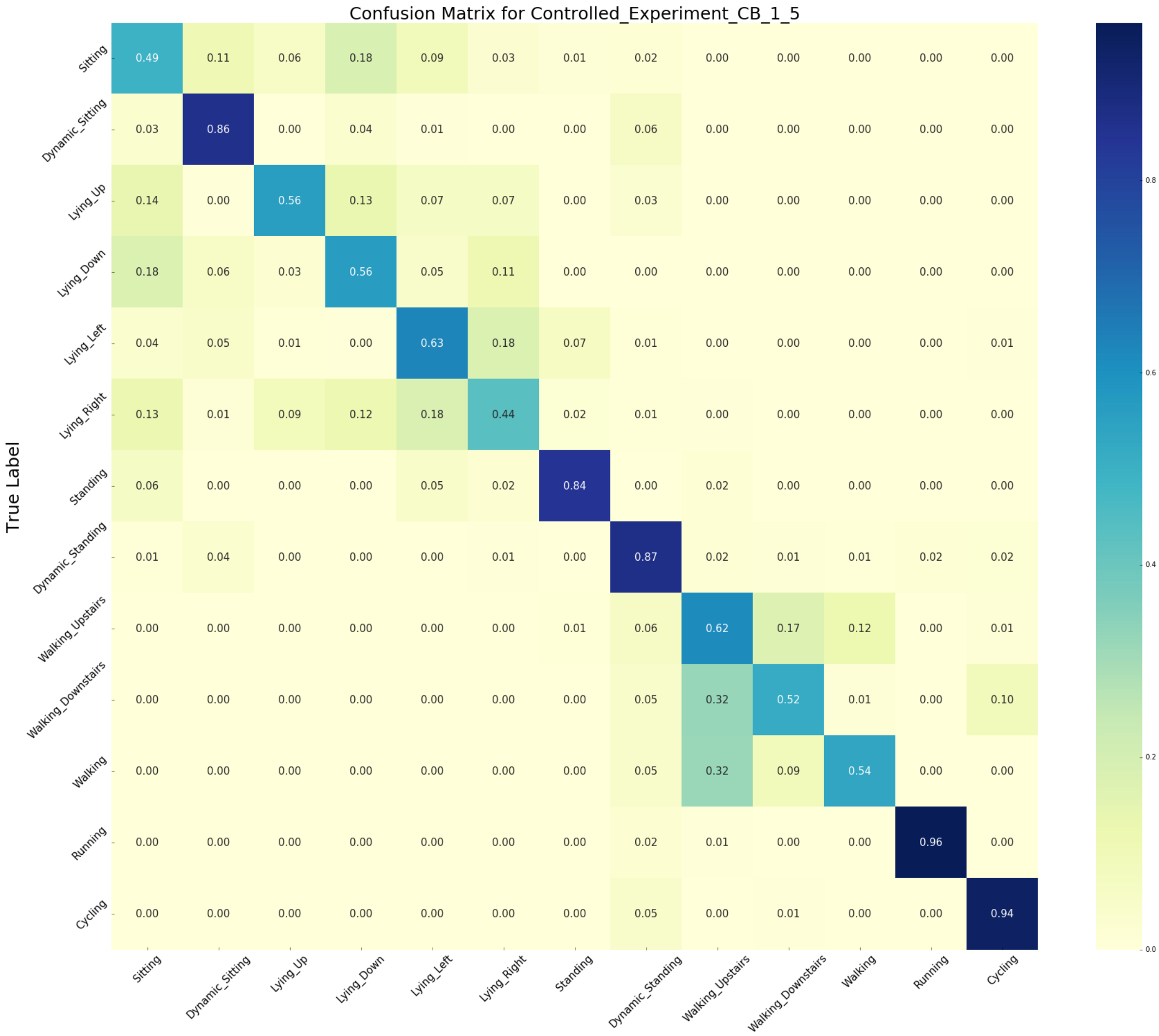


Figure A.5: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-15 for Controlled Study, by detecting 13 activities.

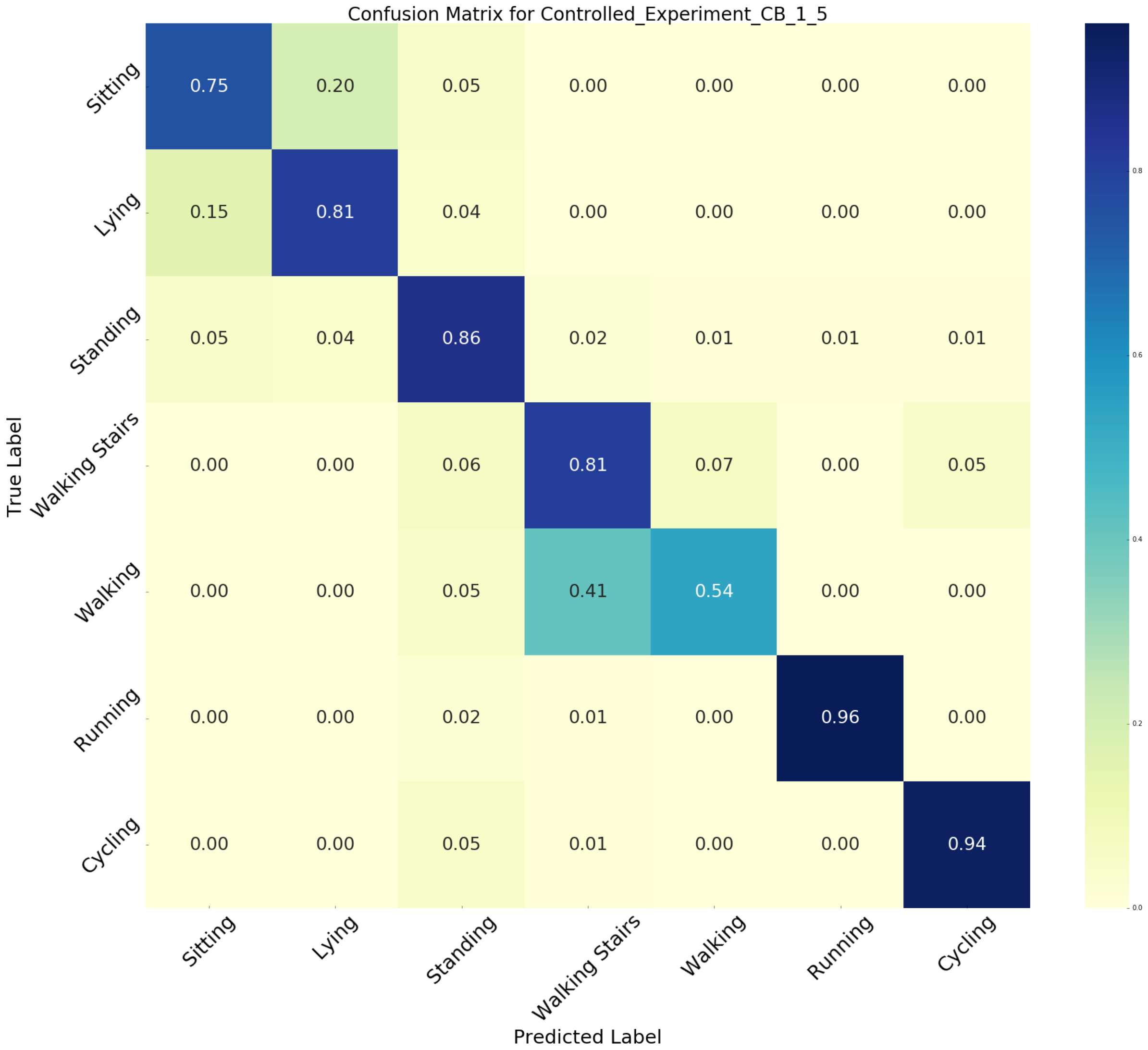


Figure A.6: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-15 for Controlled Study, by detecting 7 activities.

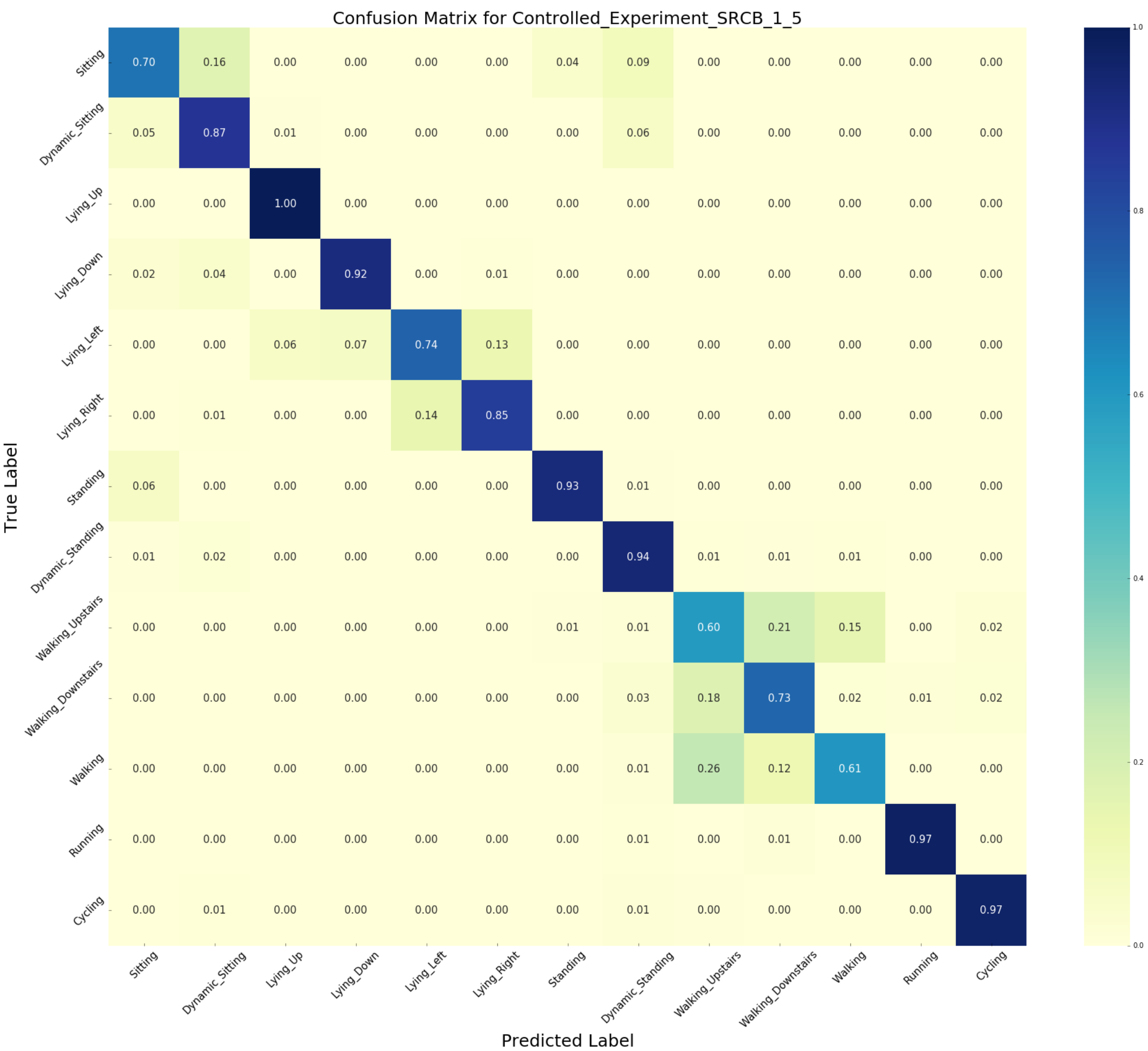


Figure A.7: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-15 for Controlled Study, by detecting 13 activities.

