Identifying Children's Activities

Development of a wearable to assess the activities performed by free-living children

Karen Rijnders

A proof-of-concept study on movement identification by experimentally obtained sensor data from adults performing instructed activities under free-living conditions

TUDelft

Identifying Children's Activities

Development of a wearable to assess the activities performed by free-living children

by

Karen Rijnders

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday October 7, 2020 at 1:00 pm

Student number:	4374010	
Project duration:	February 2020 – September 2020	
Thesis committee:	Prof. dr. H.E.J. Veeger,	TU Delft (3ME), Supervisor
	MSc. M.A. Gielen,	TU Delft (IDE), Supervisor
	MSc. L. Gomaz,	TU Delft (EWI)

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

IDENTIFYING CHILDREN'S ACTIVITIES

VERZAMEL JE FIT

DEVELOPMENT OF A LOW-COST WEARABLE TO ASSESS ACTIVITIES PERFORMED BY CHILDREN

A TU Delft Initiative

Introduction

Moving frequently and variably is important for mental and physical health. It is associated with an improved confidence as well as a reduction in anxiety, stress and depression. Physical benefits of exercise are lower HDL- cholesterol levels, lower blood pressure, decreased chances of obesity, stronger bones and a better immune system. Children are therefore recommended to move at least one hour a day at a moderate intensity level and additionally implement muscle and bone strengthening activities for minimally three time per week.

The Problem

Children in The Netherlands do not achieve a sufficient amount of physical activity. Therefore, measures must be taken to intrinsically motivate Dutch children to move more frequently. Meanwhile, it is useful to collect data on physical activity patterns of children to determine whether interventions are necessary and whether interventions are effective. This information can additionally be used as part of the intervention when rewarding children based on positive changes in their behaviour.

Verzamel je Fit is part of an effort to motivate children to move more frequently and perform varying movements.

Prospected Returns:

- Feedback on amount and type of physical activity, delivered by the user's favorite athletes
- Data collection on movements made by children for policy-making and evaluating the effect of interventions

Why Invest In "Verzamel je Fit"?

Prospected Features

- Cool gadget on the right wrist
- Uses triaxial accelerometer and heart rate measurements
- Transfer data to smartphone application through an NFC tag
- Charge battery wirelessly via a smartphone
- Up to 2 days of data on movements
- High tech smartphone application with machine learning algorithm
- Athlete of choice encourages variety and increased quantity of movements

Estimated Costs*

Triaxial Accelerometer:	€ 0.50
Heart Rate sensor:	€ 0.70
NFC tag:	€ 0.20
Battery:	€ 1.00
Wrist band:	€ 0.10
Micro controller (with storage):	€ 2.00
	€ 4.50

IMPROVE & VALIDATE

Investements are needed to further develop the algorithm and wearable

- Initial results suggest it will be difficult to accurately identify various activites in free-living children
- However, with more data from both genders and all relevant ages it seems possible

Currently Accuracy: 84%, Precision: 70-91%
Desired Precision: >90%

TU Delft Sports Engineering Institute

* All price indications are derived from products with suitable specifications listed on alibaba.com
Images of children playing: https://www.freepik.com/vectors/children, Children vector created by brgfx

Summary

Dutch children achieve an insufficient amount of physical activity. For that reason, Delft University of Technology is looking to develop a wearable that contributes to motivating children to move more frequently and more intensely while collecting data on the movements made by its wearer. In this thesis it was investigated whether it would be possible to identify the activities performed by free-living children and what type of data from which body placement should be obtained in order to do so. Due to restrictions resulting from COVID-19, triaxial accelerometer- and gyroscope measurements of typical child play activities were carried out on various body parts of eight adults. A Long-Short Term Memory algorithm was applied to short sequential sections of summarized accelerometer data. This algorithm gives a prediction of the activity executed by the wearer of a triaxial accelerometer for every 10 seconds in time. The effects of the wear-site (left wrist, right wrist, right hip, left ankle or right ankle), type of accelerometer (low noise or wide range) and epoch length (0.25, 0.33, 0.5 or 1 second) were studied. The shortest epoch length was found to result in the most precise predictions per activity and the highest overall accuracy for classifying the activities. The hip and right wrist placement perform better than the other locations. A wrist placement is favored over the hip because a heart rate sensor can be added to the former. Measuring the heart rate in combination with classifying the activities performed gives insight in both intensity and variety of movements made by children.

To further increase the performance of these classifications, predictions made with a score below a certain threshold, 0.775 in this thesis, can be excluded. This will decrease the amount of classifications made but it improves the accuracy as well as causing the precision with which each activity is recognized to rise. Without excluding results with a score below this threshold value, classifications of low noise 0.25 second epoch data from the right wrists have an accuracy of 74.8% and a precision per activity of 54.5% - 82.1%. Removing the more uncertain predictions yields an accuracy of 84.0% and a precision between 57.8% - 91.5%. Clustering specific activities, such as sitting and lying down, increases the precision considerably. The accuracy and precisions when applying sitting and lying down as one cluster in combination with removing the uncertain predictions for the low noise sensor become 85.4% and 70.1% - 92.2%. Before this algorithm can be successfully implemented in combination with the intended wearable, the precision with which each individual activity is identified should be >90%. To transfer the collected accelerometer- and heart rate data to a smartphone, it is recommended to use NFC technology. Through a smartphone application it can then be made available for research purposes and the children can get feedback on the variety and quantity of their physical activity of the past two days. A battery that is appropriately sized for a child's wrist wearable will only be able to power the wearable for a few days. This is an insufficient battery life for the intended use, so the battery has to be charged. NFC technology additionally offers energy harvesting capabilities, making it possible to wirelessly charge the wearable via a smartphone. From this exploratory research it can be concluded that it will be challenging to develop a low-cost wearable that can identify activities and measure how frequently free-living children are physically active. This is the case for both software, the algorithm that predicts which activities were performed, and hardware, where the most pressing challenge is the battery life. However, it is believed that with more extensive research it is possible to create a fully operational wearable.

Contents

1 Introduction **5**
1.1 Background and Problem analysis . . . 5
1.2 Existing devices . . . 5
1.3 Framing the research. . . . 6
1.4 Research question . . . 7
1.5 Nota Bene . . . 7

2 Methods **9**
2.1 Data collection . . . 9
2.2 Data preparation and activity identification algorithm . . . 10
2.3 Evaluation of the model . . . 11

3 Results **13**
3.1 General outcomes . . . 13
3.2 Gyroscope, magnitude and multiple sensor information . . . 13
3.3 Identification of activities . . . 15
3.4 Activity-Intensity identification . . . 17
3.5 Categorizing uncertain predictions as ”unknown”. . . . 17
3.6 Testing the algorithm on children’s data . . . 18
3.7 Effect of more training data on classification accuracy . . . 18
3.8 Calculations for data transfer capacity and power demand . . . 19

4 Discussion **21**
4.1 General remarks . . . 21
4.2 Gyroscope, magnitude and multiple sensors . . . 21
4.3 Identification of activities . . . 22
4.4 Supplementary categorization of intensities. . . . 23
4.5 Children’s data and quantity of data . . . 23
4.6 Comparison with existing research . . . 24
4.7 Data transfer and power supply implications . . . 24

5 Conclusion **25**

Bibliography **27**

Appendices **29**

A Literature review **31**

B Scripts **52**
B.1 Reading data . . . 52
B.2 Segmenting data . . . 56
B.3 Preparing data . . . 59
B.4 Reshape, Randomize, Train and Test data with overlap through bootstrapping. . . . 62
B.5 Support functions . . . 67
B.5.1 CreateCFS_SW_3D . . . 67
B.5.2 LabelActSequences . . . 67
B.5.3 SlidingWindow_FT . . . 68

C Confusion matrices **69**

D Detailed results **138**

1 Introduction

1.1. Background and Problem analysis

Physical activity (PA) is associated with an improved confidence and a reduction in anxiety, depression and stress [4]. Exercise is additionally beneficial for children's physical health, both in their youth and later life. Lower HDL-cholesterol levels, a lower blood pressure, less obesity, a higher bone-mineral density and a better immune system are examples of observations made among children that participate in PA more often and have healthier eating habits [10], [16], [22]. Moreover, obesity increases the likelihood of degenerative diseases later in life [7]. Other severe consequences of childhood obesity are higher chances of cardiovascular diseases and even cancer in adulthood [27]. In order to reap the aforementioned mental and physical health benefits, or prevent the detriments resulting from an unhealthy lifestyle, all children are recommended to move at least one hour a day at a moderate intensity level, perform at least three times a week bone and muscle strengthening activities and limit the time they spend sedentary (SB) [17]. An important side note is that moving more frequently than these recommendations yields more benefit for the child's body and mind. This study will focus on the Dutch context in particular and for Dutch children the same guidelines are upheld [26]. Dutch children do however not achieve a sufficient amount of PA[1], neither do children in other parts of the world [8]. For that reason, it is essential to encourage children to engage more in PA.

To realize more activity among children, Delft University of Technology is looking to develop a wearable that assists in intrinsically motivating children to perform movements. The choice for creating a wearable solution was made because this type of product additionally enables data capturing on the movement patterns of free-living children. The intention is that the wearable will be used on a large scale, which can be accomplished by finding a suitable partner for distribution and ensuring that the device is affordable for a large population. Using the data collected by the wearable when worn by children, decisions can be made regarding interventions to stimulate children to move adequately. More specifically, from gathered data it can be determined in principle whether an intervention is necessary, whether a certain intervention is effective and the wearable itself can even be part of the intervention when the personal activity records are used to stimulate exercise in children. Examples of the latter are rewarding children for performing more various movements or for being more active than some time before. The sample of children used for policy making should be large enough to make valid assumptions, which can be facilitated by ensuring that the device is inexpensive enough to be widely adopted. Furthermore, measurements should be somewhat accurate to avoid pointless or mistaken interventions. The two main design requirements of the prospected wearable are therefore affordability and sufficient accuracy.

1.2. Existing devices

There are many ways to encourage children in PA, ranging from (subsidised) social initiatives, for example by enabling free participation in sport classes[2] and implementing more physical activity during a school day[3] to commercial toy-like products, such as trampolines and inline skates, and from team-sports to active video games. It is however impossible to get an accurate overview of movements made by free-living children through solely analysing data from these practices, because data on executed movements are (most of the time) not recorded. The exception is active video gaming, which in fact could generate measurements of PA in children. Nevertheless, it is not feasible to achieve a thorough analysis of free-living children since movements that occur when they are not connected to their active video game are not registered. Similarly, despite their measurement accuracy, smartphones are not a suitable source for measurement data because that would require the smartphone to be at a known placement on the body throughout the entire day. Often children are not even carrying a smartphone while exercising, since the size and weight can cause inconvenience or

[1] www.rijksoverheid.nl/onderwerpen/sport-en-bewegen/sporten-en-bewegen-voor-kinderen
[2] www.oranjefonds.nl/kracht-van-sport
[3] www.jongerenopgezondgewicht.nl/initiatieven/the-daily-mile

because they do not want to damage their device.

An example of a commercially available device that is used to collect data on movements made by adults or children is the Fitbit. The Fitbit Ace 2 is the activity monitor manufactured specifically for children. This device can be purchased for €69.95[4]. Fitbit has implemented incentives for children to get them to move more frequently, such as rewards when an activity goal is achieved and a step challenge, in which they can compete for a trophy in their Fitbit-app against other Fitbit wearers. From the literature study that was conducted as part of this thesis project, it however appeared that Fitbits do not provide an accurate method for measuring intensities of physical activity in free-living children, the review can be found in Appendix A. The Ace 2 has not been investigated by any of the studies included in the literature study, however two other models (Fitbit Zip and Fitbit Charge HR) did not perform well in the field of accuracy. Due to the relatively high costs of the Fitbits, cheaper alternatives of less established brands are flooding the market. Prices of the majority of these models start around €20,-, but they do not have a version specific to children, are not equipped with incentives to stimulate PA in children and the measurement accuracy is questionable[5] [19], [25].

In conclusion, Fitbits are not as accessible as desired for this application due to their costs and even cheaper alternatives are still relatively expensive. Therefore, making general assumptions based on the subgroup that is equipped with a Fitbit yields a potential participation bias: Children that do wear a Fitbit (or alternative) could be more active or inspired by their parents to engage in sports. This population could also contain overweight children that need to keep better track of their activity patterns. Nonetheless, this subgroup will most likely not form a valid representation of society. It is thus desired to develop a wearable in the price range of a few euros that provides children with stimuli to move.

Because of this ambition, a prototype has already been developed. The current prototype allows for learning absolute counts of PA, but it cannot be used to assess different intensity levels or the variety of movements. When the child connects the wearable to a smartphone, the absolute number representing activity count receives a time stamp. Using the difference in time between two exchanges of the measured data and the difference in activity counts between the two moments in time, the child can get 'rewarded' for moving sufficiently within the passed period of time. This is a start towards the ultimate goal and in this research a method will be proposed to get even more knowledge on the PA in free-living children, while keeping the possibility of encouragement open. This thesis will present a concept for a wearable, including a proposition for the placement of the wearable, and supplementary algorithm that can collectively assess the movements made by freely playing children.

1.3. Framing the research

Based on the literature study that was conducted prior to the start of this thesis (Appendix A), some aspects were formulated that were used as input for this thesis. These topics consist of the measurement bias resulting from particular movements and study protocol, the choice of wear-site and the decision regarding sensor types. These subjects will now be explained in more detail.

Starting with the latter, it appears that triaxial accelerometry, biaxial accelerometry and Inertial Measurement Unit (IMU) measurements can potentially determine both quantity and variety of PA in free-living children. Therefore, it was decided to use IMUs for the collection of data in this research, since an IMU contains a triaxial accelerometer, as well as a triaxial gyroscope and triaxial magnetometer. These sensors can respectively measure linear accelerations, angular velocities and orientation with respect to the earth, all over three axes. Because affordability of the wearable is essential, an objective of this thesis is to work with the least amount of information in order to minimize the number of required sensors and data that has to be transferred. For that reason trade-offs between the amount of data that has to be transferred and the prediction-accuracy will be discussed. To the researcher's knowledge, no literature has yet been published on identification of activities from summarized data.

Sensor placements sometimes show contradictory results when compared to other wear-sites. Placing sensors on the hips and around the waist yields similar results, while outcomes of the dominant wrist differed from accelerations found at the non-dominant wrist, Appendix A. The ankle placement has only been researched

[4] `www.fitbit.com/nl/ace2`

[5] `www.gizbot.com/wearable-technology/features/are-fitness-trackers-accurate-in-measuring-vitals/articlecontent-pf117636-068064.html`

three times (Appendix A: Louie et al., Clark, Duncan M. et al.), however these studies report favorable outcomes with respect to the hip/waist and wrist placements. Due to the ambiguity in investigated literature, the choice was made to compare the accuracies of predictions made based on information from the various wear-sites. Because of these considerations and the amount of IMUs that was available, it was decided to distribute the four IMUs over the bodies of the participants as follows: one on each wrist, one on the right hip and the last one on an ankle. No comparison was found between dominant and non-dominant ankle in the literature study, which is why both ankles were investigated.

Some of the researches reported specific activities in which the assessment of intensity levels turned out to be difficult, Appendix A. This was due to the combination of certain sensor placements and the activity that had to be performed. Mounting an accelerometer on the hip while measuring the intensity of various upper body strength games, basketball or active gaming does not result in a high measurement accuracy, for example (Appendix A: Canete Garcia-Prieto et al., Pulsford et al.). Cycling appeared to be difficult to estimate in general, each of the explored wear sites turned out to be unfit for accurate determination of PA intensities (Appendix A: Duncan, M. et al., Kang et al.). Identifying the activities made by youth should contribute to more accurate estimation of PA intensity levels for all types of movements. This study will therefore investigate movements that involve principally the upper body as well as motions that mainly occur in the lower body. Additionally, a cycling trial and a walking/running trial that facilitated capturing the associated forward acceleration, as opposed to measurements on a home trainer or treadmill, were part of the measurement protocol.

1.4. Research question

The overarching theme of this thesis is that TU Delft is aiming to develop a wearable as part of an effort to intrinsically motivate children to move more often and vary the movements they make, while keeping track of the quantity and variety in their PA. This thesis will contribute towards this goal by providing insight into methods for determining which activities children have engaged in. Therefore, the following questions will be answered in the remaining chapters:

- Which data are required for recognizing the movements made by free-living children?
- What is an appropriate method to determine the type of activity from the measurement data?
- How can this data be transferred from the wearable device to a platform that is accessible for research purposes, in a safe and affordable manner?
- What is a feasible way to incorporate these measurement and data transfer techniques in a fully operational wearable device?

In the end of this thesis, practical feasibility of the wearable will be investigated by estimating the required power supply and data transfer capacities for a potential finalized concept. Recommendations regarding future research for improving the algorithm's predictions will furthermore be given.

1.5. Nota Bene

Due to the measures resulting from the COVID-19 pandemic that were taken during the execution of this research, unfortunately it seemed unethical to perform measurements on children. instead, collected data originate from movements made by adults. Data collected from two children in the researcher's personal environment have additionally been used to investigate applicability of the algorithm to measurements on children.

2
Methods

2.1. Data collection

Eight individuals (age: 33.25 ±17.48, 5 females) participated in the trials that were conducted as part of this study. None of the participants were visibly severely over- or underweight. The participants executed eleven different activities that are known to be made by free-living children[1]. All activities were executed in varying orders and the participants had time to rest in between. Some activities were performed at multiple intensities. An overview of the activities, their intensities and duration can be found in table 2.1.