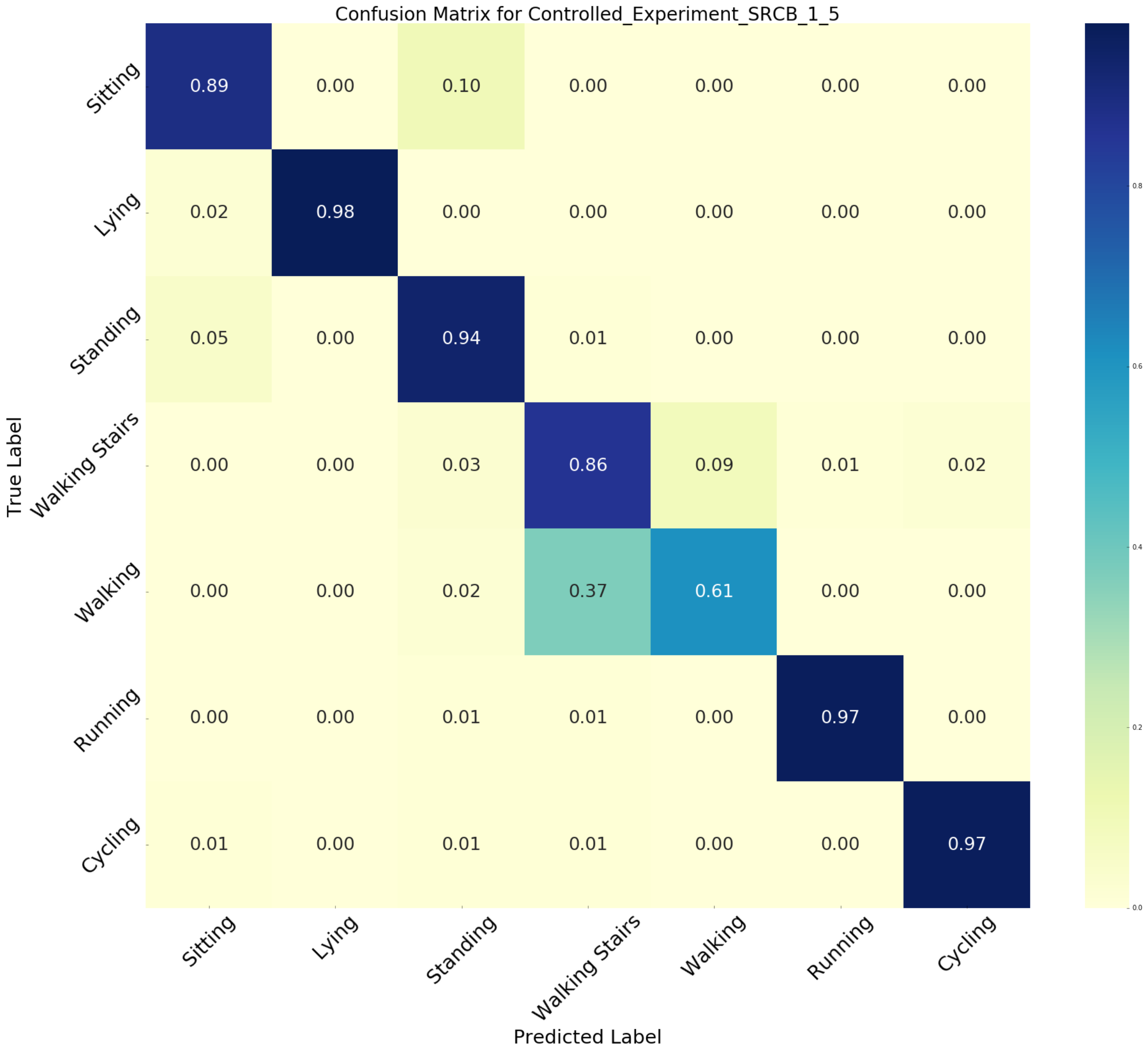


Figure A.8: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-15 for Controlled Study, by detecting 7 activities.

B

Controlled Study: Validation

B.1. Actual Vs Predicted: 13 Activities

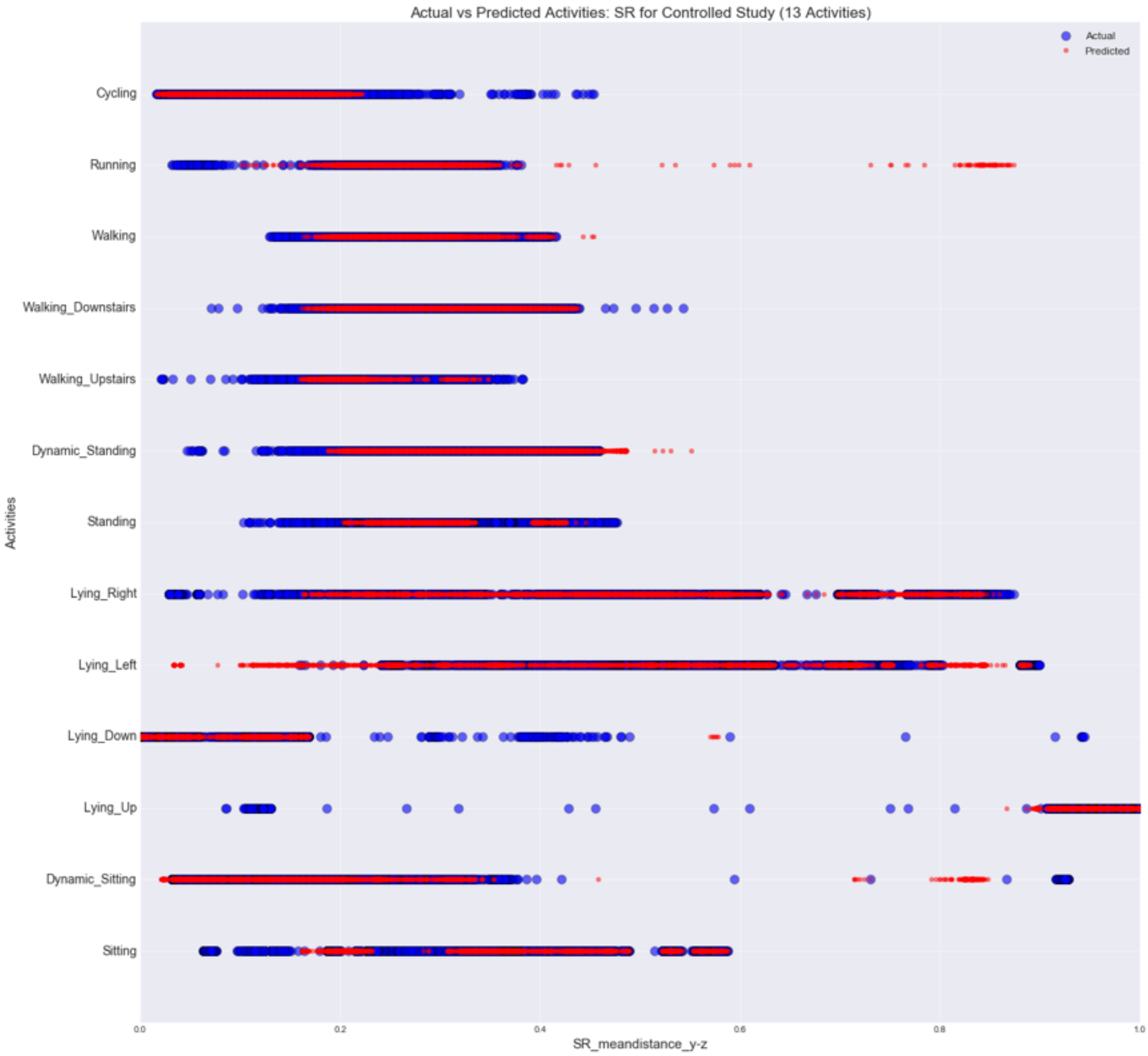


Figure B.1: Actual versus predicted activities for the test set in the Controlled Study. The prediction is based on SR sensor for detecting 13 activities.

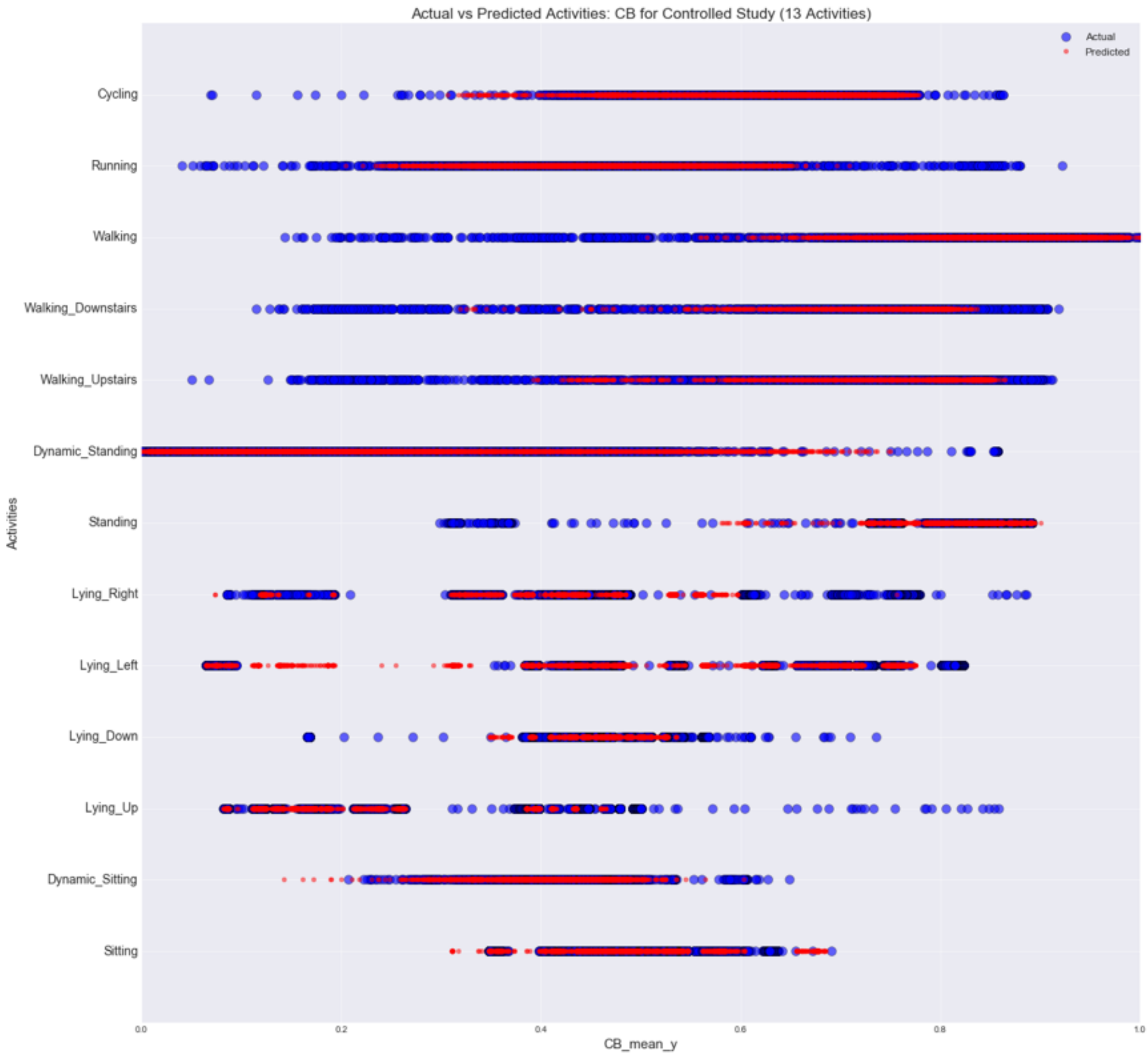


Figure B.2: Actual versus predicted activities for the test set in the Controlled Study. The prediction is based on CB sensor for detecting 13 activities.

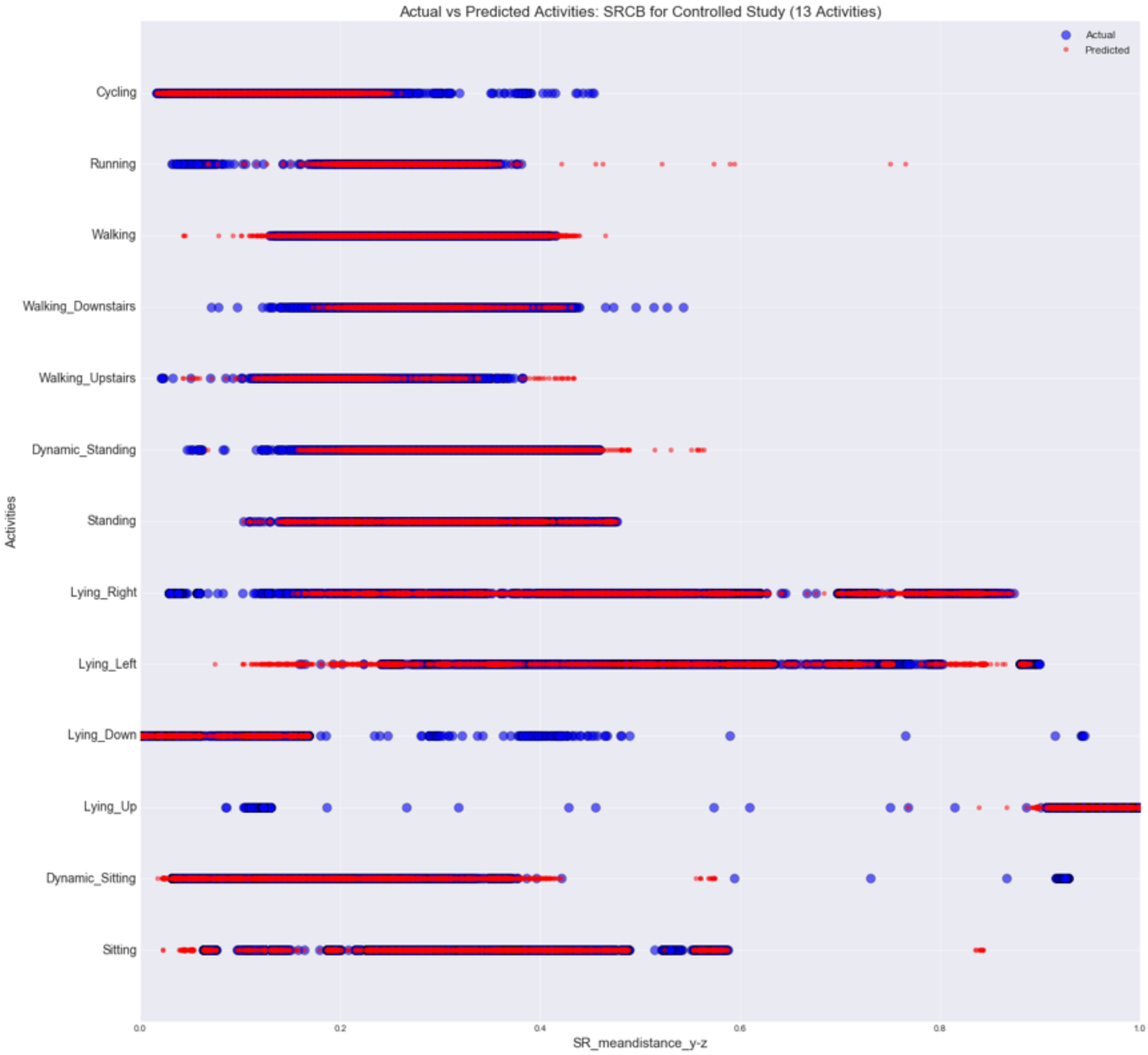


Figure B.3: Actual versus predicted activities for the test set in the Controlled Study. The prediction is based on SRCB sensors for detecting 13 activities.

B.2. ACTUAL VS PREDICTED: 7 ACTIVITIES

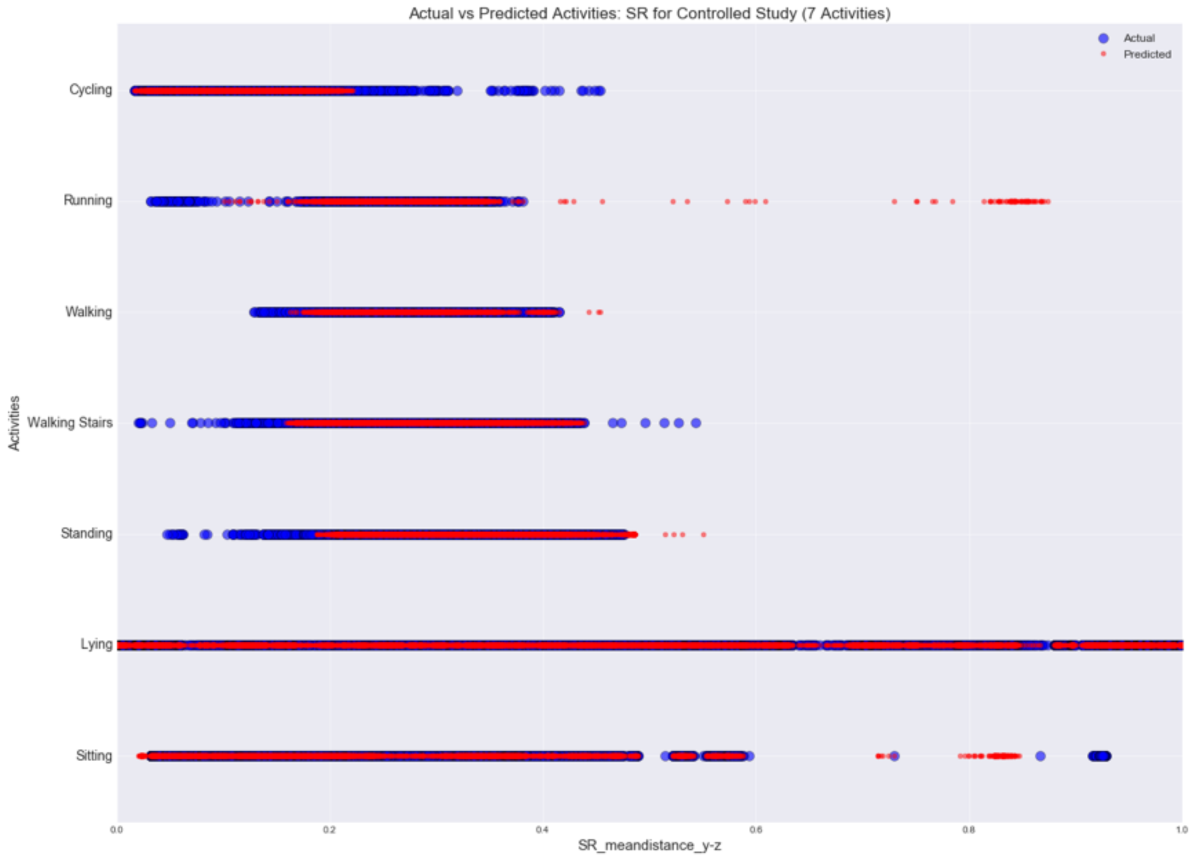


Figure B.4: Actual versus predicted activities for the test set in the Controlled Study. The prediction is based on SR sensor for detecting 7 activities.

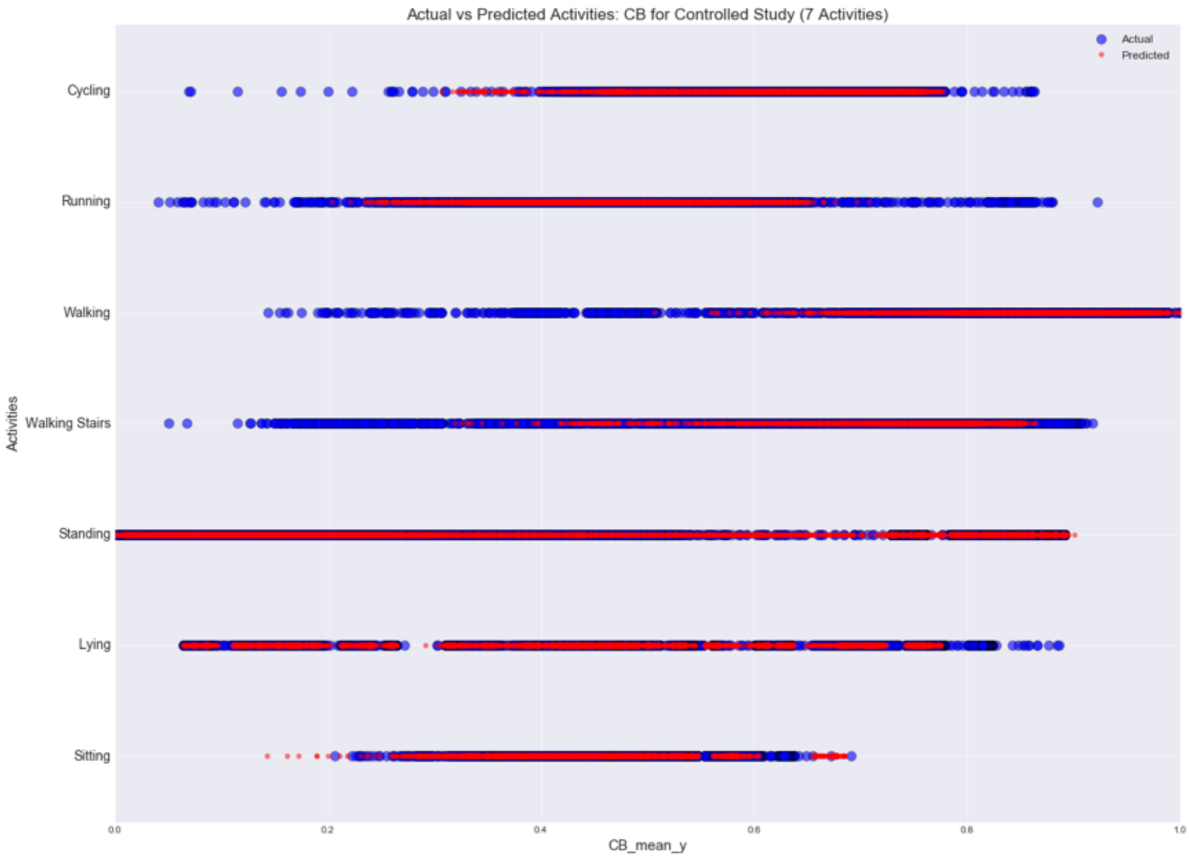


Figure B.5: Actual versus predicted activities for the test set in the Controlled Study. The prediction is based on CB sensor for detecting 7 activities.

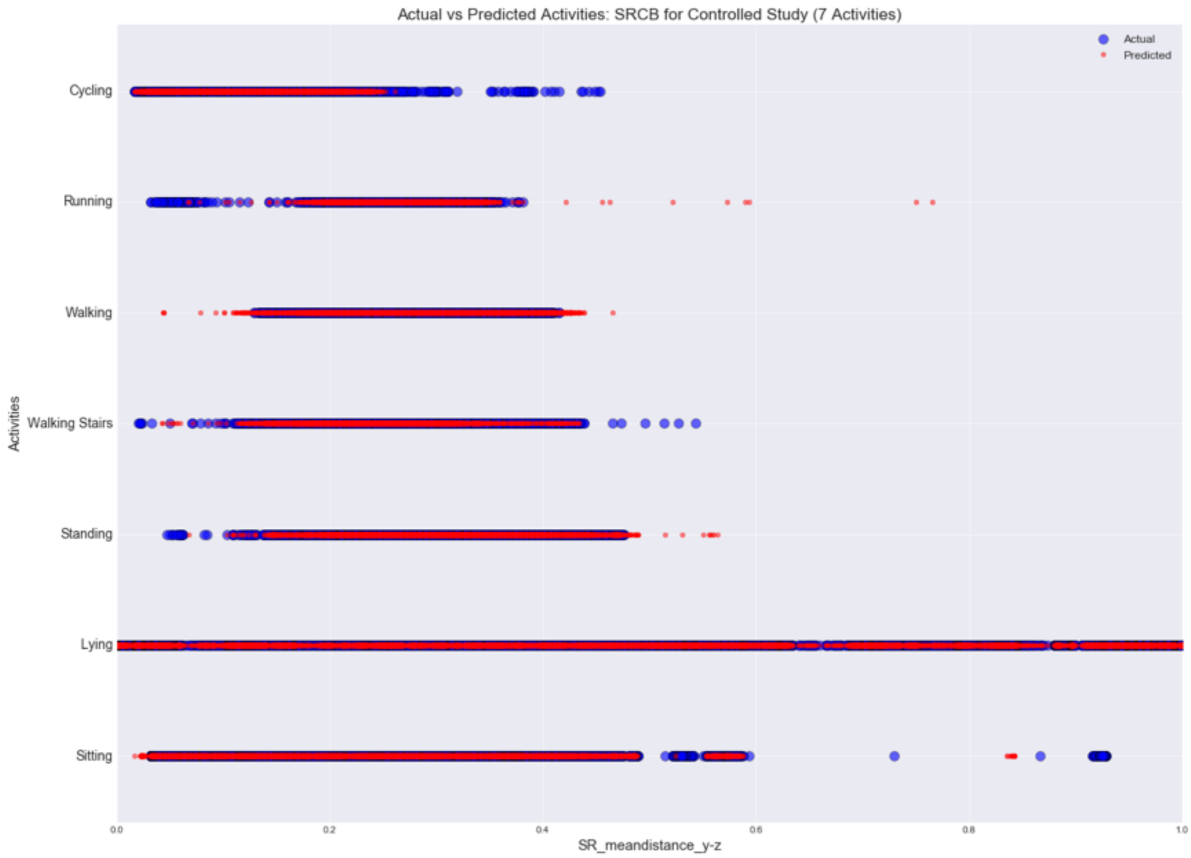


Figure B.6: Actual versus predicted activities for the test set in the Controlled Study. The prediction is based on SRCB sensors for detecting 7 activities.

C

Free-Living Study: Learning Model

C.1. FEATURES DISTRIBUTION

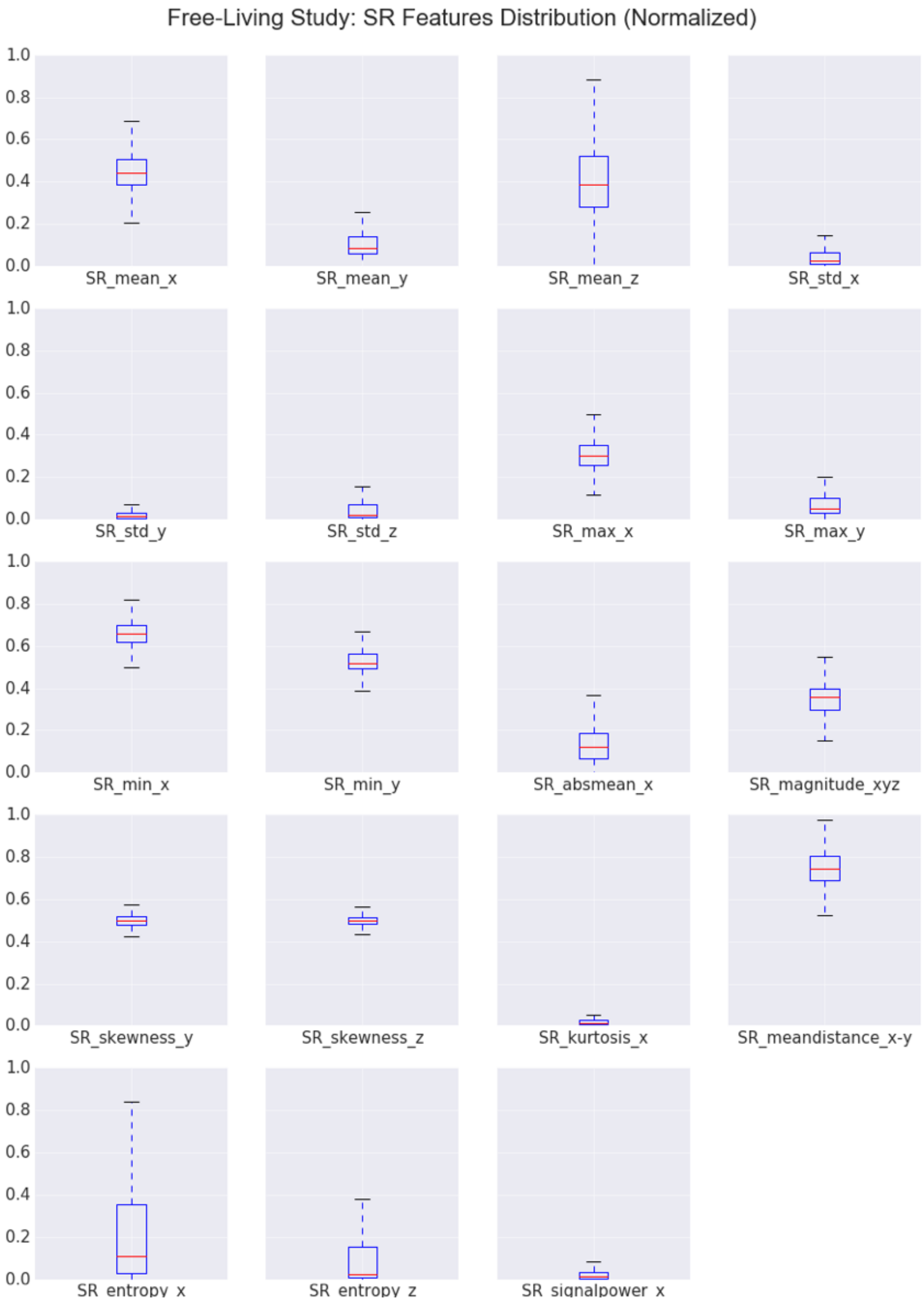


Figure C.1: SR features distribution for the learning model (Free-Living Study).