Activity	Speed or Distance	Duration	Intensity level
Lying down	-	1 minute	SB
Sitting	-	1 minute	SB
Writing/colouring	-	1 minute	LPA
Boxing	Slowly	30 seconds	VPA
	As fast as possible	15 seconds	VPA
Cycling	16 km/h	4 minutes	MPA
	20 km/h	4 minutes	VPA
Throwing a toy ball	3 meter	5 throws	LPA
	5 meter	5 throws	MPA
	8 meter	5 throws	MPA
Kicking a soccer ball	3 meter	5 kicks	MPA
	5 meter	5 kicks	VPA
	8 meter	5 kicks	VPA
Jumping over objects	few centimeters	3 jumps	MPA
	20 centimeters	3 jumps	MPA
	35 centimeters	3 jumps	MPA
	45 centimeters	3 jumps	VPA
Rope jumping	One jump per spin	15 seconds	VPA
	Two jumps per spin	15 seconds	VPA
Walking and running	3 km/h	2 minutes	LPA
	5.5 km/h	2 minutes	MPA
	8 km/h	2 minutes	VPA
	10 km/h	2 minutes	VPA

Table 2.1: Activities, instructions and duration as performed by the participants with corresponding intensity levels (Sedentary Behaviour, Light PA, Moderate PA and Vigorous PA)[2] based on estimated Metabolic Equivalents from comparable activities[14], [13], [2]. Boxing, ball throwing and kicking estimates are based on values determined on adults

With the exception of lying down, sitting and writing, all activities were performed outdoors. Materials and protocols used during the activities were chosen to maximally resemble free-living conditions for children. The ball used for throwing was a toy ball of 22 cm diameter, for kicking a size 5 soccer ball was used. Walking and running trials were carried out as modified shuttle tests to include the forward acceleration and simulate turns and short bursts of activity, which would not have been possible had this trial been conducted on a treadmill.

[1] www.mayoclinic.org/healthy-lifestyle/childrens-health/in-depth/fitness/art-20048027

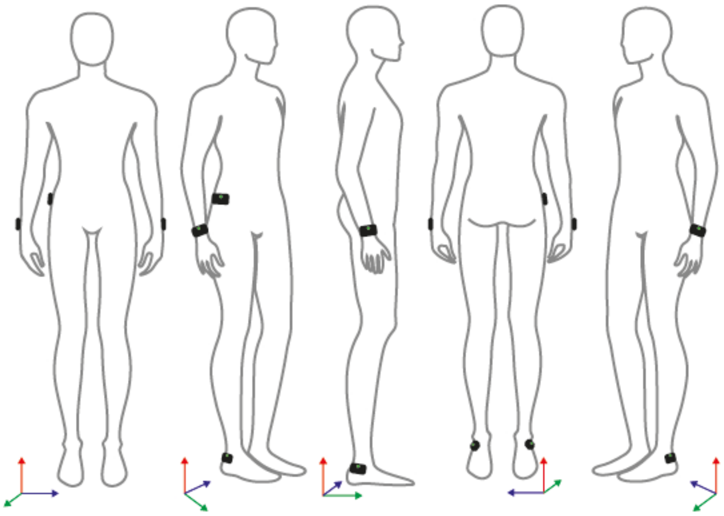


Figure 2.1: IMUs are indicated by the black rectangle, green dots represent the direction of the IMU's LEDs. The blue, green and red arrows represent respectively the positive X, Y and Z directions of the global coordinate system.

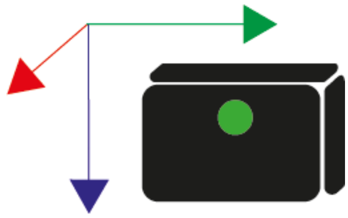


Figure 2.2: The blue, green and red arrows represent respectively the positive X, Y and Z directions of the IMU's coordinate system [20].

For similar reasons, the cycling trials were done on the participants' own bikes, which were equipped with a smartphone holder on the steer. The smartphone was placed in the holder and its screen was locked on the application 'GPS Speedometer', showing the cycling speed and a timer to the participant. The participants were instructed to ride around 16 km/h for four minutes and then 20 km/h for another four minutes while the screen was filmed by an external camera (Nikkei XTREME X6) affixed to the collar of their shirts. All other activities were recorded from a distance, by means of the same video camera held by the researcher. The recordings included a timestamp that was set to the 'real time' accurate to the second, which was used to segment the IMU measurements and link them to the corresponding activity. The time for each activity was monitored via the stopwatch function on a smartphone.

Each participant had given verbal consent to wearing the four IMUs (Shimmer3, Realtime Technologies ltd., Dublin, Ireland) and being recorded with a camera, after being instructed through a protocol approved by HREC (TU Delft's Human Research Ethical Committee). The IMUs were placed on the left wrist, right wrist, right hip and either left or right ankle. Half of the participants was assigned right ankle placement, the other half wore the IMU on the left ankle. The IMUs were positioned with their green LED-lights facing outwards and upwards. In case of the IMUs mounted on the back-hand side of the wrist, the green lights thus pointed in direction of the elbows. The ankle IMUs were placed just behind the lateral malleolus to ensure maximal comfort during movements and minimal movement of the IMU with respect to the participant's body. The image in figure 2.1 shows the positioning of the IMUs. The firmware on the Shimmers was updated to SDLog v0.19.0 and the Shimmers were configured to record at 51.2 Hz, with Undock/Dock as the Start/Stop logging method. For all IMUs, the Low Noise (LN) Accelermeter, Wide Range (WR) Accelerometer, Gyroscope and Magnetometer were enabled. After the measurements were conducted, the data were imported from the Shimmer3's SD cards as an uncalibrated *.mat* file with unix time through the ConsensysBASIC v.1.6.0. software.

2.2. Data preparation and activity identification algorithm

The raw measurement data were imported in MATLAB R2018b and segmented per activity on the second of the IMU's unix time. Data were labeled according to the time registered on the video recordings. The longer sequences resulting from the running, cycling, lying down, sitting, writing and walking were trimmed to 30 seconds of data to minimize unbalanced distributions between the various activities. Based on real time, the data were summarized as the average value per axis over one-fourth, one-third, half and whole seconds, thus 0.25, 0.33, 0.5 and 1 second epochs. From hereon, only the Low Noise and Wide Range accelerometer data were used to develop an algorithm. Per participant, the data were structured as a labeled sequence containing all the activities in a row, in alphabetical order. This order was then randomized and broken down into four short sequences containing three or four activities. Four of these activities appeared in two combinations to make sure that the algorithm would not only learn the activities that were combined, however doing this for all activities would have increased the risk of overfitting. Each of the combinations was then repeated thirteen times, while the order of the activities within one combination was randomized per repetition. Thereafter, these sequences were divided into windows, using a sliding window approach, of a width of ten seconds and an overlap of seven to eight seconds, depending on the epoch length. For each of these windows a scalogram was created through a continuous wavelet operation with filterbank set to twelve voices per octave[3]. The pooling activations of the GoogleNet ”pool5-7x7_s1” layer were applied to each scalogram[4], returning a column vector of 1024 features per window. For each window, the mode of the labels present in the window was calculated and used as the label corresponding to the features of that window. This means that if the window consisted of eight seconds running and two seconds jumping, the label would be ”run”. These features and labels were divided over one test and seven training data sets alternating over the participants, so the data of each participant would be used once for testing and seven times for training. A Long-Short Term Memory (LSTM) network, consisting of a bidirectional LSTM layer with 500 hidden units, was trained on the training sets for each alteration and validated on the test data which the network was still naive to. All MATLAB scripts that were written for the preparation of data as well as training of the neural network and classifying the validation dataset can be found in Appendix B.

2.3. Evaluation of the model

For every composition of seven train and one test data set, the classification accuracies of each short sequence were calculated. The accuracy was defined as the percentage of all windows within one short sequence for which the predicted activity was the same as the actual activity. For each test set 52 accuracies were found. Since each participant wore four sensors on different locations, 208 accuracies were found per sensortype and participant. IBM SPSS Statistics 25 was used to make statistical analyses on the prediction accuracies of each individual. ANOVAs with Tukey’s post-hoc tests were applied to compare difference in accuracy for the various sensor placements, sensor types and epoch lengths. Via paired T Tests it was investigated if the addition of gyroscope data to the accelerations would significantly and relevantly improve the result and if using the magnitude of the three accelerometer axes instead of the three-dimensional data would decrease the performance of the algorithm. Also, combining data of multiple placements was compared to using data from a single placement. Addition of sensor measurements was performed by concatenating the continuous wavelet transforms with filterbanks before the GoogleNet layer activations were executed.

Whereas these are important characteristics for future model and data collection choices, it is also important to examine the performance of the model on identifying the combined and individual activities. An ANOVA with Tukey’s post hoc test was applied to find how well the algorithm could recognize each combination of activities. To investigate the performance on the identification of individual activities and find out which activities were more easy or difficult to classify, recall and precision were calculated for each activity. Hereafter, it was investigated whether this algorithm could predict at which intensity level a certain activity was performed. Hereto the data were divided over 23 classes, representing an activity executed at a specific intensity.

In addition to making a prediction, the classification algorithm was also capable of returning the posterior probabilities of each test window for each of the activities. Based on these scores a simple classification algorithm was developed to minimize faulty classification in case of overall low posterior probabilities by defining these activities as ’unknown’ and excluding them when computing the accuracy. Whether the implementation of an ’unknown’ class contributed to the model’s performance was investigated by means of a paired T Test. In order to get a better understanding of the influence of the amount of data on the performance of the model, linear regression analyses were executed for LSTM networks trained on different amounts of data sets. Classification accuracies of one randomly chosen validation data set were computed for algorithms trained on one to seven datasets. Investigating this regression enables observing how the performance evolves over the incremental size of training data and if performance could be enhanced by generating and training on more data.

Finally, the LSTM network was trained on all eight measurements that were performed on adults while data from two girls (age = 12.1 ±0, fraternal twins) was used for validation. The means of the resulting accuracies were analysed against the means of the accuracies from the adults that were classified using a network trained on seven datasets by means of an independent T Test. For each statistical test the significance level was set to 0.05.

[3] www.mathworks.com/help/wavelet/examples/classify-time-series-using-wavelet-analysis-and-deep-learning.html
[4] www.mathworks.com/help/deeplearning/ug/classify-videos-using-deep-learning.html

3 Results

3.1. General outcomes

Initially, 80 LSTM networks were trained, each with a training time of approximately 50 minutes. A total of 13312 accuracies resulted from the initial classifications of four epoch lengths, two sensor types (LN and WR accelerometer), five wear-sites, eight participants and thirteen alterations of each of the four combinations of activities. Because the mode was used as the operation to label the individual activities within a short sequence, the occurrence of each activity varied slightly. Approximately, the activities represented the following percentages of the total dataset:

Box	**Cycle**	**Jump**	**Kick**	**Lie**	**Rope jump**	**Run**	**Sit**	**Throw**	**Walk**	**Write**
10.9%	7.4%	7.8%	22.9%	3.5%	3.0%	12.4%	3.4%	10.5%	14.6%	3.6%

A summarized comparison of the accuracies for the different wear-sites and epoch lengths can be found in figure 3.1. All epoch lengths are significantly different from each other ($p<0.01$). The use of a WR or a LN accelerometer has no significant influence on the resulting accuracy ($p = 0.151$). The hip placement yielded significantly higher results than the other placements ($p \leq 0.010$), whereas the placements on the left side of the body showed no significant difference in effect of wear-site ($p = 0.065$), as did the right wrist and ankle ($p = 0.781$). Analysis of the combined effects of sensor type and placement for the 0.25 second epoch length reveals that the combinations of WR & Hip, LN & Hip, WR & Right wrist, LN & Right wrist do not show significant differences with respect to each other ($p \geq 0.219$). When investigating the outcomes of the 0.33 second epoch length, it appeared that the same combinations yielded no significantly dissimilar results ($p \geq 0.570$). More resembling combinations can be found, however the above mentioned are the only ones in which for all configurations of sensor type and placement the effects were not significant. Therefore, only the 0.33 and 0.25 second epoch data from the hip and right wrist will be analysed from here on.

3.2. Gyroscope, magnitude and multiple sensor information

The results for summarizing the three dimensional measurements in a magnitude, adding the gyroscope data to the accelerometer information and merging the magnitude of the accelerometer and gyroscope data are visualized in figure 3.2. In all cases, enriching the data with raw gyroscope measurements did not improve the accuracy. Computing the magnitude resulted without exception in a significantly decreased performance with respect to the three dimensional accelerometer data ($p \leq 0.021$). Likewise, the combination of the accelerometer and gyroscope magnitudes showed a significant reduction in prediction accuracies (all $p < 0.001$). Merging the right wrist and hip data does not result in a significant improvement of the accuracies found relative to data from only the hip or wrist, for any of the epoch lengths and sensor types. The changes in mean accuracy in relation to the single sensors and corresponding P values is listed in table 3.1

Placement	With respect to	Low noise accelerometer		Wide range accelerometer	
		0.33 seconds	0.25 seconds	0.33 seconds	0.25 seconds
Right wrist and Hip	Hip	-1.1% (p=0.999)	-9.3% (p<0.001)	-1.4% (p=0.995)	-5.0% (p=0.005)
	Right wrist	-4.2% (p=0.045)	-7.8% (p<0.001)	-0.9% (p=1.000)	-0.1% (p=1.000)

Table 3.1: Difference and P values for the accuracies resulting from the combination of right wrist and hip data with respect to the accuracie for the hip and right wrist individually.

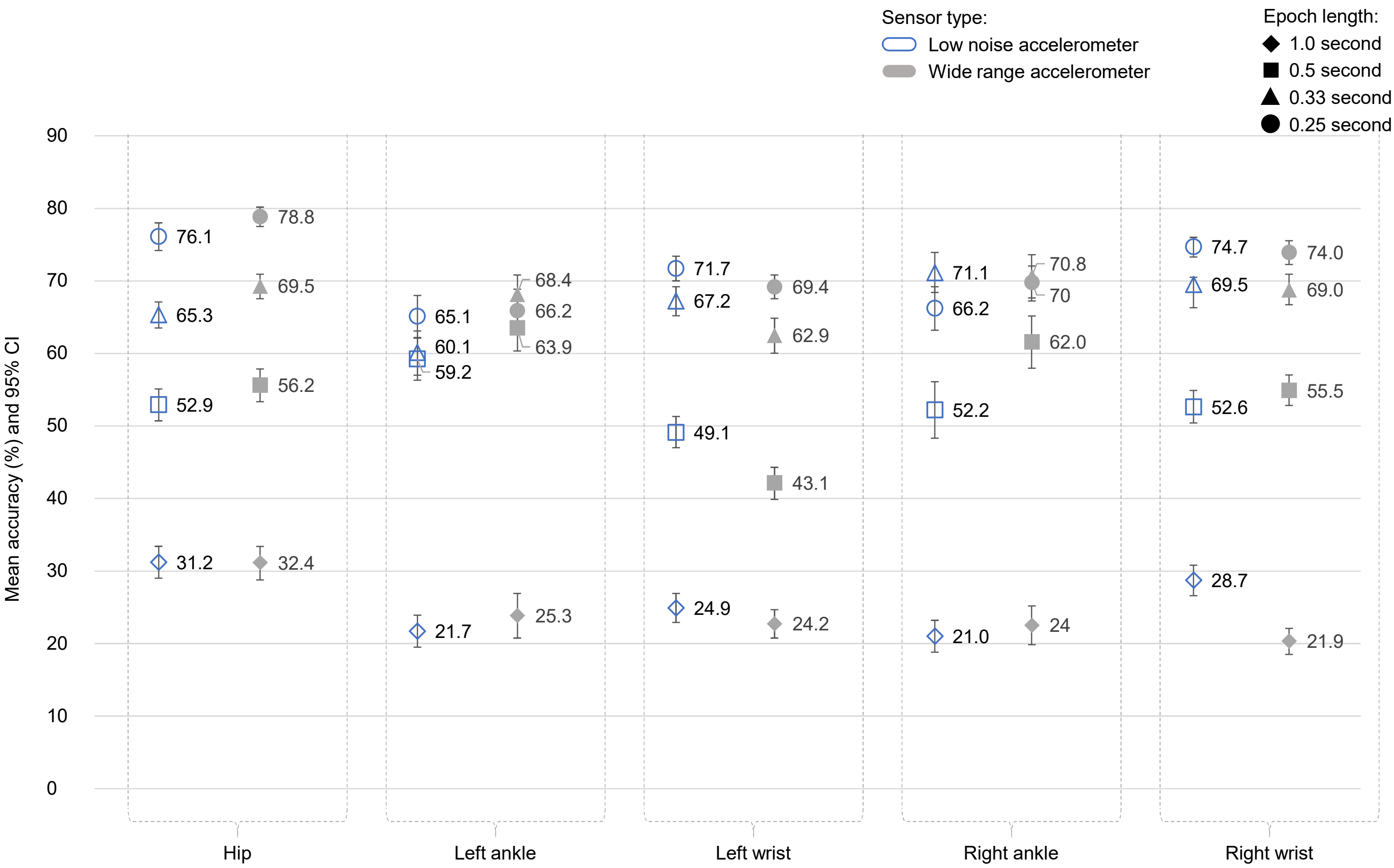


Figure 3.1: The mean accuracy (%) with 95% confidence interval per wear-site, sensor type and epoch length, exact numbers can be found in Appendix D.