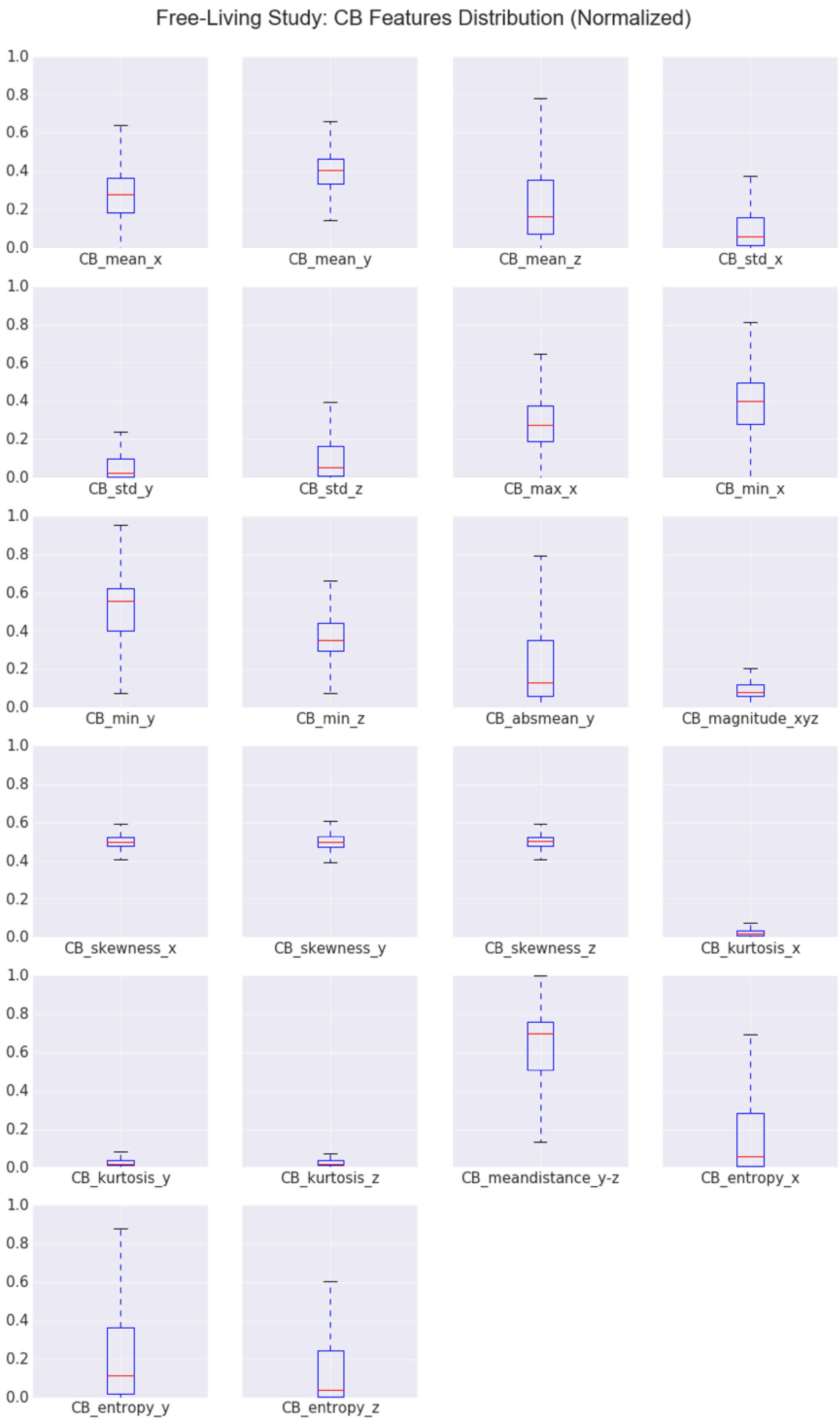


Figure C.2: CB features distribution for the learning model (Free-Living Study).

C.2. CONFUSION MATRIX

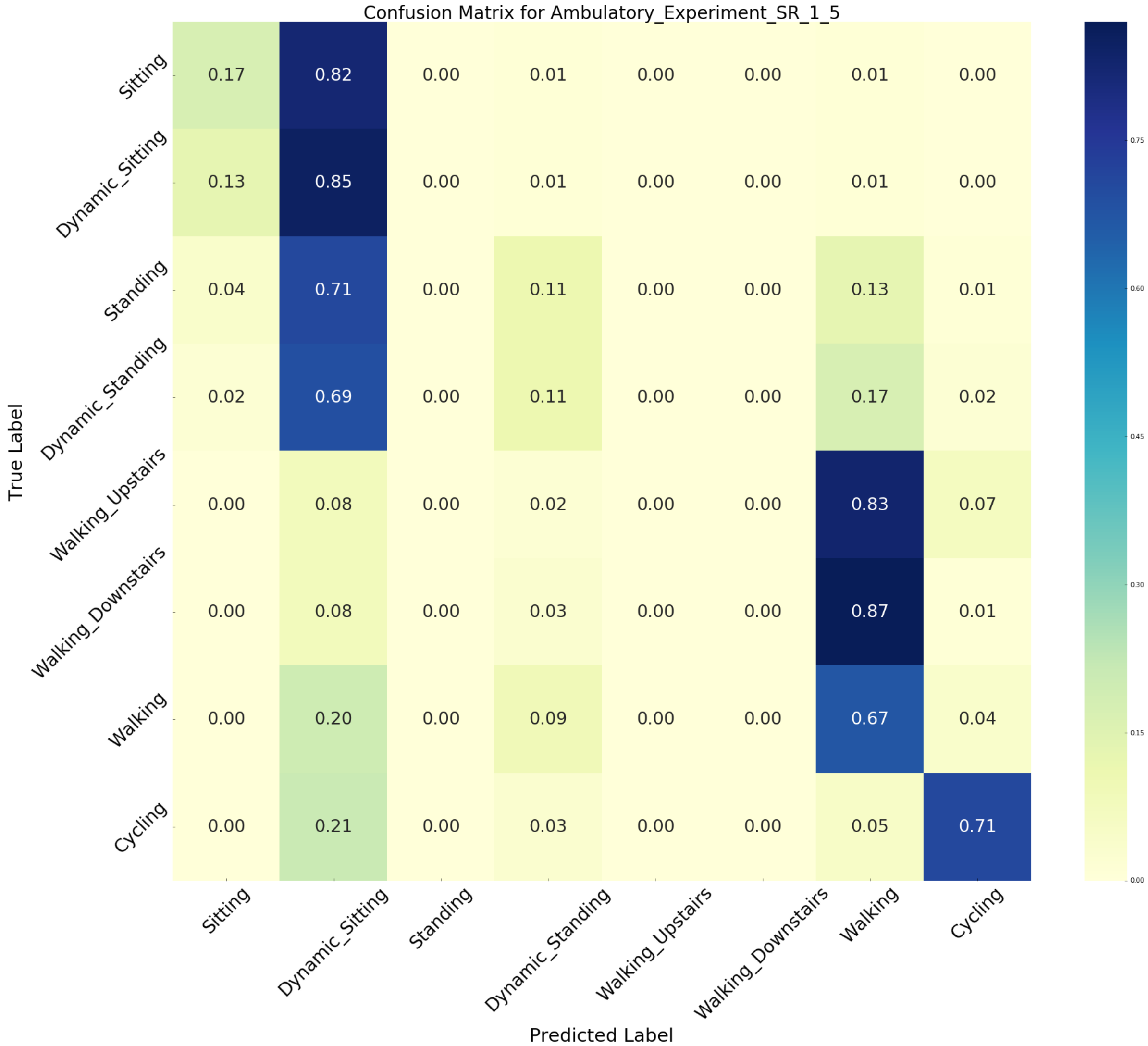


Figure C.3: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-15 for Free-Living Study, by detecting 8 activities.

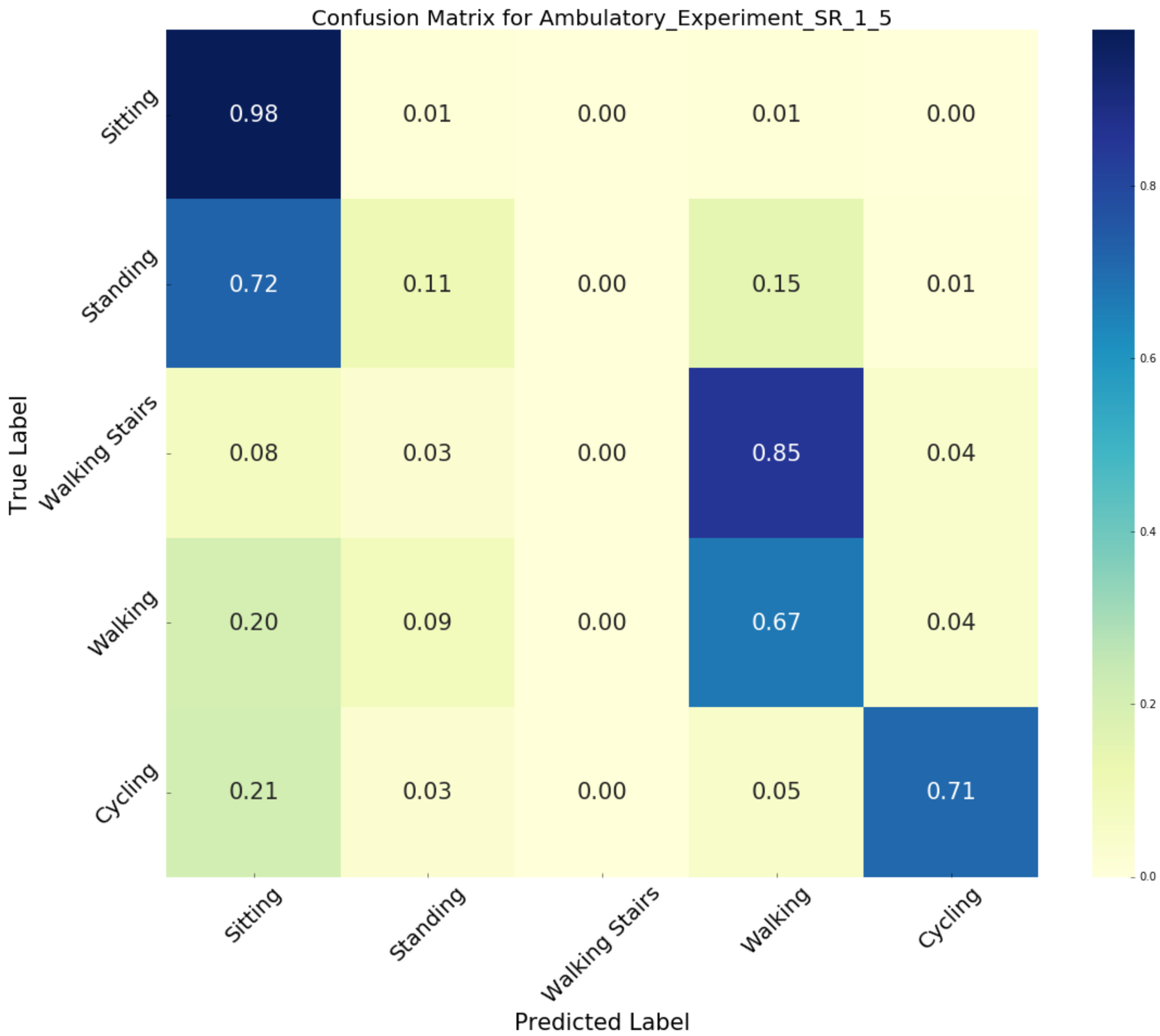


Figure C.4: The confusion matrix for the stingray (SR) sensor is presented. The validation is based on LOSOCV-15 for Free-Living Study, by detecting 5 activities.