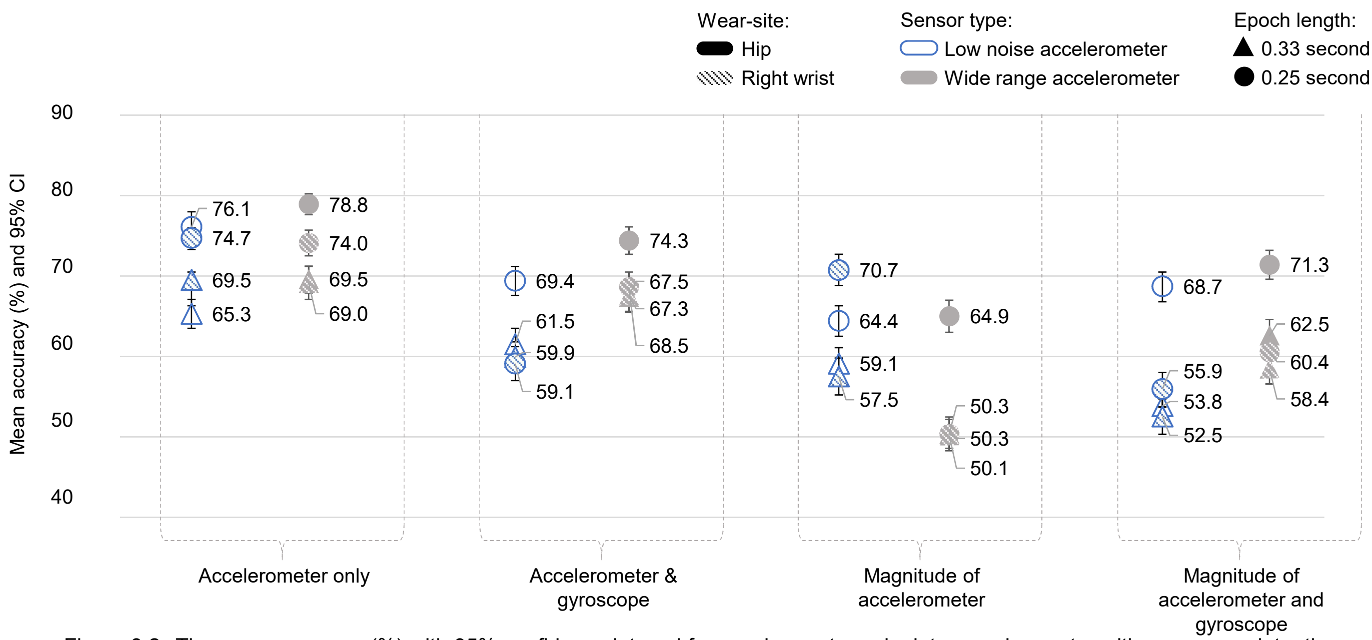


Figure 3.2: The mean accuracy (%) with 95% confidence interval for accelerometer only data, accelerometer with gyroscope data, the magnitude of accelerometer data and the magnitudes of acclereometer and gyroscope data. Exact numbers can be found in Appendix D.

3.3. Identification of activities

Furthermore, it was investigated whether there was a difference in identification accuracy between the four randomly assigned short sequences containing specific activities. The combination ”Jump - Lie - Run - Sit” and ”Box - Kick - Walk - Write” did not result in significantly different accuracies, as did the short sequences containing ”Cycle - Kick - Rope jump - Run” and ”Jump - Lie - Run - Sit” ($p=0.089$ and $p=0.890$, respectively). All other combinations differ significantly ($p \leq 0.012$).

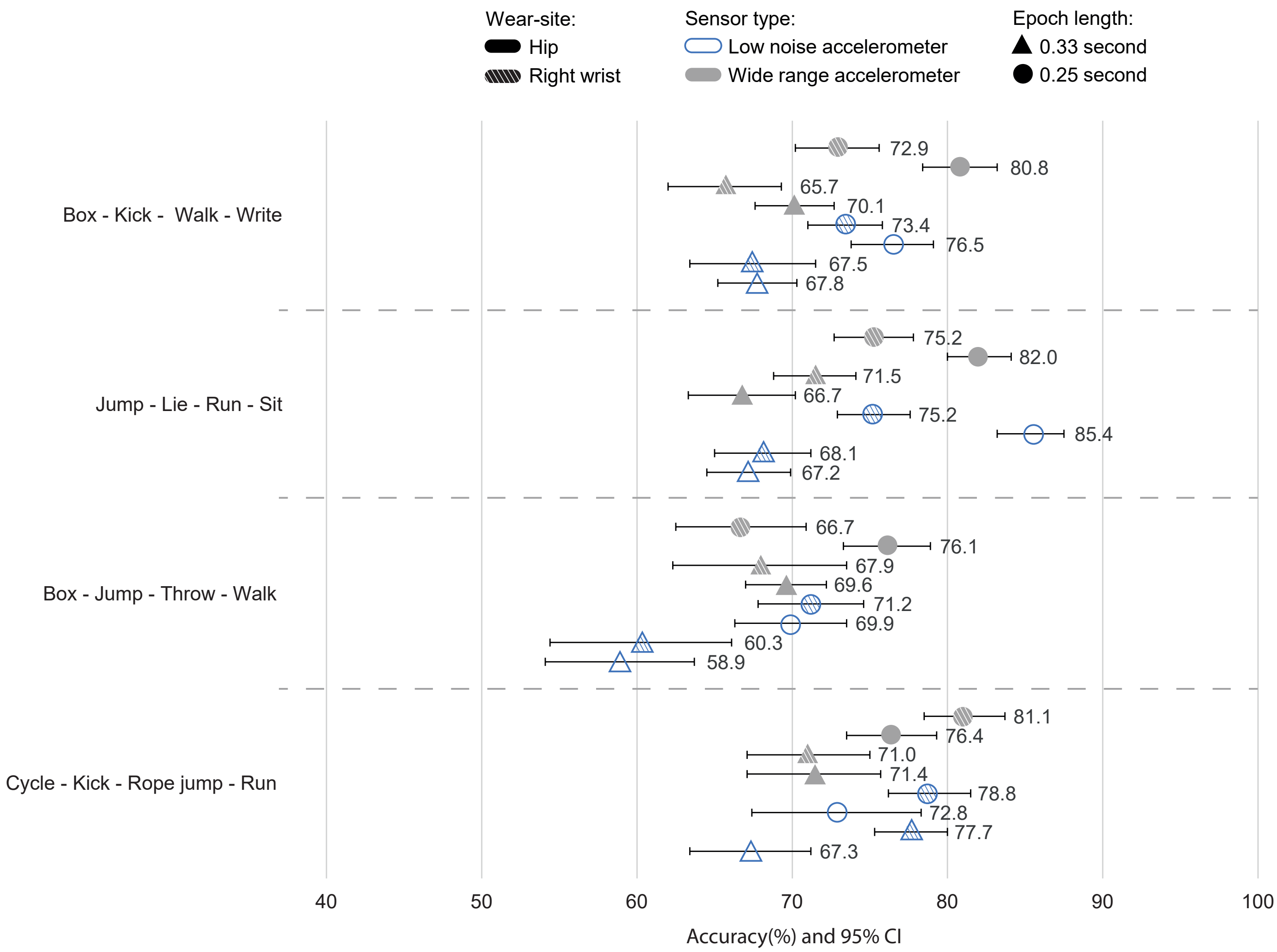


Figure 3.3: The mean accuracy (%) with standard deviation and 95% confidence interval of the predictions per short sequence of activities for measurements from both accelerometers on the right wrist and hip summarized per 0.33 and 0.25 seconds.

Two important performance characteristics of a classification algorithm are precision and recall. Precision is also known as the positive predictive value and this measure is equal to the ratio of true positives versus all positives. In this thesis precision therefore denotes how often a specific activity is correctly recognized out of all the times this activity was predicted. Recall, or sensitivity, refers to the number of activities that are correctly classified. It is calculated as the percentage of true positives among all predictions made for a certain activity. The recall and precision for the individual activities trained and validated on adult measurements are shown in figure 3.4. In this analysis all activities are investigated separately, however it could be acceptable to cluster some specific activities. One instance is the distinction between sitting and lying. Since both activities are sedentary, differentiation between the two is less relevant than identifying the difference in lower and upper body movements. In table 3.2 the changes in precision and recall as a consequence of grouping specific activities and influence on the overall accuracy are given for a few activities that can be perceived as similar. The entire overview of how each activity was classified is displayed in confusion matrices for every alternative in Appendix C.

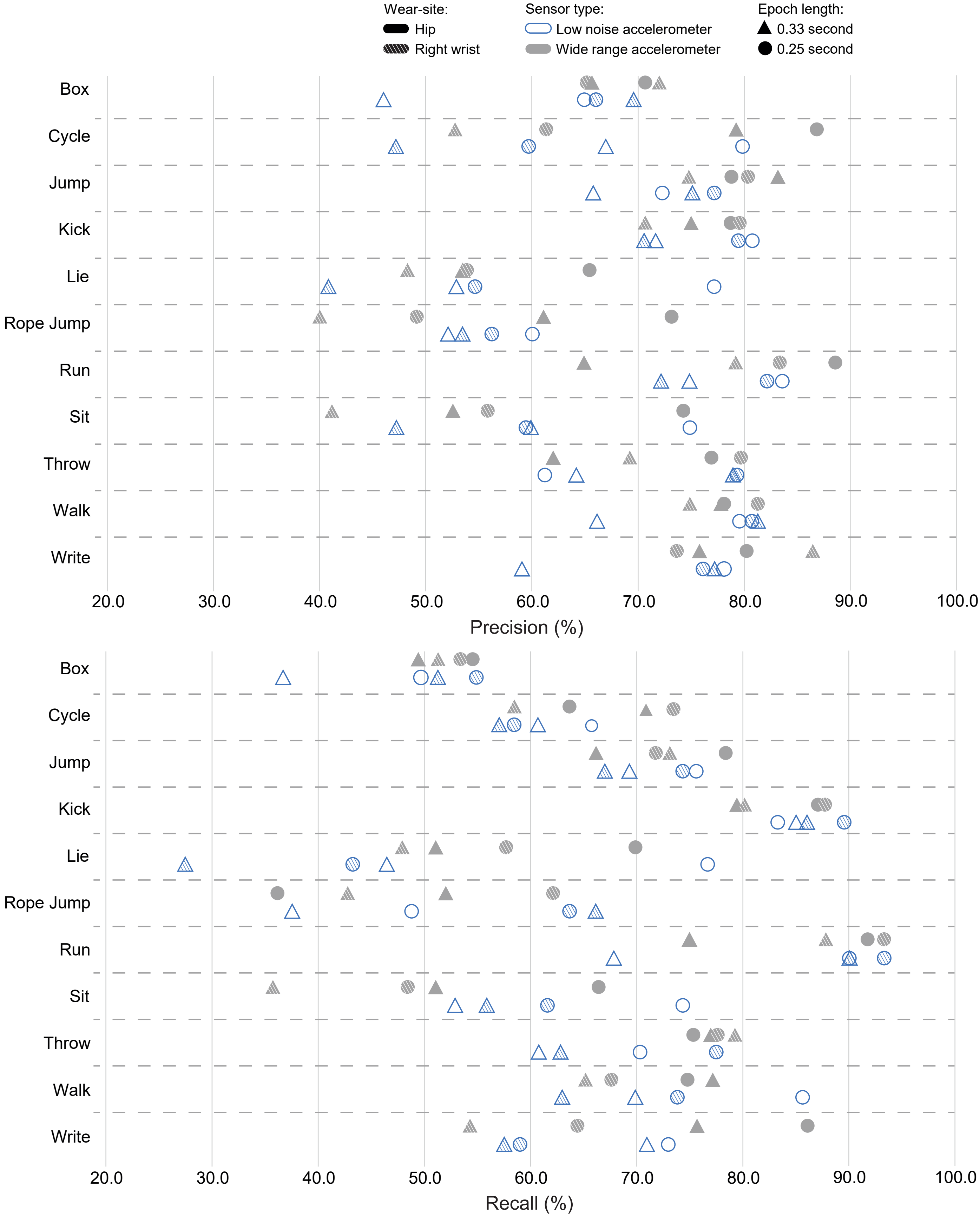


Figure 3.4: The precision (%) and recall (%) for different activities per sensor type, wear-site and epoch length.

Activity		Low noise accelerometer				Wide range accelerometer			
		0.33 s. epochs		0.25 s. epochs		0.33 s. epochs		0.25 s. epochs	
		Hip	Right wrist	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist
Sit or Lie	Prc.	91.5%	84.3%	90.9%	92.8%	88.4%	91.4%	91.8%	89.1%
	Rcl.	80.6%	90.6%	90.4%	84.8%	85.5%	85.0%	90.2%	86.5%
	Acc.	+2.2%	+3.0%	+1.1%	+2.3%	+2.4%	+2.9%	+1.6%	+2.3%
Jump or Rope jump	Prc.	69.2%	72.2%	77.1%	70.7%	80.0%	67.2%	86.4%	69.2%
	Rcl.	69.4%	68.2%	75.6%	71.5%	65.0%	67.5%	74.0%	69.2%
	Acc.	+1.0%	+0.2%	+0.9%	+0.0%	+0.3%	+0.3%	+0.8%	+0.0%
Box or Throw	Prc.	68.7%	81.1%	73.5%	80.7%	73.7%	75.9%	80.2%	79.6%
	Rcl.	59.1%	62.1%	70.3%	72.9%	73.2%	70.2%	70.1%	71.2%
	Acc.	+2.3%	+1.1%	+2.2%	+1.4%	+2.2%	+1.1%	+1.2%	+1.4%
Correction accuracy		+5.5%	+4.3%	+4.2%	+3.7%	+4.9%	+4.3%	+3.6%	+3.7%

Table 3.2: Implications of grouping comparable activities.

Number of classifications	Low noise accelerometer				Wide range accelerometer			
	0.33 s. epochs		0.25 s. epochs		0.33 s. epochs		0.25 s. epochs	
	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist
KNN	23688	24237	26054	25825	24349	24118	26094	26008
Thr.	21569	22780	25633	25488	23581	23147	25906	25332
Original	31811	31811	33605	33605	31811	31811	33605	33605
	Differences with original classifications							
KNN	8123	7574	7551	7780	7462	7693	7511	7597
Thr.	10242	9031	7972	8117	8230	8664	7699	8273

Table 3.3: The amount of classifications made without and with the addition of an unknown class via two different approaches.

3.4. Activity-Intensity identification

For some activities an additional discrimination can be made based on the different intensities at which an activity was executed. Various metabolic equivalents (METs) have been assigned to each activity as mentioned in table 2.1. The eleven activity classes were converted into sixteen classes: Box VPA, cycle MPA, cycle VPA, jump MPA, jump VPA, kick MPA, kick VPA, lie SB, rope jump VPA, run VPA, sit SB, throw LPA, throw MPA, walk LPA, walk MPA, write. The Hip's 0.25 second epoch data produced significantly better results than the right wrist or 0.33 second epoch measurements, with a mean accuracy for LN of 60.1% and WR of 63.2% ($p \leq 0.008$). Comparing the activity-intensity predictions of the 0.25 second epoch LN and WR hip accelerometer to the same configuration of activity-only predictions shows that the mean accuracy is over 15% lower when intensities are included in the network training and prediction ($p<0.001$).

3.5. Categorizing uncertain predictions as ”unknown”

On the 0.25 and 0.33 second epoch length data from the LN and WR accelerometers from the hip and right wrist it was studied how the classification precision per activity changes when doubtful predicitons would not be classified. A simple K-Nearest-Neighbours (KNN) machine learning algorithm was constructed to redefine some predictions as ”unknown” based on their maximal score, an additional output returned by MATLAB's classify-function. Additionally, a script was written using a threshold value for the maximal score of 0.775. This value was was experimentally obtained by changing it iteratively for a subset of measurements and looking at the resulting overall predictive accuracy and number of excluded classifications. This procedure thus classified every prediction with a score below the chosen threshold as ”unknown”. The influence of these two approaches on the precision of the initial classification is shown in figure 3.5. The number of remaining classifications was also recorded and this can be found for both approaches and all configuration in table 3.3. With exception of the 0.25 second epoch WR right wrist placement, the threshold approach resulted in significantly better precisions than when no predictions are excluded ($p<0.017$). The KNN algorithm only significantly increased the precisions with respect to the original values four out of eight times. For all hip data and the right wrist's LN 0.25 second epoch measurements, the threshold method returned significantly higher precisions than the KNN algorithm. The confusion matrices for all individual sensors and for the predictions with and without the addition of an unknown class can be found in Appendix C.

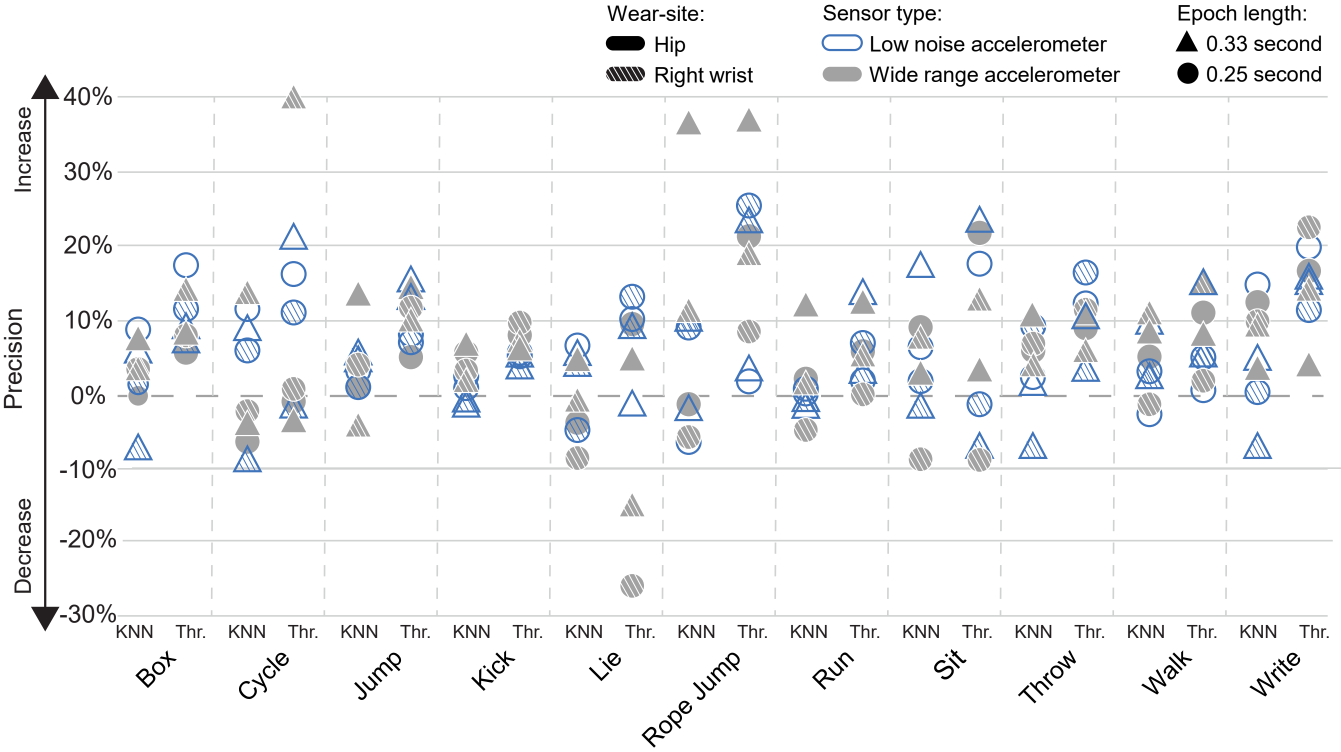


Figure 3.5: The changes in precision per activity with respect to the classification without defining an unknown class for the KNN and the Threshold (Thr.) algorithms.

3.6. Testing the algorithm on children's data

The data collected on children were used as validation dataset for a network trained on all eight adult datasets. The accuracies for each sensor type, placement and epoch length were significantly lower than the accuracies when validating with an adult data set on an adult data trained network (p <0.001), table 3.4. Also a network was trained on each of the two children and validated with the other child's dataset to which the network was naive. The highest mean accuracy belongs to the 0.25 second epoch data from the right wrist's WR accelerometer and was found to be 53.0%, significantly outperforming most other configurations. Only the right wrist LN 0.25 second epoch data (mean accuracy = 47.1%) and the hip LN 0.33 second epoch (mean accuracy = 48.1%) yielded no significantly lower results with respect to the best performing sensor (p = 0.219 and p = 0.450, respectively). The accuracies of the low noise 0.25 second epoch hip accelerometer from the children were compared against those of adults' predictions obtained when training the LSTM on the same measurements from one adult. The accuracies of the children's data are significantly lower with a mean of 40.9% (±20.6, 95% CI: 36.7 - 44.9) as opposed to 53.7% (±22.2, 95% CI:49.5 - 58.0) for the adults' data set (p <0.001). When only child data are used for network training and validation, the accuracies (overall average = 42.5%) are significantly lower (p =0.047) than when the adult trained network, which used more data for training, is used to classify the children's measurements (overall average = 44.5%).

Descriptives	Low noise accelerometer				Wide range accelerometer			
of	0.33 s. epochs		0.25 s. epochs		0.33 s. epochs		0.25 s. epochs	
accuracy	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist
Mean	54.1%	45.3%	33.0%	40.1%	54.7%	46.7%	53.1%	29.4%
Std.	24.8	16.4	24.6	17.8	25.3	17.6	23.7	22.9
95% CI	(49.3 -58.9)	(42.1 - 48.5)	(28.1 - 37.8)	(36.6 - 43.5)	(49.8 - 59.6)	(43.3 - 50.1)	(48.5 - 57.7)	(24.9 - 33.8)

Table 3.4: Mean, standard deviation and 95% CI for the accuracies when classifying data from children on a network trained on adults.

3.7. Effect of more training data on classification accuracy

Seven networks were trained on 0.25 second epoch low noise accelerometer data from the hip sensor, each with a different number of training data sets, ranging from one to seven. A linear regression analysis was performed on the mean accuracies that were obtained after classification of the remaining validation sets with each network. A correlation coefficient (R^2) of 0.9405 (p = 0.016) was found between the mean prediction accuracy and the number of training sets that was used for training the neural network. The plot with corresponding trendline is included in figure 3.6. The equation for the trendline is: $y = 0.0327x + 0.534$.

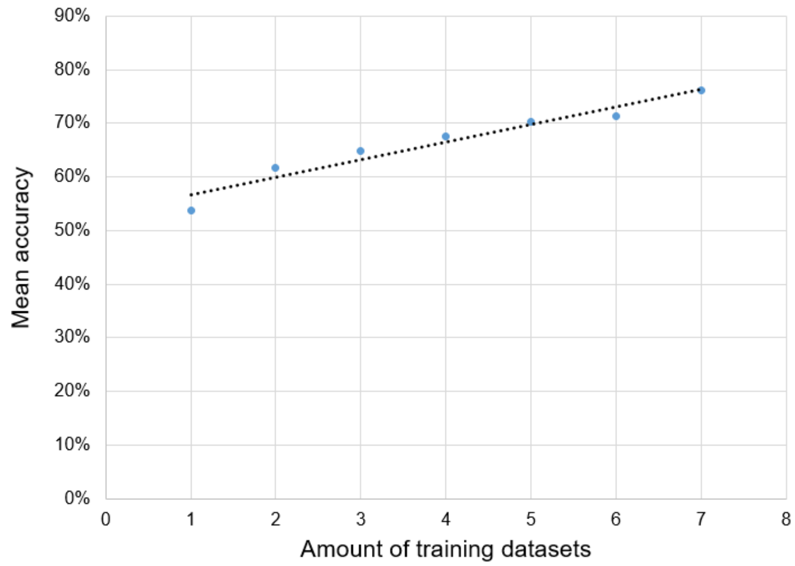


Figure 3.6: Influence of the number of training data on the prediction accuracies of the validation datasets.

3.8. Calculations for data transfer capacity and power demand

For the following calculations it was assumed that the device is programmed to only collect data when the child is moving, which is estimated to be at most six hours a day. Furthermore it is believed that two days of accelerometer data yield a sufficient amount of information for research purposes. Calculations were performed for uploading triaxial accelerometer data summarized in 0.25 second epochs, enriched with one HR measurement per minute to a smartphone. It was calculated that this would result in 518400 accelerometer data points and 720 HR data points. The accelerometer's values range between -4096 and +4096, which means 13 bits are required to store one data point. A value of 255 is the maximum to be stored in 8 bits, which is sufficient for HR measurements. The required data storage and transfer capacity is therefore *518400 * 13 + 720 * 8 = 6744960* bits, or *6744960 / 1024 = 6586.875* Kbits.

Three wireless methods have been investigated, namely Near Field Communication (NFC), Bluetooth Low Energy (BLE) and Light-Fidelity (Li-Fi). The maximum transfer speed of NFC is 424 Kbit/s, meaning that it would take approximately 15.5 seconds to transfer the specified amount of data [11]. NFC tags do not require power to operate. Through BLE the signal could be sent more continuously, however this requires energy. If data are transferred with a frequency of 4 Hz, which would take around 8 ms per second[1], assuming the micro controller requires 20 mA while transferring and 5 μA when on standby, the necessary current per second would be *0.008 * 20* mA + *0.992 * 5*μA = *0.165* mA. If two days worth of data would be sent at once, instead of continuously, the data could be transferred in 6.4 seconds because BLE has a speed of around 1.0 Mbit/s [24]. However, BLE would require around 4 extra seconds to establish the connection between the device and the wearable, resulting in an estimated transfer time of 10.4 seconds. Moreover, the current associated with two days of measurements would then be *10.4 * 20* mA = *208* mA. Li-Fi uses flashing LED lights to transmit a bar code-like signal to a smartphone camera [21]. One LED can transfer data at a speed of 1550 bit/s when held at a distance of maximally 5 cm from the smartphone camera [6]. It would thus take 4351.6 seconds to communicate all measurements from twelve hours to the smartphone when only using one LED. With five LEDs the transfer time would drop to 870.3 seconds and for 64 LEDs this value would be 68.0 seconds. Each LED operates on around 20 mA[2]. Table 3.5 shows the amount of seconds it would cost to transfer one hour

[1] stackoverflow.com/questions/36725225/how-can-i-estimate-the-power-consumption-of-a-ble-module

[2] www.allekabels.nl/led-diode/7369/1096839/led-led5gln.html?gclid=EAIaIQobChMI8py0x7jo6wIVmuF3Ch1orQrJEAQYASABEgJzPf
BwE

Technique	Transfer time (s for 1 hour of data)	Connection time (s)
NFC	1.30	-
BLE	0.54	4
Li-Fi: 1 LED	362.63	-
Li-Fi: 5 LEDs	72.53	-
Li-Fi: 64 LEDs	5.67	-

Table 3.5: Seconds of time needed to transmit data from the wearable to a smartphone per hour of collected data.

of measurements for each of the methods.

An expert in the field of electronics was consulted for the estimations regarding the electrical characteristics of the micro controller unit (MCU) and accelerometer. Performing measurements at 51 Hz with the estimation that it would take 0.5 ms to read the values causes the MCU and accelerometer to be in high power mode *51 * 0.0005= 2.55%* of the time. According to the ShimmerSensing datasheet, the wearable would ask 162 μA in high power mode, 10 μA in low power mode and 2.5 μA in power down mode [15]. Taking 6 hours a day where the wearable will be active, of which 2.5% in high power mode, gives a daily power consumption, excluding data transfer, of 461.7 mA. Larger batteries have higher capacities, so a trade-off can be made regarding the size of the wearable and battery life. For example, a cylindrical lithium ion 18650 battery with a size of 18*65mm and a capacity of 2600 mAh[3] could power the wearable for *2600 / 461.7 = 5.6* days. Another cylindrical lithium ion battery, the 32650, has dimensions of 32*70mm and 6000 mAh capacity[4], meaning it can power the device for 13.0 days. A different battery delivers not even three days worth of power, but is considerably smaller with a cuboidal size of 1.0*30*40mm[5]. For the BLE and Li-Fi technologies, the life expectancy of the batteries will be less than calculated in this paragraph since these transfer methods require additional power.

[3] www.alibaba.com/product-detail/Battery-Lithium-Rechargable-Battery-3-7v_62252618977.html?spm=a2700.galleryofferlist.0.0.6d473f37yoglox&s=p

[4] www.alibaba.com/product-detail/Battery-32650-Battery-6000mah-High-Capacity_60812240465.html?spm=a2700.galleryofferlist.0.0.6d473f37yoglox&s=p

[5] www.alibaba.com/product-detail/Battery-103040-3-7v-1200mah-Lipo_60764366650.html?spm=a2700.galleryofferlist.0.0.6d473f37yoglox&s=p

4 Discussion

Before the published results of this research were conceived, some exploratory analyses were performed to gain better understanding of the data. The implications of these investigations will briefly be discussed before reviewing the outcomes disclosed in the previous section. Preceding the decision to randomize the order of activities it was observed that consecutive activities were often confused. Moreover, the order of activities was alphabetical by default and boxing, alphabetically the first activity label, appeared to be recognized with the highest accuracy in all cases. When the order of activities was randomized while still forming a full sequence of eleven activities, accuracies dropped drastically from 85% to around 20%. Therefore it was concluded that the order-dependency characteristic of LSTM networks[1] caused the system to remember these structural regularities instead of the movement-specific features. Because of the relatively small amount of available data and the risk of overfitting, two alternatives were considered to work with this neural network property: Separating the data into individual activities and randomizing the arrangement of these activities or consecutively merging the activities and dividing this series over multiple shorter sequences while randomizing the order of activities within a short sequence. Children are known to move in short bursts rather than for example running a marathon. Since the aim of this research is to develop a wearable for children, it was decided that merging activities would thus be more realistic. The transition periods between individual activities will comprise a significant portion of the data captured in real life free-living children and an activity identification algorithm should thus be able to operate adequately when these transitions are present in the data. This is the approach used to generate all results reported in this thesis. For future research it is important to be aware of the pitfall of order-dependency, the activities must be randomized to prevent this type of algorithm from simply learning the order of events rather than movement specific features. Exact reproduction of sequences has no relevance for this thesis' intended application of activity recognition in the real world.

4.1. General remarks

Despite trimming the data from some activities, the data set is still unbalanced due to the high prevalence of kicking. When the activities combined in short sequences were randomly selected, some activities were picked to occur in two combinations, so these measurements are used twice among all data. This contributed to the high presence of kicking, since this activity occured in two short sequences. The other reason is that only long sequences were trimmed, but the full length of the shorter burst activities were included. In hind sight, a part of the kicking data could have been discarded to get a more equal distribution of data. Kicking does get recognized with moderate precision, however there are other activities with lower prevalences that are identified with lower recall than kicking. If the algorithm regularly classifies other movements as kicking, recall would decrease. Seen the values for recall in figure 3.4 this is not yet an issue with kicking.

4.2. Gyroscope, magnitude and multiple sensors

Using raw gyroscope data does not increase the performance of the proposed LSTM algorithm with respect to only using three dimensional accelerometer data. If the raw gyrometer data would be used to compute the actual angular velocities, it could potentially support the differentiation between sitting and lying. This could be done by establishing a short period of sensor-stationarity to reduce the expanding errors resulting from double integration [3]. Calculating the actual displacements could affect the classification accuracy of other activities as well, however as most activities were executed in an upright position the differentiation between activities should most likely be based on changes in angular velocity, which are also present in raw measurements. Employing the magnitude of the three accelerometer axes yields relevantly decreased accuracies, since the accuracy significantly drops between 4.0% and 19.0%. Adding the magnitude of the gyroscope data to this information continues to produce significantly inferior results. Employing two accelerometers to capture data on activities does not generate more accurate predictions than using measurements from one accelerometer

[1] `colah.github.io/posts/2015-08-Understanding-LSTMs/`

on either the right hip or the right wrist. This is an unexpected result that is probably due to the method used to join data from multiple sensors. It is likely there is a better approach to combine measurements from two sensors, but investigating this was outside the scope of this thesis. However, from these results it can be concluded that data from one sensor are sufficient for identifying activities. This is also preferable since less data have to be transferred. A choice thus has to be made between the hip and the right wrist placement.

4.3. Identification of activities

Of the four combinations of activities, only ”Box-Jump-Throw-Walk” generates significantly and relevantly lower accuracies than the other short sequences. Looking at all classifications done within this sequence, it becomes obvious that the activities that are most often confused are throwing, classified as kicking (4307 times) or boxing (1898 times), and boxing, predicted as throwing (2740 times). Also, this is the only combination that shares three activities with other sequences (Box, Jump and Walk) while throwing only appears in this configuration. Both the hip and right wrist placement confuses throwing mostly with kicking, however it seems that the right wrist placement is slightly better at distinguishing the two, see Appendix C. It makes sense that these two are often confused, as the movement pattern apart from the throw or kick itself is fairly similar. The participants had to jog towards the ball after throwing or kicking, return to their initial position and have another shot. The fact that the jogging pattern is interrupted by throwing or kicking a ball makes it probably distinctive from walking or running, but the periodicity of both ball activities is analogous. For the 0.25 second epochs from a wide range accelerometer, precision and recall for throwing are between 75% and 80%, which is compared to other activities a decent score, see figure 3.4. Despite mixing throwing up with kicking sometimes, the algorithm is still reasonably capable of recognizing the individual activity. The fact that boxing and throwing are mixed up can be due to the involvement of the upper body. Even though the movement itself is quite different, because throwing exists of shorter bursts of the upper body alternated with a jogging activity whereas boxing is a continuous motion of the upper body, it is possible that the algorithm acknowledges the relative absence of lower body movements at certain intervals.

Precision and recall have been used to attribute performance of the algorithm for the individual activities. The bigger picture is that the algorithm can be used for a wearable that motivates children to move more often by giving them feedback which movements they made. For this application it is therefore more important to ensure that every time a movement is classified as running the child was in fact running and it is less relevant to identify every 10 seconds a child spends running. This implies that for this algorithm high precision is favored over high recall. When studying the results previous to adding an unknown class, it is visible that even though the WR accelerometer might not perform significantly better, it does recognize certain activities better than a LN accelerometer. The difference between these types of accelerometers is that the former allows for measurements of short events with high frequencies, thus explosive burst activities, while LN accelerometers have a better resolution among lower frequencies, the more sedentary activities, and generally consume less power[2]. Albeit not significant, for jumping, cycling and running the WR accelerometer seems to classify with higher precision than the LN accelerometer. Lying, on the other hand, is most precisely recognized by the 0.25 second LN hip accelerometer. This is thus in accordance with the theory explained before. It would be interesting to investigate the performance of both sensor types for an increased ratio of jumping, cycling and running activities. Children are more often sedentary than performing high frequent burst activities, however the wearable is intended to be in power-down mode when a child is inactive. The ratio of high frequent burst activities to more sedentary activities that are recorded by the wearable is thus unknown. If there are relatively a lot of high frequent activities among the measurements, the WR could theoretically significantly outperform a LN accelerometer. Nonetheless, this cannot be concluded from the outcome of this thesis. From these results it appears that it does not matter whether a WR or LN accelerometer is used.

From figure 3.4 it is evident that lying, sitting and ropejumping are the hardest activities to identify when looking at the precision score. However, relevant improvements are made among these classes if clustered with another comparable activity. The precision of both sedentary activities reaches around 90% and using a WR accelerometer on the hip both jumping activities get recognized with precisions ≥80.0%, see table 3.2. Jumping and rope jumping are less often confused for the right wrist than for the hip placement, which is probably due to the substantial wrist movements involved in spinning the rope. An important question to answer is thus how much we need to know, since there is a trade-off in distinctiveness and precision, and which clusters help increase the performance of the algorithm. If it is sufficient to very generally distinguish various activities,

[2] www.analog.com/en/analog-dialogue/articles/choosing-the-most-suitable-accelerometer-for-your-application-par html#

such as being sedentary, walking, running, jumping activities, activities with a ball, etc. the precision for recognizing such comprehensive classes will increase drastically. It will however remain unknown if a child was jumping or rope jumping. Because the variety of movements made by children is of interest, the relevance of identifying activities very specifically is arguably less than having a high precision when categorizing activities in global clusters. Looking more closely at the specific activities, it can be seen that the right wrist placement makes for poorer estimations of cycling, rope jumping and sitting but slightly outperforms the hip placement when it comes to boxing, throwing and walking. Striking is that the precision for writing is comparable for both wear-sites, but the recall for writing measured on the right wrist is evidently lower. All participants were right-handed which could have caused the right wrist accelerations during writing to resemble more energetic movements involving the right hand, as it is often confused with boxing, see Appendix C.