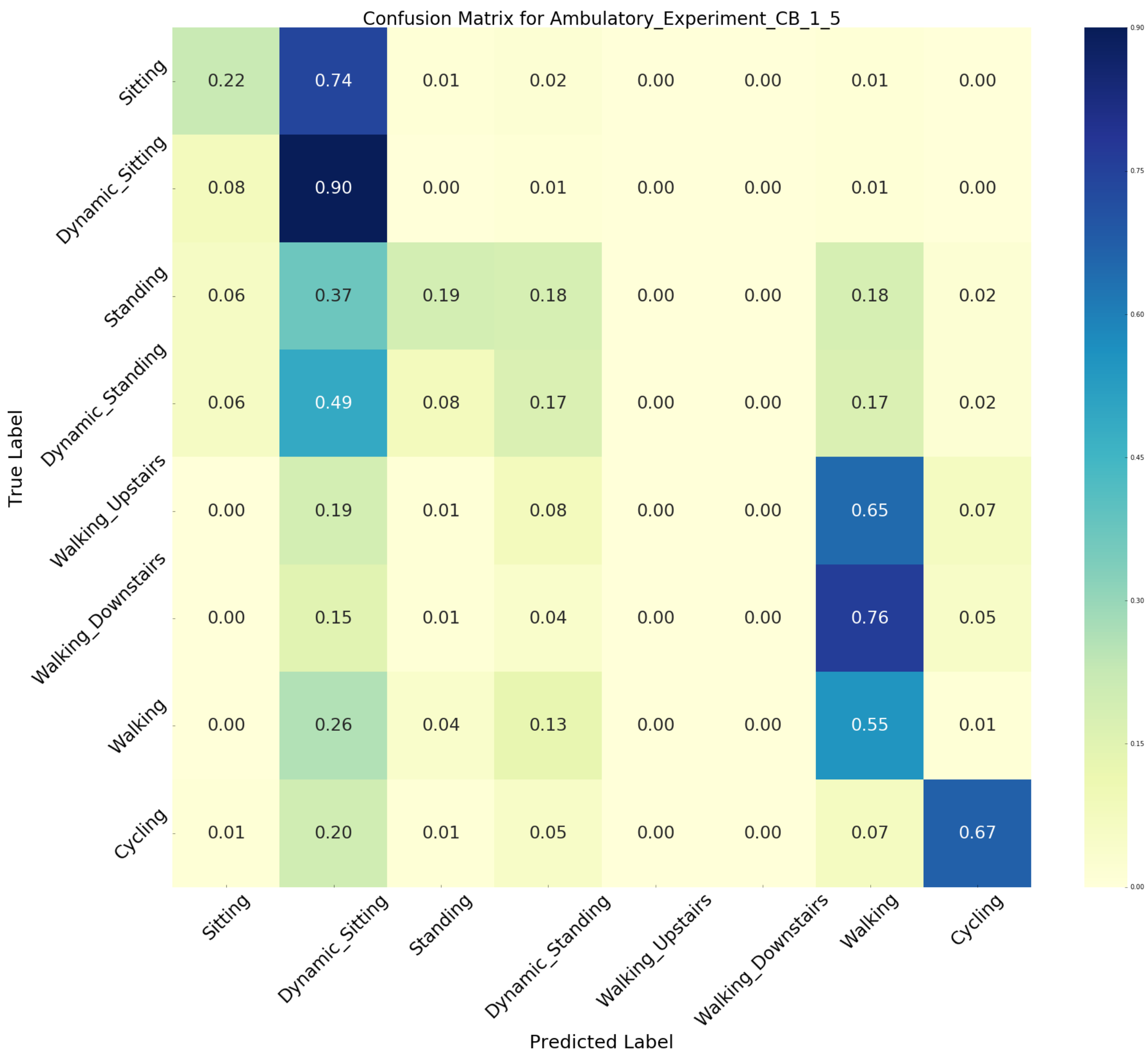


Figure C.5: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-15 for Free-Living Study, by detecting 8 activities.

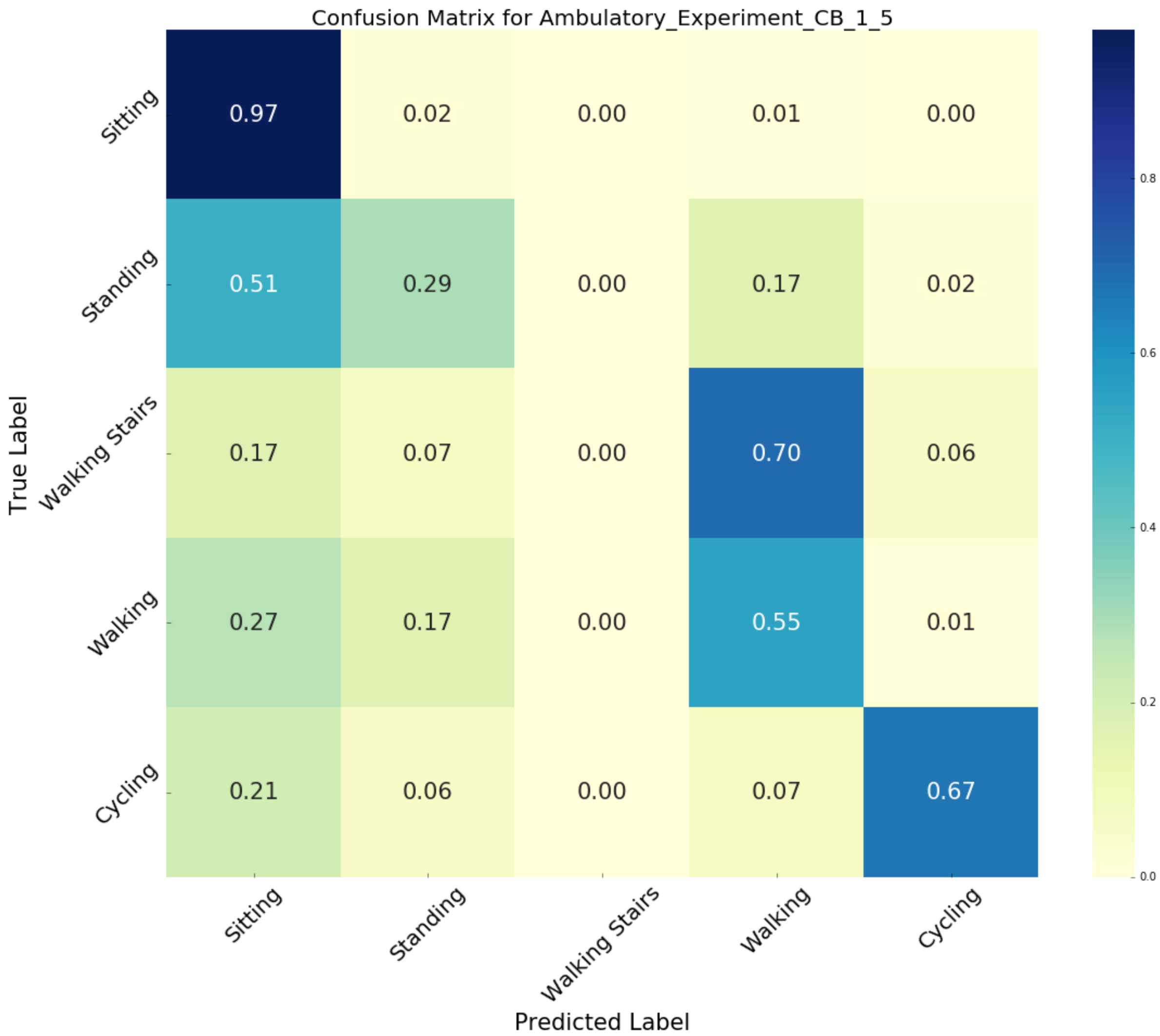


Figure C.6: The confusion matrix for the chillband (CB) sensor is presented. The validation is based on LOSOCV-15 for Free-Living Study, by detecting 5 activities.

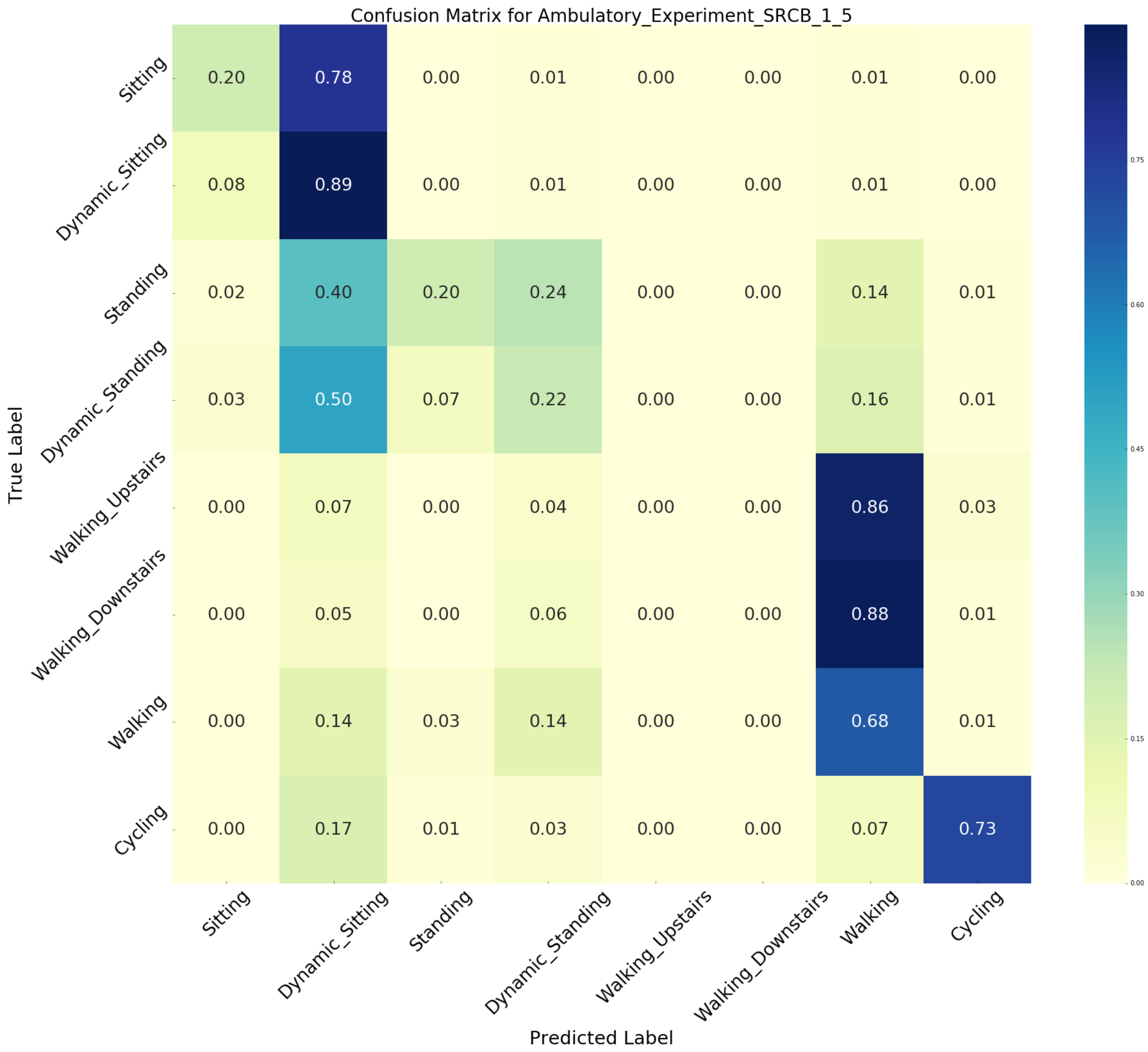


Figure C.7: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-15 for Free-Living Study, by detecting 8 activities.