With the activities grouped as described in table 3.2, the precision for recognition of every activity becomes higher than 80% for multiple combinations of sensor type, placement and epoch length for all of the (grouped) activities. The precision, accuracy and recall rise even further if activities with a low prediction score are discarded as unknown, see figure 3.5. The 0.33 second epoch length data in particular seems to improve for both ways of implementing an unknown class, more than the 0.25 second epoch data. Potentially there is greater uncertainty in predictions among the larger epoch length. This claim is supported because 0.33 second epoch data are relatively to the original amount of data more frequently classified as ”unknown”, as reported in table 3.3. Furthermore, it can be seen that lying responds the worst to this procedure. From the confusion matrices in Appendix C it can be derived that this is due to the persistent mix-ups with sitting. The algorithm is repeatedly mistakenly convinced that lying is sitting and vice versa. This is not simply due to hesitation, which would cause lower prediction scores and thus increase precision when a threshold is implemented, but because the neural network cannot differentiate sufficiently between the two. Also an obvious improvement is seen in the precision of classifying rope jumping while being the most difficult activity to identify according to the original data, if lying and sitting are considered a grouped activity. This indicates that the algorithm is generally unsure when making the prediction of rope jumping rather than being convinced it was something else. Rope jumping thus occasionally does not show amply pronounced features to be classified with satisfactory confidence.

Despite the fact that omitting inconclusive predictions allows for a significantly better assessment of which activities indeed were executed, there is now also a lot of data that is not taken into account. Approximately 23.4% and 26.1% of the predictions that were actually correct have on average been classified unknown by respectively the machine learning and threshold approaches 3.3. Nevertheless, abiding by the same reasoning why high precision is preferred over high recall, making more precise and accurate predictions is favored over more negligently classifying every ten seconds of the day as an activity. It is therefore recommended to perform a more thorough investigation into an algorithm that can exclude the most ambiguous predictions without discarding too much of the correctly classified data. The algorithm calculating the time that a child was active should accordingly be tweaked to ensure that the loss of predictions does not result in a severe underestimation of the amount of PA.

4.4. Supplementary categorization of intensities

Identifying the activity and intensity based on the triaxial accelerometer data from the hip and right wrist provided substantially lower accuracies than only classifying activities. Another method for estimating intensity levels is by measuring the heart rate (HR). Disadvantages of this feature are it’s subjection to emotional distress and that it exhibits a plateau effect at very high intensities [1]. The former objection can be reversed by knowing the activity that was executed at the time of measuring the HR. When the child in question is sitting while simultaneously displaying a high HR, it can be concluded that he or she might have been watching a thrilling movie. The matter of the plateau effect cannot be abated in the same way, however as this phenomenon only occurs at very high intensities it is not that relevant. That is to say, at high HR it is already certain that a child is performing VPA and the relevance of the measurements is not affected by missing these slight deviations within this level of intensity. Also, especially when omitting some classifications based on a low classification score, the HR data can provide more information on the quantity of physical activity performed by a child. HR measurements can nowadays easily be obtained from a wrist location, putting the right wrist placement in a favorable position over the hip placement [23],[18].

4.5. Children’s data and quantity of data

Using an algorithm trained on measurements from adults, it can be determined whether a child is sedentary with sufficient precision. However, the remaining activities are not appropriately being recognized, meaning that for the classification of activities performed by children, a network should be used that was trained on data generated by children. Because age or physical development could apparently be pertinent factors for variance in accelerometry measurements, it is additionally recommended to collect data over the range of ages that are of interest for the development of this wearable. Potentially the classifications would become more accurate if the age or body height of the wearer is known and the smartphone application would classify measurements by employing a neural network trained on information from children with similar characteristics. Even though there exists a significant difference between the accuracies found when applying children's data to an adult trained network and the outcomes returned by training and testing a network on children data, as the difference in means is only 2.0%, it is not a relevant distinction. It is essential information that the adult network was trained on eight data sets, while the child network was trained on one. More data on children will make for a better performing network, there will be further elaborated on this topic in the following paragraph.

Training the 0.25 second epoch data on the hip's LN accelerometer with three training sets resulted in a validation accuracy of 64.8%, figure 3.6. The right and left ankle had no significant difference in performance, but showed higher accuracies for three training sets. Upon further analysis, it was found that the ankles did not achieve results that were significantly different from the hip when trained on three data sets ($p \leq 0.745$). In future research it should be considered to examine one ankle alongside the hip and right wrist placements, with more data for each of the wear-sites. By means of the high correlation coefficient of 0.94 associated with figure 3.6 it can furthermore be assumed that more data in general would contribute to improved accuracies.

4.6. Comparison with existing research

One study was discovered that applied a pre-trained Convolutional Neural network (CNN) to both an existing and a real-world data set [5]. The reason for this was that the researchers wanted to evaluate the pre-trained CNN for realistic circumstances, because the existing data set was thought to under-represent important challenges present in real-world data. They found that the existing data set was classified with an accuracy of 91.98% while the realistic measurements only returned an accuracy of 67.51%. Acknowledging that the LSTM network proposed in this thesis was trained and validated on "real-world" data, the activity-classification accuracies achieved by using summarized triaxial accelerometer data from the hip and right wrist are promising compared to the previously mentioned study.

4.7. Data transfer and power supply implications

It is desired to save two days worth of data before deleting the earliest recordings in order to limit the potential bias resulting from children who only connect on days that they were relatively active. If the smartphone is connected to the smartphone with intervals larger than two days, it will take more time to complete the transfer of data than when data is more often transmitted. Li-Fi technology is infeasible for the intended application because it will take too long to transfer the predicted amount of data. For larger amounts of data, NFC takes relatively longer to perform the data transfer than BLE. It is however expected that NFC will be possible at higher rates in the future [9]. When looking at the time required for the transmission of one hour of data and taking into account the additional time needed to establish a connection, see table 3.5, NFC is found to be faster than BLE if the wearable transfers less than 5.26 hours data. There are thus advantages regarding the transfer times for both techniques, but BLE requires extra power to operate in addition to the considerable amount of power estimated for the MCU and accelerometer. NFC is thus favored over BLE.

Even when the wearable is equipped with a passive data transfer component, the wearable would last merely a couple of days with a reasonably sized battery. NFC technology actually allows for energy harvesting. In this procedure, the power of an active electronic device is used to power the battery of a device that carries a NFC tag. This is not a fast procedure for charging a battery, but if the wearable would be in close proximity of a smartphone, for example at nights, it would be possible to recharge it [12]. This is preferred to replacing the battery every few days or making the wearable disposable, due to user comfort and sustainability constraints respectively. An additional benefit of NFC technology is that it is inexpensive, at cost of $0.05 to %0.10 USD[3]. One more advantage is the associated safety, since data can only be transferred over a range of a couple of centimeters. The disadvantage of a lengthy smartphone-communication time could be partially solved by encouraging children to update more regularly or by finding a method that demands less data. Using 0.33 second epoch lengths reduces the time needed for data transfer by almost 25%.

[3] https://medium.com/blue-bite/your-questions-answered-7-common-misconceptions-about-nfc-9c580fd66635

5 Conclusion

With the proposed framework it is possible to identify an activity for every 10 seconds of summarized triaxial accelerometer data. Based on this research alone, it appears that the most promising way of assessing both intensity levels and the specific activities performed by free-living adults is through 0.25 second epochs of right wrist triaxial accelerometer measurements enriched with counts from a HR monitor. This can be done by creating a scalogram of the accelerometer data, computing GoogleNet activations of this frequency plot and using this as input for a neural network consisting of a bidirectional Long-Short Term Memory layer with 500 hidden units. Results indicate that the algorithm can distinguish the eleven activities that are part of this thesis, but the accuracies obtained by only taking the previously mentioned steps are not sufficient. The accuracies of LN and WR 0.25 second epoch accelerometer data of the right wrist are respectively 74.8% and 74.4%. The associated precisions per activity range from 54.5% to 82.1% and from 49.1% to 83.3%. The precision for recognizing the various activities should be minimally 90% for each activity. If the prediction is often incorrect, children might get demotivated and not use the application as much. High precisions are thus necessary to contribute to motivating children to move more frequently. If the information is used for research or policy-making purposes, the classifications should as well be of sufficient quality in order to make valid assumptions. Achieving the desired precisions will be difficult, but in the researcher's opinion it is not inconceivable. It is therefore strongly recommended to carry out future data collection on children over the full range of relevant ages and to accumulate a lot more data than what was used in this thesis. To improve the performance of the algorithm, two techniques should be applied: Removing ambiguous classifications and clustering activities for which the distinction is not relevant.

Discarding classifications with a low predictive certainty is strongly recommended, although additional research is needed to develop a suitable procedure for this. When classifications of right wrist LN measurements with a score below 0.775 are eliminated, an accuracy of up to 84.0% and precisions per activity between 57.8% and 91.5% can be achieved. These results will further increase when specific activities are clustered, for example sitting and lying down, as both are sedentary. Clustering sitting and lying yields a total accuracy of 85.4% and the lowest precision would change to 70.1%. Activities should thus be clustered into relevant groups and classifications with a low certainty should be excluded. Research should be done to determine which activities are relevant to cluster. A promising cluster is lying down and sitting, since they are often confused and both activities are sedentary and therefore the distinction is not that interesting for the interaction between the child and the smartphone application or for the policy-making. Moreover, kicking and throwing are often confused, clustering these would additionally improve the algorithm's performance. However, it is debatable whether the latter would form an appropriate cluster since determining physical relevance for distinguishing specific activities was outside the scope of this thesis. Groups of activities should be carefully chosen in a way that the beneficial effects of the movements within a cluster are comparable and that these activities are mixed up by the software.

Not only development of the algorithm will be challenging, also the hardware has to be considered. A battery life of several days is a drawback of the current concept, since children will have to think about charging the wearable quite often. If they forget to charge it the wearable will shut down and no measurements are recorded. This could cause children to lose their motivation to move more frequently and variably. When selecting the electronics for the end product, there should be focused on a maximal reduction of power consumption. For now, it is advised to use NFC technology as means of transferring the data from the wearable to a smartphone application. This is because an NFC tag does not require power and it allows for energy harvesting, thus enabling wireless charging of the wearable's battery.

Bibliography

[1] Juul Achten and Asker E Jeukendrup. Heart rate monitoring. *Sports medicine*, 33(7):517–538, 2003.

[2] Barbara E Ainsworth, William L Haskell, Arthur S Leon, David R Jacobs Jr, Henry J Montoye, James F Sallis, and Ralph S Paffenbarger Jr. Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and science in sports and exercise*, 25(1):71–80, 1993.

[3] Michael Angermann, Patrick Robertson, Thomas Kemptner, and Mohammed Khider. A high precision reference data set for pedestrian navigation using foot-mounted inertial sensors. In *2010 International Conference on Indoor Positioning and Indoor Navigation*, pages 1–6. IEEE, 2010.

[4] Stuart Biddle. Psychological benefits of excercise and physical activity. *Revista de psicología del deporte*, 2(2), 1993.

[5] Federico Cruciani, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. Feature learning for human activity recognition using convolutional neural networks. *CCF Transactions on Pervasive Computing and Interaction*, 2(1):18–32, 2020.

[6] Alexis Duque, Razvan Stanica, Herve Rivano, and Adrien Desportes. Unleashing the power of led-to-camera communications for iot devices. In *Proceedings of the 3rd Workshop on Visible Light Communication Systems*, pages 55–60, 2016.

[7] Jeff Galloway. *Fit Kids, Smarter Kids*. Meyer & Meyer Verlag, 2006.

[8] Regina Guthold, Gretchen A Stevens, Leanne M Riley, and Fiona C Bull. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1· 6 million participants. *The Lancet Child & Adolescent Health*, 2019.

[9] Garima Jain and Sanjeet Dahiya. Nfc?: Advantages, limits and future scope. *vol*, 4:1–12, 2015.

[10] Ian Janssen and Allana G LeBlanc. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International journal of behavioral nutrition and physical activity*, 7(1): 40, 2010.

[11] Antti Lahtela, Marko Hassinen, and Virpi Jylha. Rfid and nfc in healthcare: Safety of hospitals medication care. In *2008 Second International Conference on Pervasive Computing Technologies for Healthcare*, pages 241–244. IEEE, 2008.

[12] Antonio Lazaro, Ramon Villarino, and David Girbau. A survey of nfc sensors based on energy harvesting for iot applications. *Sensors*, 18(11):3746, 2018.

[13] Kate Lyden, Sarah Kozey Keadle, John Staudenmayer, Patty Freedson, and Sofiya Alhassan. Energy cost of common activities in children and adolescents. *Journal of Physical Activity and Health*, 10(1): 62–69, 2013.

[14] Maedeh Mansoubi, Natalie Pearson, Stacy A Clemes, Stuart JH Biddle, Danielle H Bodicoat, Keith Tolfrey, Charlotte L Edwardson, and Thomas Yates. Energy expenditure during common sitting and standing tasks: examining the 1.5 met definition of sedentary behaviour. *BMC public health*, 15(1):516, 2015.

[15] ST Microelectronics. Lsm303agr—ultra-compact high-performance ecompass module: Ultra-low-power 3d accelerometer and 3d magnetometer, 2018.

[16] Geraldine Naughton, David Greene, Daniel Courteix, and Adam Baxter-Jones. Resilient, responsive, and healthy developing bones: The good news about exercise and bone in children and youth, 2017.

[17] World Health Organization et al. *Global recommendations on physical activity for health*. World Health Organization, 2010.

[18] Jakub Parak, Maria Uuskoski, Jan Machek, and Ilkka Korhonen. Estimating heart rate, energy expenditure, and physical performance with a wrist photoplethysmographic device during running. *JMIR mHealth and uHealth*, 5(7):e97, 2017.

[19] Lorenzo Scalise and Gloria Cosoli. Wearables for health and fitness: Measurement characteristics and accuracy. In *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2018.

[20] ShimmerSensing. *IMU User Manual Rev. 1.4*. Realtime Technologies Ltd, 2016.

[21] S Shreyas, G Shreyas, and J Pattar Praveen. Implementation of li-fi for data transfer in android devices. *International Journal of Scientific Research and Review*, pages 5–10, 2019.

[22] MS Sothern, M Loftin, RM Suskind, JN Udall, and U Blecker. The health benefits of physical activity in children and adolescents: implications for chronic disease prevention. *European journal of pediatrics*, 158(4):271–274, 1999.

[23] Sarah E Stahl, Hyun-Sung An, Danae M Dinkel, John M Noble, and Jung-Min Lee. How accurate are the wrist-based heart rate monitors during walking and running activities? are they accurate enough? *BMJ open sport & exercise medicine*, 2(1):e000106, 2016.

[24] Rohan Tabish, Adel Ben Mnaouer, Farid Touati, and Abdulaziz M Ghaleb. A comparative analysis of ble and 6lowpan for u-healthcare applications. In *2013 7th IEEE GCC Conference and Exhibition (GCC)*, pages 286–291. IEEE, 2013.

[25] Salvatore Tedesco, Marco Sica, Andrea Ancillao, Suzanne Timmons, John Barton, and Brendan O'Flynn. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. *PloS one*, 14(5), 2019.

[26] Victor van der Meer. Bewegen is goed, meer bewegen is beter. *Huisarts en wetenschap*, 60(12):618–618, 2017.

[27] Susann Weihrauch-Blüher, Peter Schwarz, and Jan-Henning Klusmann. Childhood obesity: increased risk for cardiometabolic disease and cancer in adulthood. *Metabolism*, 92:147–152, 2019.

Appendices

A
Literature review

How to accurately measure quantity and variety of physical activity in free-living children: a systematic review

K. C. Rijnders[a]

[a] *Student Biomedical Engineering, track Musculoskeletal Biomechanics & Sports Engineering, Delft University of Technology, The Netherlands*

Abstract

Objective
Regular physical activity (PA) is associated with improved mental and physical well-being in children. It is however approximated that 81% of youth world wide does not achieve the desired amount of activity. Therefore, the ambition exists to develop a device that intrinsically encourages youth to move more frequently. Because children additionally need to vary the movements they make, this gadget should register how often children move at which intensities as well as give information on the diverse movements that are executed by the wearer. For that reason, this review will give an oversight of the accuracies of different available sensor-types and placements on the body of children that have been used for PA recording in children.

Methods
The database Web of Science was explored for documents on measuring quantity and/or variety of PA using wearable devices in children aged 6 to 12 years. A pre-set search string was used to retrieve relevant literature. Articles were included by means of predetermined exclusion criteria. In the analysis, study objectives, protocols and outcomes were reviewed, without a statistical analysis.

Results
All documents that turned up were systematically reviewed for eligibility. 34 studies were included after applying the in- and exclusion criteria on all 295 documents that resulted from the search string. Some research reported on both investigated characteristics of PA, therefore 30 of the included studies discussed quantity and only 8 reported on variety. From the included literature it was concluded that a triaxial accelerometer (AM) or heart rate (HR) monitor assesses quantity with the highest accuracy in the case of free-living children. Additionally, triaxial AMs allow for estimation of motor skill development, ground reaction force (GRF) and combined with a machine learning algorithm even enable recognition of different movements. The placements of the device on the body that yield the highest accuracy are the hip and the ankle.

Discussion
From the included documents it became evident that there is often a trade-off between accuracy and cost. Triaxial AMs are the most accurate option for assessing quantity and different aspects of variety of PA in children. Because triaxial AMs are expensive, solutions requiring cheaper technology combined with algorithms facilitating extraction of information with an accuracy equal to triaxial AM measurements should be explored. It is furthermore recommended to look beyond the devices reviewed in this study and explore more innovative ways of determining PA intensities and variety of movements made by children. Waist- and ankle-worn sensors outperformed wrist worn devices based on accuracy, but when taking convenience for the user into account the wrist placement seems more appropriate.

Keywords: physical activity, children, measurement device, accelerometer, IMU, pedometer, heart rate monitor

Abbreviations	
AEE	Activity Induced Energy Expenditure
AM	Accelerometer
EE	Energy Expenditure
EMG	Electromyography
GRF	Ground Reaction Force
HR	Heart Rate
IMU	Inertial Measurement Unit
LPA	Light Physical Activity
MET	Metabolic Equivalent
MPA	Moderate Physical Activity
MVPA	Moderate to Vigorous Physical Activity
PA	Physical Activity
PM	Pedometer
SB	Sedentary Behavior
TDEE	Total Daily Energy Expenditure
TEE	Total Energy Expenditure
VPA	Vigorous Physical Activity

1 Introduction

The intention of this review is to contribute to the development of an easy-to-use, small sized and affordable wearable that will motivate children to be more active and at the same time can be used to track movement patterns in children. Physical activity (PA) is associated with an improved confidence and a reduction in anxiety, depression and stress [1]. Especially in children the psychosocial benefits of PA are evident [2]. Even at the age of 5 exposure to exercise stimulates bone accrual [3]. In addition to stimulating and maintaining a healthy body composition, the combination of moving frequently and healthy eating habits improves the functions of the heart and vessels, metabolism, bones and immune system [4]. In addition to having individual consequences, overweight and obesity also pressure finances in the medical sector [5]. Thus, moving regularly has a positive influence on both the mental and physical health of youth and their adulthood.

It is recommended for children to move at least 60 minutes a day to remain healthy [6]. More specifically, children between 5 and 17 years old should engage in at least one hour of mostly aerobic Moderate to Vigorous-intensity Physical Activity (MVPA) per day. Additionally, for a minimum of three times per week bone and muscle strength, among others, should be stimulated by performing Vigorous-intensity Physical Activity (VPA). However, it is estimated that worldwide 81% of children aged between 11 and 17 years do not reach sufficient levels of PA to comply with these recommendations [7]. Therefore, it is important to stimulate children to move more frequently and/or more intensely. Besides the amount of time that children spend being active, it is also important that they do not always execute the exact same movement. To enable a favorable muscle and bone development over their entire body, youth should engage all different body parts in PA by alternating the movements they make. This review aims to give insight into the different possible approaches that have been used by researchers to study two important characteristics of PA, which are defined as quantity and variety of the activities performed by youth.

Quantity is an indication of how often and how intensely people move. This can be calculated using minutes of being active and determining different intensity levels of PA. Variety is a more difficult concept to define and measure. To obtain data on variety, different activities should be distinguished from one another. So, jumping, cycling and running for example should be recognized and characterized as different categories of movement. This is because during these activities different bones and muscles are loaded. Another aspect of these movement is how well they are executed by the child performing a certain movement. The ability to correctly perform a motion depends on the motor skill competence of the individual child. Which muscles and how these are used to result in a certain movement can differ between kids of different ages for use of force and motions get more gradual and precise when a child gets older[1]. Furthermore, it is of importance how beneficial a certain activity is for bone and muscle development, or in other words: physical health. For example, bone growth is stimulated by GRF (Ground Reaction Force) and peak loading of the bone [8]. These loads can thus be measured to get an indication of how beneficial the PA is for the skeleton.

Sedentary Behavior (SB), Light PA (LPA), Moderate PA (MPA), MVPA or VPA are most commonly used to describe the intensity of PA. PA intensity levels can be measured subjectively, by means of surveys or diaries, and objectively, using motion detecting devices. Self-reported PA has been proven to have low reliability and validity in children multiple times [9], [10]. Also parents have been shown to over- and under-report single activities and therefore the existing parent questionnaires yield inaccurate results [11]. It is therefore preferred to use objective methods to gain information on the quantity and variety of activity in children, which is why this review is primarily concerned with ways to objectively and accurately assess PA characteristics in children. In addition to the different types of available devices, there also is a choice in location of the device on the body

Different intensities can be determined by measuring the Energy Expenditure (EE), which is measured by breathing through a gas analyser or via examining urine samples [12]. EE is a result from PA and from EE it can be observed how many calories are being

[1] https://www.stanfordchildrens.org/en/topic/default?id=the-growing-child-school-age-6-to-12-years-90-P02278

burned by the body during which intensities of exercise. PA only forms a part of the daily EE because the main contributor to EE is the resting EE, among which the thermal regulation of the body [12]. EE due to PA is the variable component of the Total EE (TEE). The rate of EE is called Metabolic Equivalent (MET). Since it is not feasible to equip a large population of children with (portable) breath analysers during free living conditions, measurements of EE and METs can solely be used to validate other measurement devices, such as accelerometers (AM), pedometers (PM), heart rate (HR) monitors and the placement of these devices on the body. To relate activity recorded by these devices to PA intensities, cut-points for intensity levels are set based upon METs and EE.

When studying the variety and execution of movements made by children, making observations is the most common approach. For a systematic way of conducting and reporting on these observations when made by humans, multiple surveys have been developed that focus on different aspects of childrens' PA. For example, there are questionnaires created for evaluating the motor skill competences of children [13] and inquiries on which movements are often made by children [14]. These examples require an other individual to watch and report on children's PA, meaning it is both a subjective and an inconvenient procedure for a large-scale investigation. An objective approach for recording the movements made by children is via optical motion capture technologies. This method uses camera recordings and special software to gain information on the pose, the combination of position and orientation, of different body parts. Additional markers on the subject's body might be necessary, however there are markerless systems available as well[2]. Despite the fact that markers are no longer a prerequisite, an optical system is not feasible for analyses of many children due to the required set-up. Therefore, it would be valuable to look into options of tracking motions using non-optical systems. Some of these non-optical techniques require an electromagnetic field, which is also impractical for extensive research. Measuring inertial data thus poses the best solution for collecting knowledge on this characteristic.

Especially from the age of six, children are known to get more involved in organized movements activities, like team sports or dancing, and become less interested in creative ones [15]. Since particularly spontaneous playing and exercise should be encouraged to get children to move regularly, it is evident to stimulate six-year-olds and over to move more outside their scheduled sports routines. With the current state of technology it is possible to accurately measure characteristics of PA in children. Even though the first motion-capture technology was invented during World War I[3] and Inertial Measurement Unit (IMU) technology started to evolve in the 1930s [16], only recently technology has developed sufficiently to allow for small sized and inexpensive measurement options [17].

Sensor type and placement on the body are substantial factors to consider when measuring characteristics of movements made by children. Previously conducted (systematic) reviews have focused on adults [18] or looked only at the use of an AM compared against measuring EE [19]. In this systematic review, the different approaches for measuring the quality and variety of PA in children between the ages of 6 and 12 will be summarized. Additionally, the accuracy of the different methods will be reported in order to select the most reliable and valid option(s) to investigate children's movements.

2 Methods

The main inclusion criterion was defined as follows: The objective of the article should be to give information on the quality of the measurements performed on children by one or multiple types of wearable activity monitors. Based upon this criterion, a few sub inclusion criteria were formulated:

- The research should address where on the body a certain device was placed
- The research should address the type of measurement device that was used
- The movements made by the participants should be natural, in a sense that the motions could also occur in free playing conditions
- Participants should be aged between 6 and 12 years old

Using these criteria, the database Web Of Science was searched on the 19th of November 2019, using the search term that can be found in appendix A. The search settings were set to find only documents in the English language and included documents from 1993 and beyond, because that was when sensors were deemed reliable enough to distinguish different body postures [20]. Five predetermined key references ([21], [22], [23], [24], [25]) and additionally applying multiple broader search strings were strategies to investigate whether all relevant studies were found. Each of the predetermined reference documents turned up when using the above mentioned search term. Applying broader search strategies did not result in more relevant documents.

[2] https://www.qualisys.com/applications/human-biomechanics/markerless-motion-capture/?gclid=Cj0KCQiAsbrxBRDpARIsAAnnz_PZk9tn-cNMWUyylwmuIcDsU66npq-CXbN9_1oleIQz7mcyGGz3CksaAtkxEALw_wcB
[3] https://engt.co/2JUozfA

The studies were assessed for eligibility by means of the subsequent exclusion criteria:

- No full text available using TU Delft resources.
- Not available in English language
- Comments, letters, reviews or case reports
- Animal studies
- Measurements were not performed on children between the age of six and twelve. In case of research with an age span predominantly inside the range and an average age within the limits of 6.0 to 12.0, the research should not be excluded based on age.
- The research does not address where on the body the movement was recorded.
- The research does not specify the type of sensor that was used.
- No wearable motion sensors were used, e.g. when only observations or questionnaires were conducted.
- All participants of the research were part of a specific patient population, e.g. obese or Cerebral Palsy.
- The participants were asked to perform a movement that could be seen as 'unnatural', meaning the movement usually does not occur in freely playing children, e.g. weight lifiting, or measurements were performed in an unnatural location, e.g. at high altitude.
- Sensors were used to measure something else than physical activity, e.g. sleep or respiratory movements.
- Results do not include measures of the quality of the data as delivered by the device.

Initially, the irrelevant document types were excluded from the list of results. Of all other records the titles and abstracts were screened for eligibility in accordance with the exclusion criteria. Thereafter, the full text of the remaining papers was analysed. During this phase, documents meeting any of the exclusion criteria were discarded, while from the eligible documents relevant data for this study were obtained. Data that were deemed relevant consisted of: title, author, year of publication, number of participants, mean age and age-span of the participants, gender of participants, type of sensor(s), location of sensor(s) on body, wear time, research settings, movements made, what the wearable was compared against and the results on quality of the measurement method.

Comparisons between different studies were made by looking at the measurement device that was used, the participant's characteristics, where the sensor was situated, during what type of activity recordings were made, device settings, used cut points and what the device was compared against to gain a better interpretation of the (differences in) results. Since the aim of this study is to provide an overview of existing literature, no statistical analyses were performed.

3 Results

The search resulted in a total of 295 documents. Irrelevant document types were excluded from this review. Of the remaining 211 documents, titles and abstracts were screened for eligibility against the exclusion criteria. The full text of all remaining papers was analysed. During this phase, relevant data for this study was obtained while simultaneously researches meeting any of the exclusion criteria were discarded. A flowchart on this selection process can be found in figure 1.

All results are summarized in table 1. Due to very dissimilar working protocols such as sensor types, placements, executed activities and outcome measures, most studies cannot be directly compared to each other. It was therefore decided it would be impractical to make a quantitative analysis. In this qualitative systematic review, multiple studies containing identical components will be grouped together in order to get a meaningful verdict on how quantity and variety of children's PA can be assessed most accurately.

Based on the information summarized in table 1 it is obvious that most documents focused on quantity of PA and relatively few on variety in movements. From the documents that did describe methods for determining the latter, these approaches often also allowed for derivation of PA intensity levels. For solely measuring quantities it is unnecessary to know the movement that was executed. However, for gaining knowledge on variety in movements a possible method is to determine which movement was made. If there is a desire to conclude on motor skill development skills with certainty, it is even essential to know the exact movement that was performed. Because information on the gender and age is required in addition to knowing the specific activity that was executed, determining motor skill competence on a large-scale in free-living children is probably impossible. From the results it is evident that movements can be distinguished to a certain extent and bone loading can be estimated using wearable motion trackers.

Pedometers

Looking more thoroughly to the individual conclusions of the different included studies, several connections can be made between the different researches. Two important variables for quantification of PA in children are the sensor type and placement of the device. When looking at the appropriateness of the different types of available sensors it becomes evident that there are contradicting results described in the included studies. Especially on the topic of using PMs for determining PA levels there is disagreement in literature: Zeng et al. [26], Beets et al. [27] and Duncan, S. et al. [28] conclude that waist-worn PMs do not provide a reliable

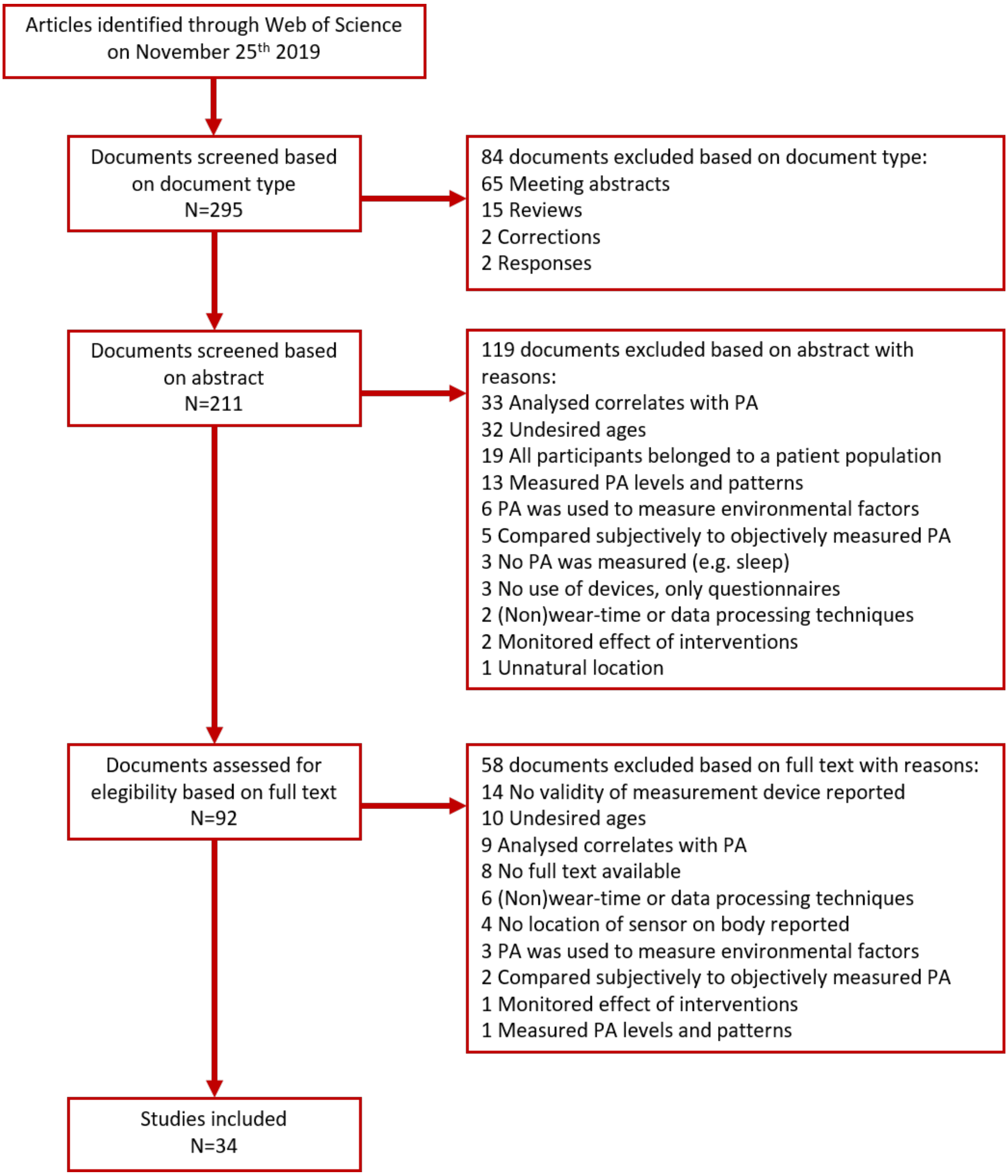


Figure 1: Flowchart of inclusion process

way for measuring intensity, whereas Trapp et al. [29], Louie et al. [30], Kilanowski et al. [31] and Ye et al. [32] suggest that it would offer a good alternative to HR monitors, uni- or triaxial AMs. A notable distinction can however be made between these two groups, all studies reporting PMs to be inaccurate compared the PM to a uni- or triaxial AM, whereas the studies that decided in favour of the PM used EE or observations with beforehand estimated METs or PA intensity levels as criterion value. Ye et al. [32] and Kilanowski et al. [31] stated that the PM and AM correlate better at higher intensities. It is plausible that during the free-living circumstances no sufficient amount of high intensity levels were reached. It is therefore also possible that the negative outcomes are due to the accuracy of the device the PM was compared against, rather than the accuracy of the PM itself. If the results of Zeng et al. [26] are left out of this comparison, for reasons that will be explained later, another distinction can be made: Beets et al. [27] and Duncan, S. et al. [28] tested the PM in free-living circumstances, while the other studies analyzed instructed activities or walking and running. It can thus not be ascertained whether the criterion value, the performed activities or a combination caused the ambiguity in conclusions on waist-worn PMs. Based on this uncertainty in literature, it cannot be concluded that waist-worn PMs present a good alternative to AMs or HR monitors for measuring PA intensity levels.