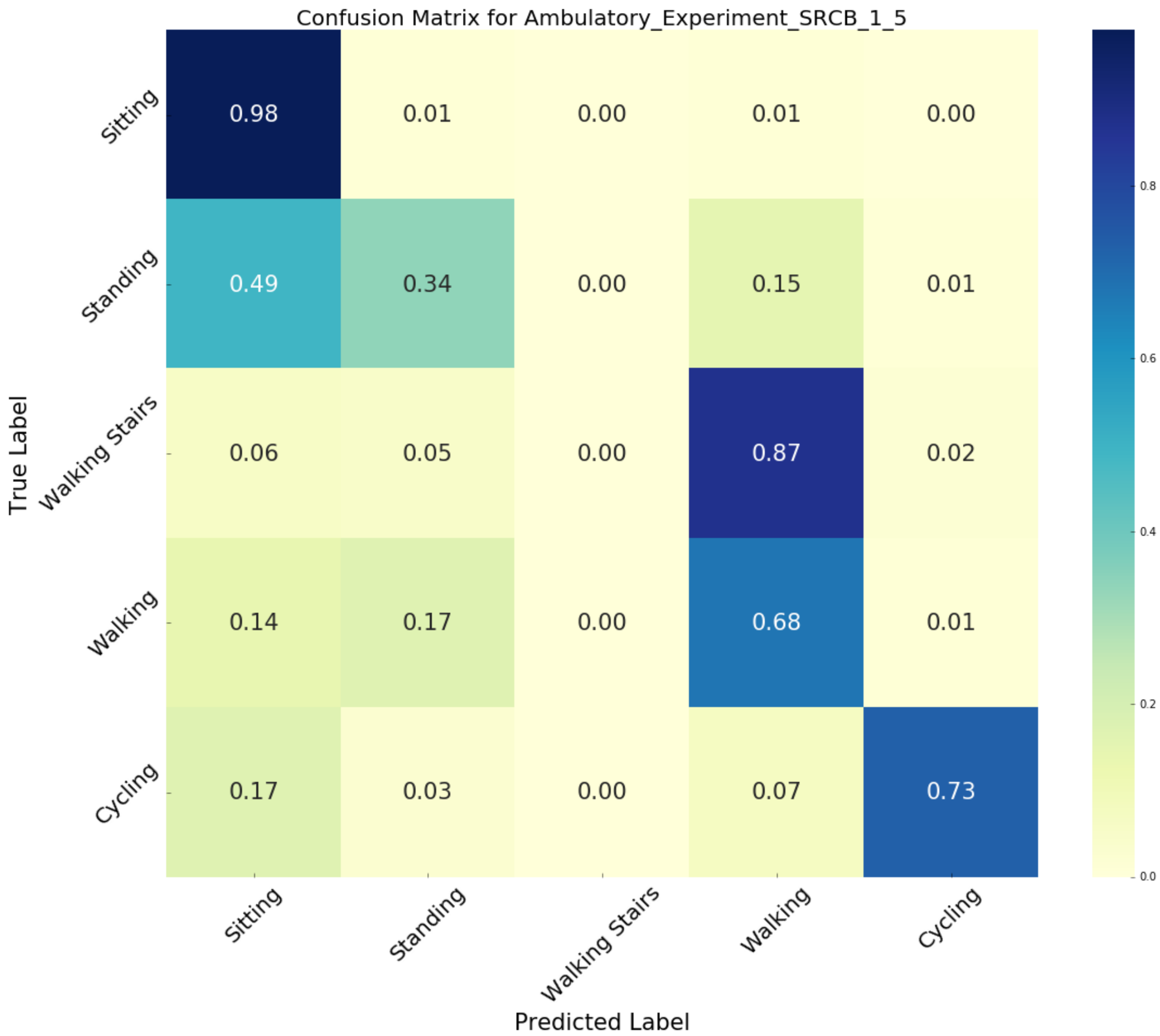


Figure C.8: The confusion matrix for the stingray-chillband (SRCB) sensor is presented. The validation is based on LOSOCV-15 for Free-Living Study, by detecting 5 activities.

D

Free-Living Study: Validation Model A

D.1. ACTUAL VS PREDICTED: 8 ACTIVITIES

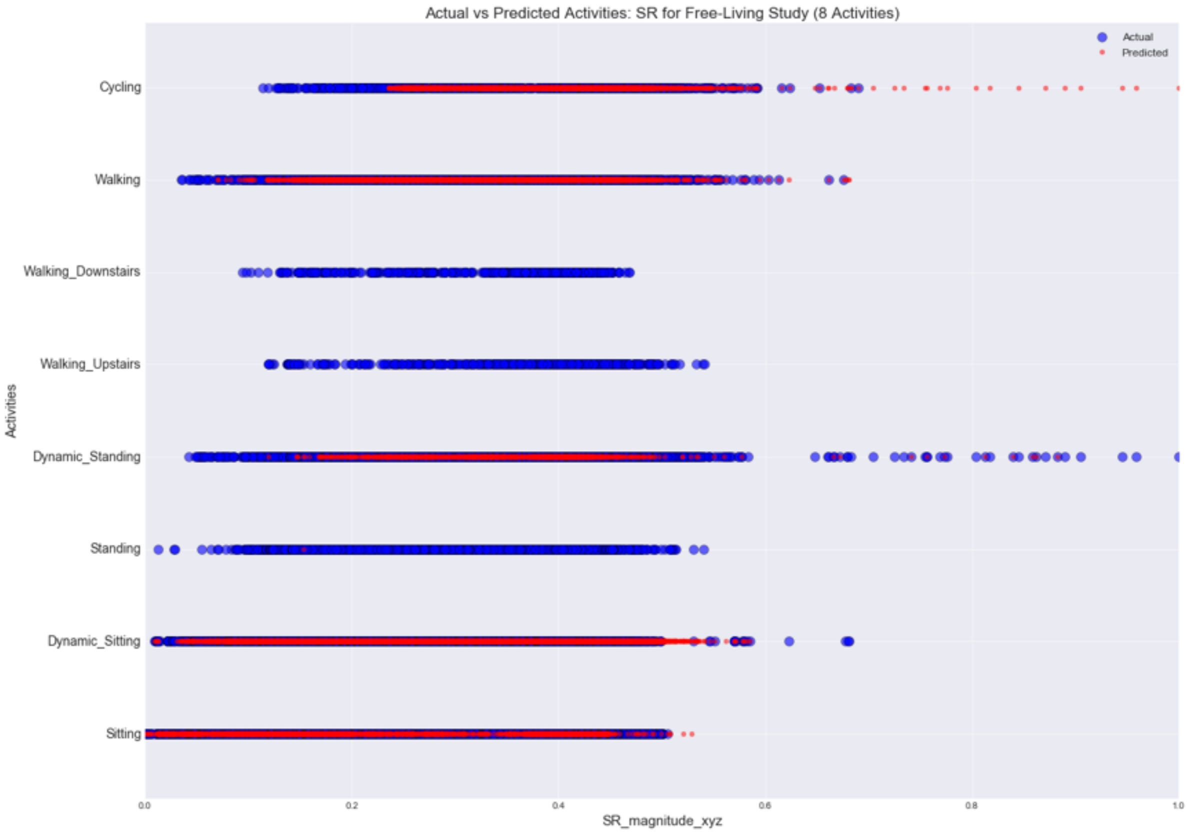


Figure D.1: Actual versus predicted activities for the test set in the Free-Living Study. The prediction is based on SR sensor for detecting 8 activities.

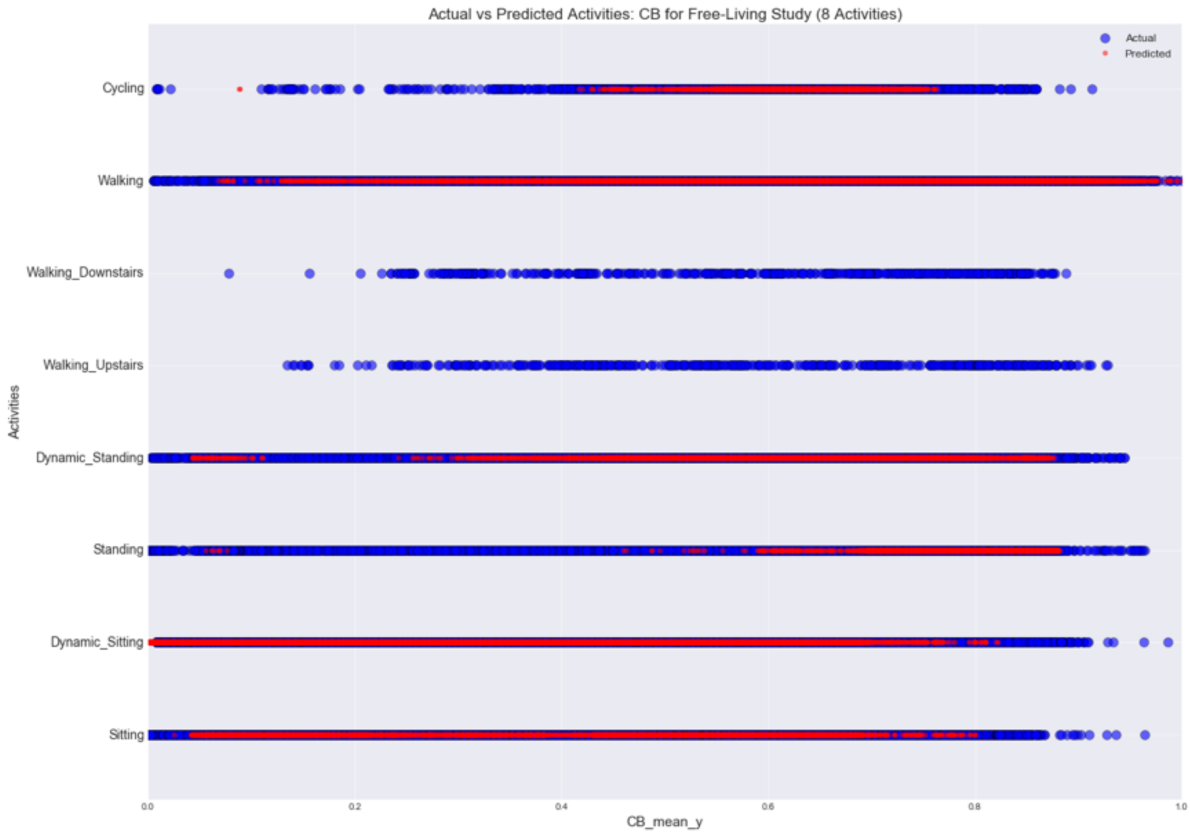


Figure D.2: Actual versus predicted activities for the test set in the Free-Living Study. The prediction is based on CB sensor for detecting 8 activities.

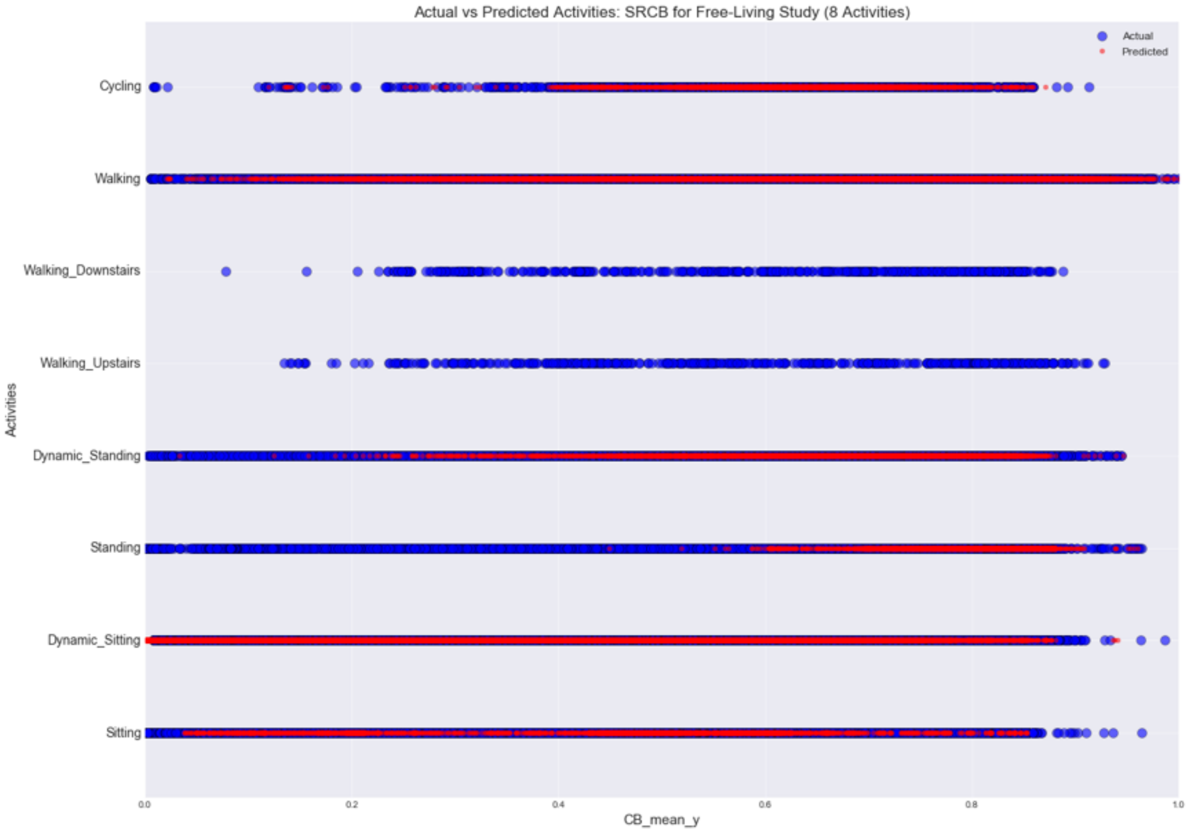


Figure D.3: Actual versus predicted activities for the test set in the Free-Living Study. The prediction is based on SRCB sensors for detecting 8 activities.