Authors (year)	Objective	Criterion Value	Conclusion	Remarks	Setting
Kilanowski et al. (1999) [31]	Compare PA quantities from PM on the left hip and a triaxial AM on the right hip	Observations and estimated intensity levels	PM, AM and observations were highly correlated for the combination of classroom and recreational activities and for only recreational activities, with r>0.94 (P<0.001) for all comparisons. However, for the lower levels of activity that occurred in the classroom activities the correlations were lower.	PM and AM yield more accurate results for higher intensity activities	Instructed activities
Louie et al. (1999) [30]	Validation of a HR monitor, PMs on the ankle, hip and wrist, a uniaxial AM on the left hip and a triaxial AM on the right hip for determining quantity of PA	EE from indirect calorimetry	All measurements except for those originating from the wrist PM showed significant correlations with EE (P<0.001). The triaxial AM predicted EE most accurately, but HR monitor and the hip-worn PM are also good predictors. The uniaxial AM performed worse than the hip PM. Combining the triaxial AM with the HR monitor resulted in the strongest combination, in which 91.6% of the variance was explained (P<0.01).	Only boys participated in this study. A HR monitor, a PM on the hip or a triaxial AM on the hip yield comparable results, but a PM outperforms a uniaxial AM. Adding HR monitor improved validity	Instructed activities
Ott et al. (2000) [33]	Compare PA quantities from a uniaxial AM on right hip, triaxial AM on the left hip and a HR monitor	Estimated activity intensities	The triaxial AM showed the highest correlation with the estimated intensity whereas HR exhibited the best correlation with METs.	Triaxial AM and HR are better predictors of respectively Intensity levels and METs then uniaxial AM	Instructed activities
Nilsson et al. (2002) [34]	Compare lower back to right hip placement of a uniaxial AM for measuring quantity of PA		No significant difference between both placements for high and very high intensity levels (P=.58 and P=.17 respectively), only at moderate intensities with 5 second epochs there is a significant effect of the wear location (P<0.01)	Uniaxial AM on lower back and right hip yield comparable results, epoch settings are more important to take into consideration than placement	Free-living
Hoos et al. (2003) [35]	Validation of a triaxial AM on the lower back for determining quantity of PA	TDEE from doubly labeled water method	TDEE and the triaxial AM worn on the lower back correlated well with each other (r=0.79 (P<0.01))	Placing a triaxial AM on the lower back is a valid method for measuring PA intensity levels in free-living children	Free-living

Lanningham-Foster et al. (2005) [24]	Validation of inclinometers and triaxial AMs for the assessment of posture allocation	Observations and EE from indirect calorimetry	Pants containing two inclinometers positioned on each mid-thigh and one AM at the lower back could not distinguish sitting and lying from each other, but as both are SB it is possible to discriminate SB from non-SB. Pants counting four inclinometers, two on the lower thigh and two on the waist, and two AMs at the lower back did not give any faults when measuring body position. Triaxial AM data correlated strongly with EE ($r>0.97$ for all children)	Distinctions between lying, sitting and standing can be made using a minimum of 4 inclinometers, triaxial AM on the lower back allows for accurate measurement of PA intensities	Lie down, sitting and standing
Tanaka et al. (2007) [36]	Validation of triaxial AM for assessing quantity of PA & Comparing dominant wrist to left hip placement	EE from indirect calorimetry	The correlations found between AM counts and EE were between 0.878 and 0.932. However, for stair climbing and ball tossing, predictions derived from AM counts underestimated EE by over 30%. The results from the device on the left hip correlated better with EE than the data from the one on the wrist.	Left hip placement outperforms wrist placement in instructed activities. Stair climbing and ball toss are hard to measure with AM on left hip.	Instructed activities
Hussey et al. (2009) [37]	Validation of a triaxial AM on the right hip for assessing quantity of PA	EE from indirect calorimetry	Correlations between EE and AM counts were lowest for the slowest walking velocity ($r=0.56$ ($P<0.01$)) and highest at 9 km/h ($r=0.84$ ($P<0.01$))	Higher correlations for triaxial AM counts and EE at higher PA intensity levels	Walking and running
Krishnaveni et al. (2009) [38]	Validation of uniaxial AM on the right hip	TEE from doubly labeled water method	There is no significant relationship between TEE and uniaxial AM counts	Possibly there is a relationship with PA intensities	Free-living
Ye et al. (2010) [32]	Compare PM counts on left and right hip & Compare PA quantities from uniaxial AM and PM on the left hip	Estimated METs	PM counts from the two locations were strongly correlated, the calculated intra class coefficient was at least 0.812 ($P<0.05$). For all speeds PM and AM results were significantly correlated ($r=0.850$, $r=0.829$ and $r=0.685$ for 4, 6 and 8 km/h (all $P<0.05$)). The correlation between PM and AM was higher during PE ($r=0.819$. $P=<0.05$) than at recess ($r=0.703$, $P=<0.05$). At higher speeds PM counts from left and right hip were less consistent	With increasing speeds the correlation between PM and AM becpmes weaker, PM might be less accurate at increased velocities	Walking and running
Beets et al. (2011) [27]	Compare PA quantities from uniaxial AM and PM on the right hip		PM consistently underestimates minutes spent in MVPA when compared to a uniaxial AM	PM and uniaxial AM outcomes differ	Free-living

Duncan, S et al. (2011) [28]	Compare PA quantities from a PM and an omnidirectional AM worn on the hip		The PM reported significantly lower MVPA than the AM, as MVPA was underestimated by 37 to 45% ($P<0.001$), also precision of the PM was concluded to be low because the 95% limits of agreement were wide with a difference of around 200% between the lower and upper boundary	PM underestimates MVPA with respect to omnidirectional AM counts	Free-living
Pulsford et al. (2011) [39]	Validation of uniaxial AM on the right hip for measuring PA quantities	EE from indirect calorimetry	Except for jogging, all non-sedentary activities revealed significant relationships with METs and AM counts ($P<0.001$). There was a strong association between AM counts and EE, although sensitivity and specificity were higher for SB(99% and 97%) and VPA(95% and 91%) than for LPA(60% and 83%) and MPA(61% and 76%). Activities in which the upper body was more involved, basketball for example, yielded too low AM counts for their PA level.	Uniaxial AM can predict SB and VPA with high accuracy, LPA and MPA can still be predicted but with lower accuracy. Hip placement is not suitable for movements in which the upper body is more evidently involved	Instructed activities
Trost et al. (2011) [21]	Validation of a uniaxial AM on the right hip for measuring quantity of PA	EE from indirect calorimetry	Sensitivity and specificity of SB(100% and 79.3%), LPA(49.3% and 91.5%), MPA(60.0% and 88.3%), VPA(73.7% and 93.8%) and MVPA(88.3% and 91.7%) can be accurately distinguished by an uniaxial AM	There is a relatively low sensitivity for recognizing LPA and MPA with a hip mounted uniaxial AM	Instructed activities
Neugebauer et al. (2012) [40]	Validation of biaxial AM on the right hip for estimating GRF	Force plate data	GRF and AM data are positively related with one another ($r=0.967$($P<0.001$) or ($r=0.877$ ($P<0.001$), depending on the model)	GRF from walking an running can be estimated based on hip acceleration	Walking and running
Ojiambo et al. (2012) [41]	Validation of a uniaxial AM on the right hip combined with HR monitoring and a triaxial AM on the right hip for measuring quantity of PA	TEE from doubly labeled water method	The uniaxial AM combined with HR monitoring yields comparable accuracy to triaxial AM, there exists a significant positive relationship between PA levels and data from both AMs. Only using a uniaxial AM does not result in a significant relationship with EE. The combined data from a uniaxial AM and HR monitor were a significant predictor of EE, because adding HR information to uniaxial AM measurements increased the predictive validity, resulting a correlation coefficient of $r=0.61$ ($P<0.05$) between the tested devices and PA levels. TEE could be predicted with higher accuracy than AEE.	Both triaxial and uniaxial AM on the right hip had significant positive relationships with PA and TEE, adding HR measurements could improve accuracy	Free-living

Routen et al. (2012) [42]	Compare PA quantities from right hip and non-dominant wrist placement of omnidirectional AM		The total hip AM count was significantly lower than that of the wrist (P<0.01).	Wrist placement overestimates PA with respect to hip placement	Free-living
Phillips et al. (2013) [43]	Compare PA quantities from wrist and hip placement of a triaxial AM & Validation of a triaxial AM and uniaxial AM on the hip	EE from indirect calorimetry	When comparing the locations of the triaxial AMs with each other, hip placement results in a significantly higher validity against the criterion value (r=0.970 (P<0.05)). Measuring at the hip with a triaxial AM yields an overall higher sensitivity than when placed on the wrist, while the correlation of uniaxial AM with EE was equal to that of the triaxial AM on the hip. Wrist worn monitors reported high intensities during Wii games while MET values were low	Hip placed triaxial AMs are overall more accurate than wrist-worn ones, however a triaxial AM located at either wrist or right hip as well as a uniaxial AM on the hip can validly discriminate between activity intensity levels. Active gaming is overestimated by wrist-worn triaxial AM	Instructed activities
Reading et al. (2013) [44]	Validation of a triaxial AM on the right hip for determining PA quantities	EE from indirect calorimetry	The AM overestimated the time spent in SB and LPA but underestimated the occurence of MPA and VPA. The average MET activity as derived from EE was significantly higher than what was measured using AMs. There exists a non linear relationship between EE and AM counts of intermittent and non-steady state activities.	Hip placement during active video gaming, which consists often of upper body movements	Active video gaming
Trapp et al. (2013) [29]	Validation of PM on the waist for estimating PA quantities	Observations	PMs accurately measure the amount of steps that are taken. The absolute value of percentage error between the criterion and measured data was greatest at the lowest speed.	Higher intensity levels yield more accurate results for PMs on the waist	Walking and running
Rowlands et al. (2014) [45]	Compare PA quantities from wirst and hip placement of a triaxial AM & Compare uniaxial and triaxial AM on the right hip		The triaxial data from both locations correlated strongly with the uniaxial hip counts. Time spent in SB had a correlation of >0.87 (P<0.001) and time in MVPA a correlation of >0.83. In LPA the wrist-worn triaxial AM only correlated for 0.61 to 0.63 (P<0.001) and the hip-worn triaxial AM dit not have a significant relation to the uniaxial AM.	There is a difference in triaxial AM worn on the hip and wrist for estimations of MVPA (the most important intensity for the purpose of this research). Uniaxial and triaxial AM data of the right hip differ significantly in free-living children	Free-living

Jang et al. (2015) [46]	Validation of a neural network with triaxial AM data from the wrist, waist and ankle for distinguishing different movements	EE from indirect calorimetry	With a 3-stage network it was possible to categorize walking with an accuracy of 95.97%, all other activities could be recognized with an accuracy of over 90%, EE was predicted with an average accuracy of 81.91%	Not only the type of movement that is made can be identified from a neural network with triaxial AM data as input, also the PA intensity level can be determined from triaxial AM data	Instructed activities
Meyer et al. (2015) [47]	Validation of a triaxial AM at the hip for estimating GRF	Force plate data	The recordings of a triaxial AM correlated with GRF (r=0.90(95% CI=0.68-0.97)), however GRF was consistently overestimated by the AM. Tasks with higher locomotion velocities and landing heights featured higher GRFs (P <0.001). Walking had the lowest GRF, followed by jogging and running and landing tasks had the highest GRF.	GRF can be calculated from triaxial AM data, but be careful of systematical overestimation of GRF when measured this way. GRF could potentially generate information on the type of movement that is executed	Walking, running and jumping activities
Chandler et al. (2016) [48]	Validation of a triaxial AM on the wrist for measuring quantity of PA	HR reserve	A wrist-placed triaxial AM can classify SB, LPA, and MVPA excellently, MPA is a bit more difficult but can still be classified	Only study with HRR as criterion value	Instructed activities
Ren et al. (2016) [49]	Validation of a neural network with triaxial AM data from the right hip for distinguishing different movements	Classification of intensity levels	Overall, activities could be identified with an accuracy of 64% and intensity estimation was 92% accurate. Continuous activities were more easy to classify than the intermittent ones. Some video games were the hardest to recognize out of all before mentioned activities. Intensity levels of the video games and sports and exercise categories got recognized with the lowest precision.	Triaxial AM data combined with a machine learning algorithm allows for classifying different executed activities, when the activities are known beforehand and PA intensity levels can be accuratly determined from this data	Instructed activities
Bisi et al. (2017) [50]	Determine whether five triaxial IMUs can be used to objectively measure motor skill development	Observations combined with a survey	Measurements from triaxial IMUs on each ankle, each wrist and the lower back show an agreement with motor competence assessed from observations of at least 77%	Not feasible for large populations	Instructed activities
Canete Garcia-Prieto et al. (2017) [22]	Comparison of uniaxial AM on the right hip and HR monitor for determining quantity of PA & Compare endurance games and strength games	EE from indirect calorimetry	There was no significant correlation between uniaxial AM and EE for strength games (r=0.21 (P=0.574)). HR correlated better with EE (r=0.71 (P=0.032) for endurance games and r=0.48 (P=0.026) for strength games) than a uniaxial AM (r=0.48 (P=0.026) for endurance games)	The low correlations for AM and strength games could be due to the right hip placement of the AM, while most strength games involved the upper body and movements were not noticed by the device	Free-play

Gao et al. (2018) [51]	Validation of a triaxial AM worn anteriorly on the waist for estimating neuromuscular loading	EMG	AM and EMG recordings did not agree with each other for intensity levels other than SB, the AM counted more LPA and less MPA than the EMG did. AM measurements poorly represent neuromuscular activity	It is currently not possible to derive EMG information from AM measurements, would this be relevant?	Instructed activities
Kim et al. (2018) [52]	Compare PA quantities from uniaxial AM and triaxial AM on the wrist, comparing waist and wrist placement of a triaxial AM		The uniaxial AM showed moderate convergent validity when compared to both triaxial AMs for MVPA (r=0.64–0.75),and SB (r=0.45–0.66), but weak convergent validity for assessing LPA. Uniaxial AM consistently overestimated MVPA. Both triaxial AMs correlated strongly (r0.87)	Uniaxial AM and triaxial AM data differ, waist and wrist placed triaxial AM yield comparable results	Free-living
Mooses et al. (2018) [53]	Compare PA quantities from hip-worn Fitbit Zip and a triaxial AM		Fitbit Zip overestimated the number of steps and time in SB and underestimated MVPA with respect to the triaxial AM. The more active participants were, the more steps and MVPA were overestimated. When children were sedentary, MVPA was to a greater extent underestimated.	Fitbit overestimates more on higher intensity levels and underestimates more on lower intensity levels	Free-living
Zeng et al. (2018) [26]	Compare PA qunatities from a PM and triaxial AM at the waist		There is no significant relationship between the measured MVPA quantity by the PM and triaxial AM (r=0.027, P=0.597). Both devices are unreliable methods for finding PA levels of exergaming children.	The activity monitors were placed at the waist while active gaming is mostly done with the upper body, likely that movements have been missed	Active video gaming
Clark (2019) [23]	Validation of raw triaxial AM data of the ankle for assessing quality and quantity of movement	Observations by means of an optical motion tracking system	Spectral purity-derived movement quality can be computed from raw triaxial AM data. Differences in motor competence between children of separte age groups can be established. Raw accelerometry is furthermore an accurate and valid option for assessing quantity (absolute variance: $<$0.001 g, coefficient of variation: 0.004%, in all axes)	An ankle worn triaxial AM can be used for accurately determining quantity and quality of movement	Instructed activities

Duncan, M et al. (2019) [25]	Comparing placement of triaxial AMs on both wrists, dominant waist and dominant ankle for assessing quantity of PA	EE from indirect calorimetry	When cycling was included all relations between MET and AM were significantly weak, except for the ankle-worn AM which had a moderate relationship (r=0.752 (P<0.01)). When cycling was omitted from the measurements, correlations improved for all wear locations. Ankle placement was still associated the strongest with METs (r=0.790 (P<0.01)	Out of these locations, the ankle placement can best discriminate between the different PA intensity levels both with and without cycling as part of the activities	Instructed activities with cycling, other lower and upper body movements
Kang et al. (2019) [54]	Validation of wrist worn Fitbit Charge HR trackers and triaxial AM on each wrist for mesuring quantity of PA	EE from indirect calorimetry	The Fitbit achieved good results in classifying SB (acc: 80.73%) and weaker results when identifying LPA (acc: 66.1%) and MVPA (acc: 70.8%). The triaxial AM showed higher accuracy and specificity for assessing MVPA (acc: 81.9%, sp:94.3%). Omitting the cycling activity from the data caused all results to improve further. Dominant and non-dominant wrist placement yield comparable results.	Triaxial AM is more accurate than Fitbit, cylcing cannot accurately be measured from wrist worn sensors	Instructed activities

Table 1: Results of included articles

Uniaxial accelerometers
Furthermore, some studies conclude that a uniaxial AM can achieve results that are comparable to those of a triaxial AM, like Phillips et al. [43] Pulsford et al. [39] Trost et al. [21] and Louie et al. [30]. However, other research, as that conducted by Ojiambo et al. [41] and Krishnaveni et al. [38] declare that the previous statement is not true. The difference between the former and latter group is in what setting the measurements were conducted; Ojiambo et al. [41] and Krishnaveni et al. [38] both looked at free-living conditions while the others investigated uniaxial AMs for instructed activities. All uniaxial AMs were worn on the right hip except for Louie et al. [30], who placed them on the left hip. Rowlands et al. [45] and Kim et al. [52] were more ambiguous on the validity of uniaxial AMs with respect to triaxial AMs and both declared that SB and MVPA correlated significantly and strong between both devices, however this was not the case for LPA. For reasons similar to those given in the paragraph about the PM, uniaxial AMs are not proven to provide accurate measurements regarding quantity of PA in free-moving children. However, Ojiambo et al. [41] did find a significant relation between EE and the uniaxial AM when the AM data was combined with a HR monitor

Fitbits, bi- and triaxial accelerometers and heart rate monitors
Both studies that investigated a different version of the Fitbit, Kang et al. [54] and Mooses et al. [53], concluded that it was not a good predictor for activity levels other than SB. With the exception of Zeng et al. [26], all studies that investigated a triaxial AM (Duncan, M. et al. [25], Hoos et al. [35], Hussey et al. [37], Tanaka et al. [36], Chandler et al. [48], Louie et al. [30], Kang et al. [54], Clark [23], Ojiambo et al. [41], Jang et al. [46], Ren et al. [49] and Gao et al. [51]) proved that a triaxial AM on ankle, wrist and waist correlates strongly with EE and PA intensities and therefore allows to classify PA levels correctly. Moreover, comparing the results of Chandler et al. [48] to those of Trost et al. [21] and Pulsford et al. [39] reveals that a wrist worn triaxial AM better estimates LPA and MPA than a uniaxial AM on the right hip. MVPA is comparable between the two devices. Ott et al. [33] also showed that a triaxial AM and HR monitor correlate better with METs and PA intensities than a uniaxial AM. The fact that HR monitoring provides adequate measurements is supported by Louie et al. [30] and Canete Garcia-Prieto et al. [22], the latter even suggested that a HR monitor functions better than a biaxial AM on the right hip, assumed based on the higher correlations with EE. It was furthermore suggested by Louie et al. [30] and Ojiambo et al. [41] to improve the accuracy of an AM by combining these data with information from an HR monitor. Ojiambo et al. [41] proved this with a combination of a uniaxial AM and HR monitor, but Louie et al. [30] added the HR monitor to both uni- and triaxial AM data and reported that the combination with a triaxial AM resulted in the strongest correlation with EE. Canete Garcia-Prieto et al. [22] and Ott et al. [33] even concluded that solely using HR monitoring allows for an accurate assessment of PA intensities.

Placement for determination of quantity
A large amount of studies used the wrist, right hip or waist as placement site, in contrast to the few studies mounting a sensor on the ankle. Duncan, M. et al. [25] mounted a triaxial AM on the ankle of children and concluded that the ankle worn device discriminated PA levels better than waist and wrist worn triaxial AMs. The ankle achieved the best results in both cases, so when cycling was included and left out of the measurements. Clark [23] was also able to distinguish different intensities using an ankle work AM. Louie et al. [30] put a PM on the ankle and compared it to PMs on the wrist and waist, the PM at the waist performed best of all locations, ankle was second. Accuracy could thus be subjected to the combined effect of sensor type and placement site. Tanaka et al. [36] also experimented with placement and saw that a triaxial AM on the dominant wrist lead to a lower performance than the left hip. Stair climbing and ball toss were underestimated by both AMs. There was furthermore a difference in omnidirectional AM placement under free-living conditions found by Routen et al. [42], who figured out that the total hip count was significantly lower than the counts from the device on the non-dominant wrist. According to Nilsson et al. [34] a uniaxial AM on the hip and lower back provide results that do not significantly differ from one another. Ye et al. [32] proved that a PM on the left hip yields results comparable to the right hip. Apparently there is a difference between the ankle, wrists and waist as a location for triaxial and omnidirectional AMs and PMs, but uniaxial AMs or PMs placed around the waist produce results that agree with each other.