D.2. ACTUAL VS PREDICTED: 5 ACTIVITIES

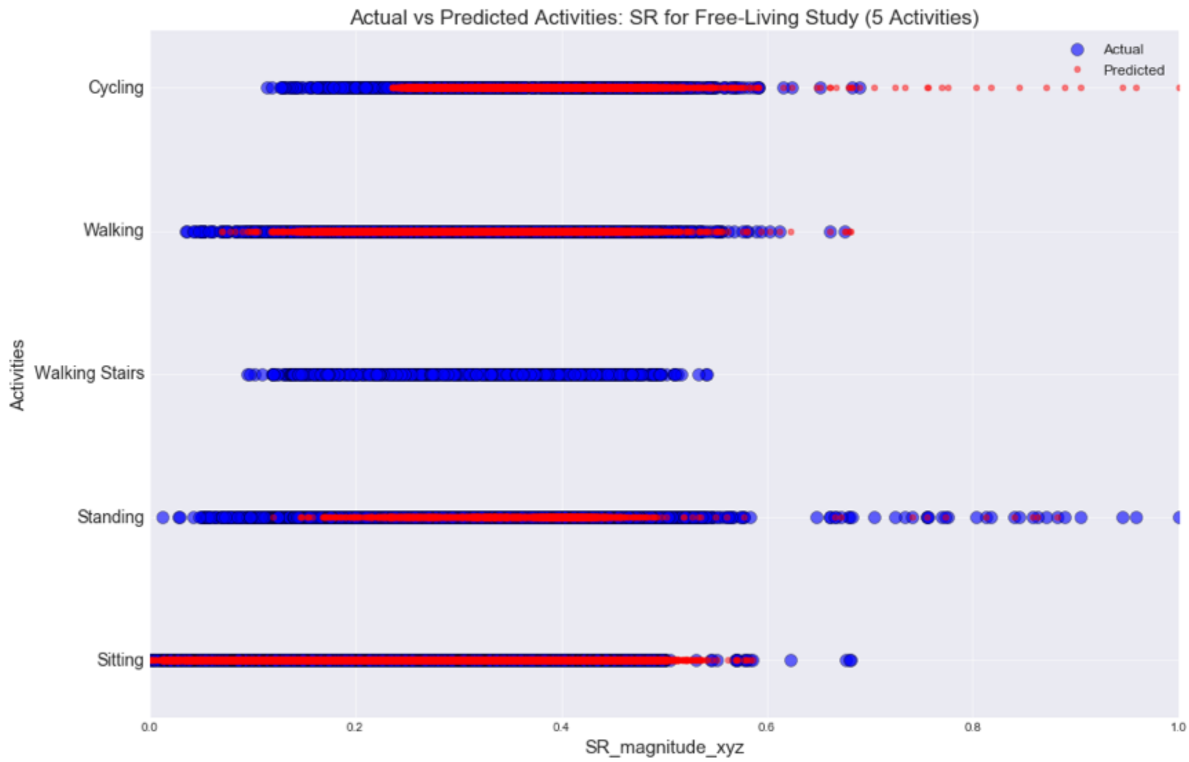


Figure D.4: Actual versus predicted activities for the test set in the Free-Living Study. The prediction is based on SR sensor for detecting 5 activities.

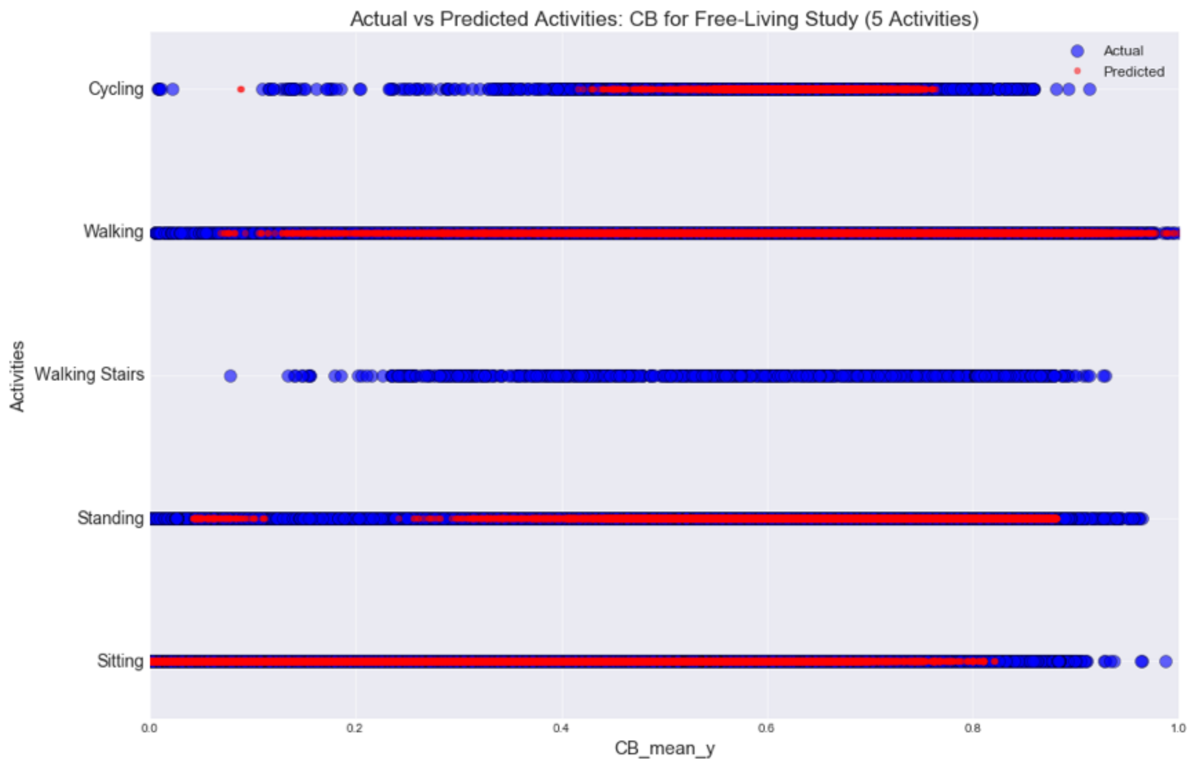


Figure D.5: Actual versus predicted activities for the test set in the Free-Living Study. The prediction is based on CB sensor for detecting 5 activities.

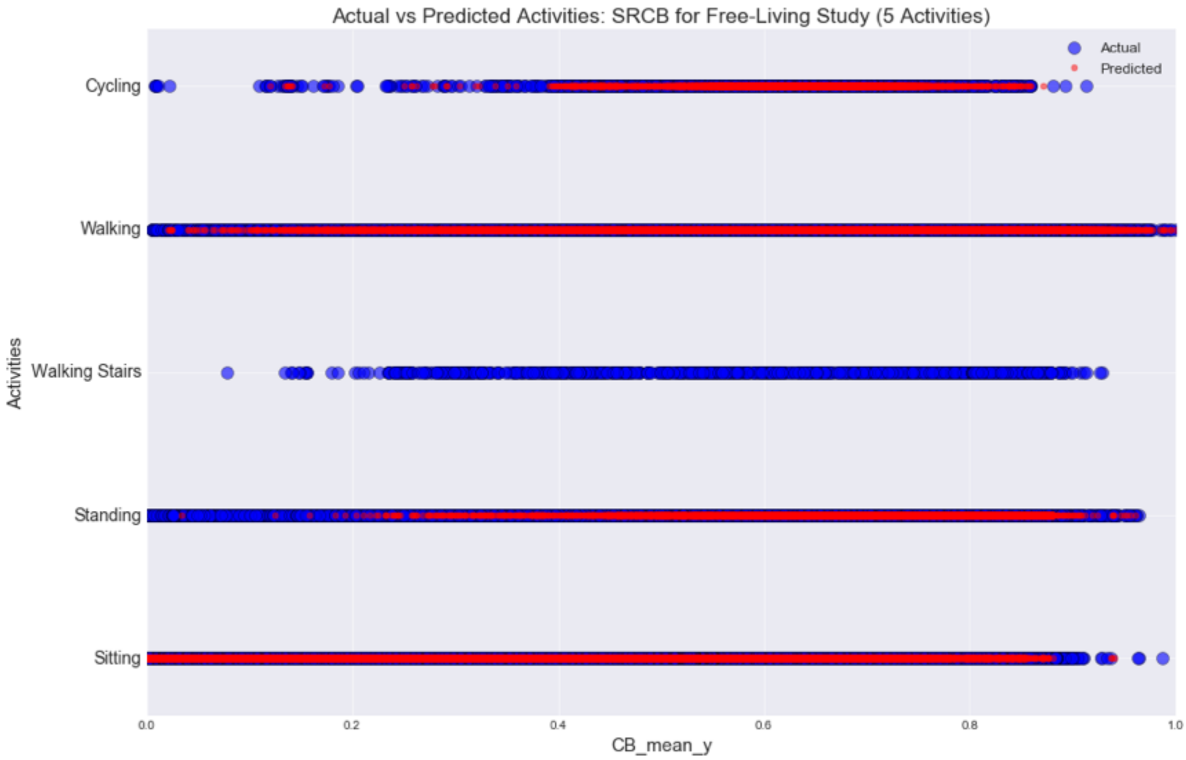


Figure D.6: Actual versus predicted activities for the test set in the Free-Living Study. The prediction is based on SRCB sensors for detecting 5 activities.

Bibliography

[1] X. F. Teng, Y. T. Zhang, C. C. Y. Poon, and P. Bonato, *"Wearable Medical Systems for p-Health,"* IEEE Reviews in Biomedical Engineering, vol. 1, pp. 62-74, 2008.

[2] P. Bonato, *"Wearable Sensors and Systems,"* IEEE Engineering in Medicine and Biology Magazine, vol. 29, no. 3, pp. 25-36, 2010.

[3] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, *"A review of wearable sensors and systems with application in rehabilitation,"* J Neuroeng Rehabil, vol. 9, p. 21, 2012.

[4] Y. Saez, A. Baldominos, and P. Isasi, *"A Comparison Study of Classifier Algorithms for Cross-Person Physical Activity Recognition,"* Sensors (Basel), vol. 17, no. 1, Jan 2017.

[5] O. D. Lara and M. Labrador, A Survey on Human Activity Recognition Using Wearable Sensors. 2013, pp. 1192-1209.

[6] S. J. H. Biddle and M. Asare, *"Physical activity and mental health in children and adolescents: a review of reviews,"* British Journal of Sports Medicine, vol. 45, no. 11, pp. 886-895, 2011.

[7] W. J. Chodzko-Zajko, *"Exercise and Physical Activity for Older Adults,"* Kinesiology Review, vol. 3, no. 1, pp. 101-106, 2014.

[8] A. L. Brosse, E. S. Sheets, H. S. Lett, and J. A. Blumenthal, *"Exercise and the treatment of clinical depression in adults: recent findings and future directions,"* Sports Med, vol. 32, no. 12, pp. 741-60, 2002.

[9] R. S. Dood, J. C. Stevens, C. Beck, R. M. Dubinsky, J. A. Kaye, L. Gwyther, R. C. Mohs, L. J. Thal, P. J. Whitehouse, S. T. DeKosky, and J. L. Cummings, *"Practice parameter: management of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology,"* Neurology, vol. 56, no. 9, pp. 1154-66, May 2001.

[10] R. J. Sigal, G. P. Kenny, D. H. Wasserman, and C. Castaneda-Sceppa, *"Physical activity/exercise and type 2 diabetes,"* Diabetes Care, vol. 27, no. 10, pp. 2518-39, Oct 2004.

[11] *"Screening for obesity in adults: recommendations and rationale,"* Ann Intern Med, vol. 139, no. 11, pp. 930-2, Dec 2003.

[12] S. Going, T. Lohman, L. Houtkooper, L. Metcalfe, H. Flint-Wagner, R. Blew, V. Stanford, E. Cussler, J. Martin, P. Teixeira, M. Harris, L. Milliken, A. Figueroa-Galvez, and J. Weber, *"Effects of exercise on bone mineral density in calcium-replete postmenopausal women with and without hormone replacement therapy,"* Osteoporos Int, vol. 14, no. 8, pp. 637-43, Aug 2003.

[13] G. Office of the Surgeon, *"Reports of the Surgeon General,"* in Bone Health and Osteoporosis: A Report of the Surgeon GeneralRockville (MD): Office of the Surgeon General (US), 2004.

[14] L. S. Pescatello, B. A. Franklin, R. Fagard, W. B. Farquhar, G. A. Kelley, and C. A. Ray, *"American College of Sports Medicine position stand. Exercise and hypertension,"* Med Sci Sports Exerc, vol. 36, no. 3, pp. 533-53, Mar 2004.

[15] P. D. Thompson, D. Buchner, I. L. Pina, G. J. Balady, M. A. Williams, B. H. Marcus, K. Berra, S. N. Blair, F. Costa, B. Franklin, G. F. Fletcher, N. F. Gordon, R. R. Pate, B. L. Rodriguez, A. K. Yancey, and N. K. Wenger, *"Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease: a statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity),"* Circulation, vol. 107, no. 24, pp. 3109-16, Jun 2003.

[16] P. D. Thompson, S. F. Crouse, B. Goodpaster, D. Kelley, N. Moyna, and L. Pescatello, *"The acute versus the chronic response to exercise,"* Med Sci Sports Exerc, vol. 33, no. 6 Suppl, pp. S438-45; discussion S452-3, Jun 2001.

[17] G. F. Fletcher, G. J. Balady, E. A. Amsterdam, B. Chaitman, R. Eckel, J. Fleg, V. F. Froelicher, A. S. Leon, I. L. Pina, R. Rodney, D. A. Simons-Morton, M. A. Williams, and T. Bazzarre, *"Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association,"* Circulation, vol. 104, no. 14, pp. 1694-740, Oct 2001.

[18] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, *"Sensor-Based Activity Recognition,"* Trans. Sys. Man Cyber Part C, vol. 42, no. 6, pp. 790-808, 2012.

[19] P. Kumari, L. Mathew, and P. Syal, *"Increasing trend of wearables and multimodal interface for human activity monitoring: A review,"* Biosens Bioelectron, vol. 90, pp. 298-307, Apr 2017.

[20] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, *"Activity classification using realistic data from wearable sensors,"* IEEE Trans Inf Technol Biomed, vol. 10, no. 1, pp. 119-28, Jan 2006.

[21] M. Arif and A. Kattan, *"Physical Activities Monitoring Using Wearable Acceleration Sensors Attached to the Body,"* PLOS ONE, vol. 10, no. 7, p. e0130851, 2015.

[22] A. Mannini and A. M. Sabatini, *"Machine learning methods for classifying human physical activity from on-body accelerometers,"* Sensors (Basel), vol. 10, no. 2, pp. 1154-75, 2010.

[23] J. Testa, *"The Thomson Reuters Journal Selection Process,"* Transnational Corporations Review, vol. 1, no. 4, pp. 59-66, 2009.

[24] A. V. Kulkarni, B. Aziz, I. Shams, and J. W. Busse, *"Comparisons of citations in web of science, scopus, and google scholar for articles published in general medical journals,"* JAMA, vol. 302, no. 10, pp. 1092-1096, 2009.

[25] G. Scholar. Inclusion Guidelines for Webmasters. [Online].
Available: https://scholar.google.com/intl/en/scholar/inclusion.html.

[26] S. C. Mukhopadhyay, *"Wearable Sensors for Human Activity Monitoring: A Review,"* IEEE Sensors Journal, vol. 15, no. 3, pp. 1321-1330, 2015.

[27] J. Skotte, M. Korshoj, J. Kristiansen, C. Hanisch, and A. Holtermann, *"Detection of physical activity types using triaxial accelerometers,"* J Phys Act Health, vol. 11, no. 1, pp. 76-84, Jan 2014.

[28] K. Altun, B. Barshan, and O. Tunçel, *"Comparative study on classifying human activities with miniature inertial and magnetic sensors,"* Pattern Recognition, vol. 43, no. 10, pp. 3605-3620, 2010.

[29] H. Leutheuser, D. Schuldhaus, and B. M. Eskofier, *"Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset,"* PLOS ONE, vol. 8, no. 10, p. e75196, 2013.

[30] D. Ashbrook and T. Starner, *"Using GPS to learn significant locations and predict movement across multiple users,"* Personal and Ubiquitous Computing, journal article vol. 7, no. 5, pp. 275-286, October 2003.

[31] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, *"Learning and inferring transportation routines,"* Artificial Intelligence, vol. 171, no. 5, pp. 311-331, 2007.

[32] D. J. Patterson, L. Liao, D. Fox, and H. Kautz, *"Inferring High-Level Behavior from Low-Level Sensors,"* in UbiComp 2003: Ubiquitous Computing: 5th International Conference, Seattle, WA, USA, October 12-15, 2003. Proceedings, A. K. Dey, A. Schmidt, and J. F. McCarthy, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 73-89, 2003.

[33] D. Riboni and C. Bettini, *"COSAR: hybrid reasoning for context-aware activity recognition,"* Personal and Ubiquitous Computing, journal article vol. 15, no. 3, pp. 271-289, 2011.

[34] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, *"Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor,"* in 2007 11th IEEE International Symposium on Wearable Computers, 2007, pp. 37-40.

[35] Ó. D. Lara, A. J. Pérez, M. A. Labrador, and J. D. Posada, *"Centinela: A human activity recognition system based on acceleration and vital sign data,"* Pervasive and Mobile Computing, vol. 8, no. 5, pp. 717-729, 2012.

[36] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, *"Physical Human Activity Recognition Using Wearable Sensors,"* Sensors, vol. 15, no. 12, p. 29858, 2015.

[37] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, and D. Finlay, *"Optimal Placement of Accelerometers for the Detection of Everyday Activities,"* Sensors, vol. 13, no. 7, p. 9183, 2013.

[38] N. Pannurat, S. Thiemjarus, E. Nantajeewarawat, and I. Anantavrasilp, *"Analysis of Optimal Sensor Positions for Activity Classification and Application on a Different Data Collection Scenario,"* Sensors (Basel, Switzerland), vol. 17, no. 4, p. 774, 2017.

[39] H. Gjoreski, M. Lustrek, and M. Gams, *"Accelerometer Placement for Posture Recognition and Fall Detection,"* presented at the Proceedings of the 2011 Seventh International Conference on Intelligent Environments, 2011.

[40] M. Gjoreski, H. Gjoreski, M. Lustrek, and M. Gams, *"How Accurately Can Your Wrist Device Recognize Daily Activities and Detect Falls?,"* Sensors (Basel), vol. 16, no. 6, June 2016.

[41] D. O. Olguın and A.S. Pentland, *"Human Activity Recognition: Accuracy Across Common Locations for Wearable Sensors. ,"* Proceedings of 2006 10th IEEE International Symposium on Wearable Computers, Montreux, Switzerland, 11–14 October 2006; pp. 11–13.

[42] F. Chamroukhi, S. Mohammed, D. Trabelsi, L. Oukhellou, and Y. Amirat, *"Joint segmentation of multivariate time series with hidden process regression for human activity recognition,"* Neurocomputing, vol. 120, no. Supplement C, pp. 633-644, 2013.

[43] C.-C. Yang and Y.-L. Hsu, *"A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring,"* Sensors (Basel, Switzerland), vol. 10, no. 8, pp. 7772-7788, 2010.

[44] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, *"Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring,"* IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 1, pp. 156-167, 2006.

[45] M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. F. Coster, *"Classification of basic daily movements using a triaxial accelerometer,"* Medical and Biological Engineering and Computing, journal article vol. 42, no. 5, pp. 679-687, Sept 2004.

[46] P. Gupta and T. Dallas, *"Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer,"* IEEE Transactions on Biomedical Engineering, vol. 61, no. 6, pp. 1780-1786, 2014.

[47] T. P. Kao, C. W. Lin, and J. S. Wang, *"Development of a portable activity detector for daily activity recognition,"* in 2009 IEEE International Symposium on Industrial Electronics, 2009, pp. 115-120.

[48] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, *"Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors,"* Sensors, vol. 16, no. 4, p. 426, 2016.

[49] S. Chernbumroong, A. S. Atkins, and H. Yu, *"Activity classification using a single wrist-worn accelerometer,"* in 2011 5th International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA) Proceedings, 2011, pp. 1-6.

[50] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, *"Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers,"* Pattern Recognition Letters, vol. 29, no. 16, pp. 2213-2220, 2008.

[51] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, *"Activity recognition and monitoring using multiple sensors on different body positions,"* in International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06), 2006, pp. 4 pp.-116.

[52] D. S. Morillo, J. L. R. Ojeda, L. F. C. Foix, and A. L. Jiménez, *"An Accelerometer-Based Device for Sleep Apnea Screening,"* IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 491-499, 2010.

[53] Y.-L. Kuo, K. M. Culhane, P. Thomason, O. Tirosh, and R. Baker, *"Measuring distance walked and step count in children with cerebral palsy: An evaluation of two portable activity monitors,"* Gait & Posture, vol. 29, no. 2, pp. 304-310, 2009.

[54] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick, *"Age-related differences in walking stability,"* Age Ageing, vol. 32, no. 2, pp. 137-42, Mar 2003.

[55] J. Bort-Roig, N. D. Gilson, A. Puig-Ribera, R. S. Contreras, and S. G. Trost, *"Measuring and influencing physical activity with smartphone technology: a systematic review,"* Sports Med, vol. 44, no. 5, pp. 671-86, May 2014.

[56] A. G. Bonomi, A. H. Goris, B. Yin, and K. R. Westerterp, *"Detection of type, duration, and intensity of physical activity using an accelerometer,"* Med Sci Sports Exerc, vol. 41, no. 9, pp. 1770-7, Sep 2009.

[57] A. G. Bonomi, G. Plasqui, A. H. Goris, and K. R. Westerterp, *"Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer,"* J Appl Physiol (1985), vol. 107, no. 3, pp. 655-61, Sep 2009.

[58] L. Bao and S. S. Intille, *"Activity Recognition from User-Annotated Acceleration Data,"* in Pervasive Computing: Second International Conference, PERVASIVE 2004, Linz/Vienna, Austria, April 21-23, 2004. Proceedings, A. Ferscha and F. Mattern, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1-17.

[59] E. Garcia-Ceja, R. Brena, J. Carrasco-Jimenez, and L. Garrido, *"Long-Term Activity Recognition from Wristwatch Accelerometer Data,"* Sensors, vol. 14, no. 12, p. 22500, 2014.

[60] A. M. Khan, Y. K. Lee, S. Y. Lee, and T. S. Kim, *"A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer,"* IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 5, pp. 1166-1172, 2010.

[61] H. Xu, J. Liu, H. Hu, and Y. Zhang, *"Wearable Sensor-Based Human Activity Recognition Method with Multi-Features Extracted from Hilbert-Huang Transform,"* Sensors (Basel, Switzerland), vol. 16, no. 12, p. 2048, 2016.

[62] B. Kikhia, M. Gomez, L. Jiménez, J. Hallberg, N. Karvonen, and K. Synnes, *"Analyzing Body Movements within the Laban Effort Framework Using a Single Accelerometer,"* Sensors, vol. 14, no. 3, p. 5725, 2014.

[63] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, *"Activity recognition from accelerometer data,"* presented at the Proceedings of the 17th conference on Innovative applications of artificial intelligence - Volume 3, Pittsburgh, Pennsylvania, 2005.

[64] S. Zhang, P. Murray, R. Zillmer, R. G. Eston, M. Catt, and A. V. Rowlands, *"Activity classification using the GENEA: optimum sampling frequency and number of axes,"* Med Sci Sports Exerc, vol. 44, no. 11, pp. 2228-34, Nov 2012.

[65] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee, and J.-S. Wang, *"Online classifier construction algorithm for human activity detection using a tri-axial accelerometer,"* Applied Mathematics and Computation, vol. 205, no. 2, pp. 849-860, 2008.

[66] M. A. Calabró, J.-M. Lee, P. F. Saint-Maurice, H. Yoo, and G. J. Welk, *"Validity of physical activity monitors for assessing lower intensity activity in adults,"* The International Journal of Behavioral Nutrition and Physical Activity, vol. 11, p. 119, 2014.

[67] B. Barshan and M. C. Yüksek, *"Recognizing Daily and Sports Activities in Two Open Source Machine Learning Environments Using Body-Worn Sensor Units,"* The Computer Journal, vol. 57, no. 11, pp. 1649-1667, 2014.

[68] M. Sysoev, A. Kos, and M. Pogačnik, *"Noninvasive stress recognition considering the current activity,"* Personal and Ubiquitous Computing, journal article vol. 19, no. 7, pp. 1045-1052, Octob 2015.

[69] S. G. Trost, Y. Zheng, and W. K. Wong, *"Machine learning for activity recognition: hip versus wrist data,"* Physiol Meas, vol. 35, no. 11, pp. 2183-9, Nov 2014.

[70] A. Moschetti, L. Fiorini, D. Esposito, P. Dario, and F. Cavallo, *"Recognition of Daily Gestures with Wearable Inertial Rings and Bracelets,"* Sensors (Basel), vol. 16, no. 8, Aug 22 2016.

[71] Scikit-learn. [Online]. Available: http://scikit-learn.org/stable/.

[72] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, *"A Survey of Online Activity Recognition Using Mobile Phones,"* Sensors, vol. 15, no. 1, p. 2059, 2015.

[73] D. Figo, P. C. Diniz, D. R. Ferreira, Jo, and o. M. Cardoso, *"Preprocessing techniques for context recognition from accelerometer data,"* Personal Ubiquitous Comput., vol. 14, no. 7, pp. 645-662, 2010.

[74] E. Munguia Tapia, *"Using machine learning for real-time activity recognition and estimation of energy expenditure,"* Massachusetts Institute of Technology. Dept. of Architecture. Program in Media Arts and Sciences., Massachusetts Institute of Technology, 2008.

[75] 1.13. Feature selection. [Online]. Available: http://scikit-learn.org/stable/modules/feature_selection.html.

[76] Scikit-learn: How to obtain True Positive, True Negative, False Positive and False Negative. [Online]. Available: https://stackoverflow.com/questions/31324218/scikit-learn-how-to-obtain-true-positive-true-negative-false-positive-and-fal.

[77] sklearn.metrics.cohen_kappa_score. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html#sklearn.metrics.cohen_kappa_score.

[78] J. Cohen, *"A Coefficient of Agreement for Nominal Scales,"* Educational and Psychological Measurement, vol. 20, no. 1, pp. 37-46, 1960.

[79] Y. Vaizman, K. Ellis, and G. Lanckriet, *"Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches,"* IEEE Pervasive Computing, vol. 16, no. 4, pp. 62-74, 2017.

[80] sklearn.metrics.f1_score. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html.

[81] 3.2.4.3.1. sklearn.ensemble.RandomForestClassifier. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[82] sklearn.model_selection.GridSearchCV. [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

[83] imblearn.over_sampling.SMOTE. [Online]. Available: http://contrib.scikit-learn.org/imbalanced-learn/stable/generated/imblearn.over_sampling.SMOTE.html.