Placement for specific activities
Two of the included studies solely investigated active gaming, but there were more that included active gaming as one of the instructed activities. Overall, active gaming activities showed low relationships with PA intensities, for instance Reading et al. [44] and Zeng et al. [26]. Upon examination of these studies, it becomes clear that they chose a waist placement for their devices. Active gaming, for example Nintendo Wii, requires movements from the upper body. The waist underestimated the PA levels, while devices placed on the wrist overestimated the intensities. For example Phillips et al. [43], who placed a triaxial AM on the wirst of Wii playing children, commented that the active gaming was overestimated by this device. Basketball was an activity part of the study done by Pulsford et al.

[39], in this case the uniaxial AM on the right hip underestimated the motions made by the upper body as well. Therefore the decision for the placements of the devices was poor for this specific activity. It would be interesting to test ankle placement f.e. in active gaming activities or free-living conditions. Unfortunately, none of the included studies investigated either. Other instances where the type of activity contributed to the results were reported by Duncan, M. et al. [25], where higher correlations were achieved once cycling was taken out of the results, and Canete Garcia-Prieto et al. [22], where the strength games involved upper body motions but the AM was placed on right hip, causing the device to not sense all movements made.

Measuring variety
When looking at the methods used to assess the variance of activities performed by children, there appear to be different aspects that can be considered. Bone loading was investigated by Neugebauer et al. [40] and Meyer et al. [47], which is important for bone development. Bisi et al. [50] and Clark [23] assessed motor skill development in children. Lanningham-Foster et al. [24] studied differentiation of body posture, while Jang et al. [46] and Ren et al. [49] succeeded in identifying the exact type of mevement that was made. The only facet of variety in PA that was concluded to be impossible to determine by an included research was neuromuscular loading, as attempted by Gao et al. [26].

Although inclinometers on the thighs and waist can successfully distinguish lying, sitting and standing from one another [24], this is not a relevant factor for this review's purposes. The sensor types that can be used for measuring relevant elements of variety were IMUs [50] and bi- or triaxial AMs [23],[47],[40] [46],[49]. IMUs were only proven to be useful in determination of motor skill development and biaxial AMs for finding the GRF, while from triaxial AM data motor skill development, the GRF and the exact type of movement can be deducted.

Placement for determination of variety
Similar to measuring quantity of PA, the placement on a child's body is also important to consider. Motor skill development can be measured by putting the triaxial AM on the ankle and the GRF can be calculated from hip-worn triaxial AM data. For the identification of movements, both included studies successfully implemented artificial intelligence procedures. Important to note are the measurement protocols used in the two researches, because Jang et al. [46] needed three sensors, on the wrist, waist and ankle, whereas Ren et al. [49] only required one AM on the right hip. All movements that were made by the participants were chosen previous to the study.

4 Discussion

Most information on movement patterns in children can be derived from raw triaxial AM data. The few studies that investigated triaxial AMs placed on the ankle of children reported good results for PA quantity derivation, although additional research needs to be done regarding this placement in free living conditions. Via machine learning, relevant movements that are made by children can be recognized and motor skills can be determined from these movements. Using machine learning to identify specific motions was done by Ren et al. [49], but their participants wore a triaxial AM on the hip. Clark [23] succeeded in determining quality by using a triaxial AM on the ankle. The possibility of doing both with only one device should be further analyzed.

Following the recommendations regarding the amount of PA in children, only MVPA and VPA are important to determine. Therefore, it is necessary to find a sensor that can at least distinguish these intensities accurately. Uniaxial AM and PM show poorer predictive values for estimating quantities in free living conditions and in this review no evidence has been found that these can be used as indicators for variety of movements. Despite the feasibility of HR monitoring for determining quantities of PA, this method is also not capable of delivering information appropriate for tracking variety in exercise. Furthermore, HR responses can also be triggered by emotional stimuli and for very high PA levels HR responses will exhibit a plateau effect. HR monitors could however be added to AMs to increase the predictive validity for measuring PA quantity, but the clinical relevance of this improvement should be examined before this decision is made.

One downside of using triaxial AMs is that they are relatively expensive compared to PMs, uniaxial AMs and even HR monitors. Even though a triaxial AM outperforms the alternatives in the fields of accuracy and feasibility when measuring quantity and variety of PA in free-playing children, it still might not be the 'best' solution for the specific problem posed in this review due to the associated high costs. Furthermore, three axes produce a lot more data than one or two axes. Therefore it is recommended to look into the possibility of reproducing (raw) triaxial AM data based on biaxial AM measurements. It is not recommended to use an uniaxial AM because from this review it was seen these devices appear to be accurate in lab settings but this turns out to be incorrect when looking at the free-living circumstances. A fabricated device containing a uniaxial AM might thus be wrongfully validated. Biaxial AMs have been proven to enable GRF estimation, but none of the included studies used a biaxial AM to investigate quantity. This sensor type might offer potential for an accurate wearable.

Because affordability is crucial for the intended use of this review, various alternatives should be considered. The choice can be made to equip a child with multiple cheap devices at the same or different locations on the body, as opposed to a single expensive gadget. From the results it is evident that all placements around the waist, including right and left hip, lower back and anteriorly at the waist, show comparable accuracy. The one study that compared a triaxial AM at the waist-location to the ankle, favored the ankle placement over the former. Wrist was reported to have lower accuracy in all studies that compared placements using EE as criterion value, and studies comparing locations often reported that the wrist overestimated the AM counts with respect to the hip. However, the wrist might be the most convenient location seen the objectives of this study. For children it is relatively easy to put on a bracelet that has to be fixated tight enough to not move independently of the body. Putting a sensor on the hip would require a belt of some sort, because clipping it to belt loops would allow for too much movement by the device. Strapping it on the ankle is harder to take off and on, it would be a shame if the device is not used because of that.

Two studies reported that cycling was a difficult motion for determination of quantity of PA [25],[54]. Since the prototype is meant to be used in The Netherlands, a country that is known for the number of people that ride bikes, that is inconvenient. Free-moving children are very likely to bike, but these biking motions might get unnoticed, systemically biasing measurements. It is therefore, regardless which sensor and placement will be chosen in the end, recommended to investigate ways of compensating this bias.

If motor skills of an individual need to be assessed, it is required to know the exact movement that was made. Identification of executed movements is possible by application of machine learning approaches. Artificial intelligence algorithms can however only distinguish beforehand selected activities, which makes this method probably infeasible to use in free-living children. Another way of knowing the executed movement, is if the motions made while wearing the device are determined beforehand, instructed activities for instance. This could be alternated with free play, hereby estimating quantity from free play, knowing the variety because certain movements are instructed and learning quality from the execution of instructed motions. When choosing this manner of collecting data, variety of true free-living conditions cannot be observed from these recordings, but this concept potentially allows for estimation of motor skill development.

A statement made by Nilsson et al. [34] was that settings of the uniaxial AM are more important than placement. An important setting of activity monitors that varied over the studies concerning PA in children is the epoch length, the frequency at which data is registered. Since PA in youth is highly intermittent due to the median bout length of $\geq$6 seconds for low and medium and even $\geq$3 seconds in high intensity activities [55], it is advisable to choose a relatively short epoch time or high sample frequency. Longer epochs result in bias, they overestimate LPA, MPA and MVPA but underestimate SB and VPA [56]. In other words, the more extreme intensity levels are leveled out when using larger epoch lengths. Due to this effect it is recommended to measure PA in children with an epoch length of maximally 5 seconds [56].

Some other relevant topics have not been discussed in this review. Examples of this are affordability, size, easiness of use, comfort, appearance and foolproofness. This is partially related to user friendliness, which is very important because there is no value in creating a device that no one will use. This could influence the reliability of the measurements as well, when children wear the device on an incorrect location, the recorded information may be biased.

Two additional matters are gender and age. Boys and girls develop differently, however gender differences are not considered. This could also affect the conclusions drawn on quality of movement from measurements by the final device. It could be useful to either uncover the gender of the users or produce separate toys expecting that boys and girls might have different preferences regarding the looks of a gadget. The latter is of course not a certainty and could actually result in faulty measurements when the preferences do not correlate with gender. Age is also an issue because what is considered 'normal' motor skill development depends on age. It should thus be known how old the wearer of a specific gadget is. These are all considerations that should be analyzed in detail while conceiving the final design.

This review was conducted to contribute to the development of a device or toy for children that will encourage PA and record data for research purposes. The intended product will be aimed at children aged between 6 and 12 years. Besides adhering to the developmental state of the intended users of the product, the safety is an important feature to take into consideration as well. A toy with a diameter of 3.2 cm and a length ranging between 2.5 and 5.7 cm is considered a small toy [57]. It is recommended to manufacture "small toys" only for children with a minimum age of 6 years due to choking hazards in lower ages [58]. Large-scale data on children's PA would be useful for research purposes. It might for example be possible to recognize patterns in movement and work on prevention of overweight and depression among youth.

Only currently existing sensors, that are already being used for the investigated purposes, have been mentioned in this review. There might be alternatives that were not discussed because they have not yet been used to measure the objectives of this review. Therefore, it is recommended to not only fixate on the devices specified in this research, but also look for instruments that are more innovative than the accepted methods.

5 Conclusion

Assessing variety is very ambitious, if not impossible, to achieve in free-living children on a large-scale. Uniaxial AMs and PMs do not meet the desired accuracy, according to this review's results. Bi- or triaxial accelerometers have potential to gain knowledge on quantity of PA, motor skill development, bone loading and identification of movements. It is known that triaxial AMs can measure all of these for instructed activities, however the latter three have not been investigated in free-living settings by any of the included documents. Giving instructions for execution of a certain movement by the wearer is an alternative to overcome this obstacle. Biaxial AMs have only been verified for determining GRF, but as triaxial AMs are expensive it might be worth looking into this option. Although wrist placement yields lower validity than ankle or waist placements, it is a more convenient wear-site for children.

References

[1] S. Biddle, 'Psychological benefits of excercise and physical activity', *Revista de psicologıa del deporte*, vol. 2, no. 2, 1993.

[2] S. J. 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.

[3] G. Naughton, D. Greene, D. Courteix and A. Baxter-Jones, *Resilient, responsive, and healthy developing bones: The good news about exercise and bone in children and youth*, 2017.

[4] M. Sothern, M. Loftin, R. Suskind, J. Udall and U. Blecker, 'The health benefits of physical activity in children and adolescents: Implications for chronic disease prevention', *European journal of pediatrics*, vol. 158, no. 4, pp. 271–274, 1999.

[5] G. V. Russell, C. W. Pierce and L. Nunley, 'Financial implications of obesity', *Orthopedic Clinics*, vol. 42, no. 1, pp. 123–127, 2011.

[6] W. H. Organization *et al.*, *Global recommendations on physical activity for health.* World Health Organization, 2010.

[7] R. Guthold, G. A. Stevens, L. M. Riley and F. C. Bull, 'Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1· 6 million participants', *The Lancet Child & Adolescent Health*, 2019.

[8] F. Lima, V. De Falco, J. Baima, J. G. Carazzato and R. M. PEREIRA, 'Effect of impact load and active load on bone metabolism and body composition of adolescent athletes', *Medicine & Science in Sports & Exercise*, vol. 33, no. 8, pp. 1318–1323, 2001.

[9] L. M. Hidding, M. J. Chinapaw, M. N. van Poppel, L. B. Mokkink and T. M. Altenburg, 'An updated systematic review of childhood physical activity questionnaires', *Sports Medicine*, vol. 48, no. 12, pp. 2797–2842, 2018.

[10] J. F. Sallis, 'Self-report measures of children's physical activity', *Journal of School health*, vol. 61, no. 5, pp. 215–219, 1991.

[11] B. Bringolf-Isler, U. Mäder, N. Ruch, S. Kriemler, L. Grize and C. Braun-Fahrländer, 'Measuring and validating physical activity and sedentary behavior comparing a parental questionnaire to accelermeter data and diaries', *Pediatric exercise science*, vol. 24, no. 2, pp. 229–245, 2012.

[12] A. P. Hills, N. Mokhtar and N. M. Byrne, 'Assessment of physical activity and energy expenditure: An overview of objective measures', *Frontiers in nutrition*, vol. 1, p. 5, 2014.

[13] K. Libertus and R. J. Landa, 'The early motor questionnaire (emq): A parental report measure of early motor development', *Infant Behavior and Development*, vol. 36, no. 4, pp. 833–842, 2013.

[14] J.-A. K. Finegan, G. A. Niccols, J. E. Zacher and J. E. Hood, 'The play activity questionnaire: A parent report measure of children's play preferences', *Archives of sexual behavior*, vol. 20, no. 4, pp. 393–408, 1991.

[15] U. C. P. S. Commission, J. A. Therell *et al.*, *Age determination guidelines: relating children's ages to toy characteristics and play behavior.* Consumer Product Safety Commission, 2002.

[16] N. Ahmad, R. A. R. Ghazilla, N. M. Khairi and V. Kasi, 'Reviews on various inertial measurement unit (imu) sensor applications', *International Journal of Signal Processing Systems*, vol. 1, no. 2, pp. 256–262, 2013.

[17] M. Tanenhaus, D. Carhoun, T. Geis, E. Wan and A. Holland, 'Miniature imu/ins with optimally fused low drift mems gyro and accelerometers for applications in gps-denied environments', in *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, IEEE, 2012, pp. 259–264.

[18] R. O'Driscoll, J. Turicchi, K. Beaulieu, S. Scott, J. Matu, K. Deighton, G. Finlayson and J. Stubbs, 'How well do activity monitors estimate energy expenditure? a systematic review and meta-analysis of the validity of current technologies', *British journal of sports medicine*, bjsports–2018, 2018.

[19] B. A. Lynch, T. K. Kaufman, T. Rajjo I, K. Mohammed, S. Kumar, M. H. Murad, N. E. Gentile, G. A. Koepp, S. K. McCrady-Spitzer and J. A. Levine, 'Accuracy of Accelerometers for Measuring Physical Activity and Levels of Sedentary Behavior in Children: A Systematic Review', *JOURNAL OF PRIMARY CARE AND COMMUNITY HEALTH*, vol. 10, SEP 2019, ISSN: 2150-1319. DOI: {10 . 1177 / 2150132719874252}.

[20] P. H. Veltink, H. B. Bussmann, F. Koelma, H. M. Franken, W. L. Martens and R. C. van Lummel, 'The feasibility of posture and movement detection by accelerometry', in *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ*, IEEE, 1993, pp. 1230–1231.

[21] S. G. Trost, P. D. Loprinzi, R. Moore and K. A. Pfeiffer, 'Comparison of Accelerometer Cut Points for Predicting Activity Intensity in Youth', *MEDICINE AND SCIENCE IN SPORTS AND EXERCISE*, vol. 43, no. 7, 1360–1368, JUL 2011, ISSN: 0195-9131. DOI: {10.1249/MSS.0b013e318206476e}.

[22] J. Canete Garcia-Prieto, V. Martinez-Vizcaino, A. Garcia-Hermoso, M. Sanchez-Lopez, N. Arias-Palencia, J. F. Ortega Fonseca and R. Mora-Rodriguez, 'Energy Expenditure in Playground Games in Primary School Children Measured by Accelerometer and Heart Rate Monitors', *INTERNATIONAL JOURNAL OF SPORT NUTRITION AND EXERCISE METABOLISM*, vol. 27, no. 5, 467–474, OCT 2017, ISSN: 1526-484X. DOI: {10.1123/ijsnem.2016-0122}.

[23] C. C. T. Clark, 'Profiling movement and gait quality using accelerometry in children's physical activity: consider quality, not just quantity', *BRITISH JOURNAL OF SPORTS MEDICINE*, vol. 53, no. 8, 493–494, APR 2019, ISSN: 0306-3674. DOI: {10 . 1136 / bjsports-2017-098204}.

[24] L. Lanningham-Foster, T. Jensen, S. McCrady, L. Nysse, R. Foster and J. Levine, 'Laboratory measurement of posture allocation and physical activity in children', *MEDICINE AND SCIENCE IN SPORTS AND EXERCISE*, vol. 37, no. 10, 1800–1805, OCT 2005, ISSN: 0195-9131. DOI: {10 . 1249 / 01 . mss.0000175050.03506.bf}.

[25] M. J. Duncan, C. M. P. Roscoe, M. Faghy, J. Tallis and E. L. J. Eyre, 'Estimating Physical Activity in Children Aged 8-11 Years Using Accelerometry: Contributions From Fundamental Movement Skills and Different Accelerometer Placements', *FRONTIERS IN PHYSIOLOGY*, vol. 10, MAR 18 2019, ISSN: 1664-042X. DOI: {10 . 3389 / fphys . 2019 . 00242}.

[26] N. Zeng, X. Gao, Y. Liu, J. E. Lee and Z. Gao, 'Reliability of Using Motion Sensors to Measure Children's Physical Activity Levels in Exergaming', *JOURNAL OF CLINICAL MEDICINE*, vol. 7, no. 5, MAY 2018, ISSN: 2077-0383. DOI: {10.3390/jcm7050100}.

[27] M. W. Beets, C. F. Morgan, J. A. Banda, D. Bornstein, W. Byun, J. Mitchell, L. Munselle, L. Rooney, A. Beighle and H. Erwin, 'Convergent Validity of Pedometer and Accelerometer Estimates of Moderate-to-Vigorous Physical Activity of Youth', *JOURNAL OF PHYSICAL ACTIVITY & HEALTH*, vol. 8, no. 2, S295–S305, SEP 2011, ISSN: 1543-3080. DOI: {10.1123/jpah.8.s2.s295}.

[28] S. Duncan, K. White, L. Sa'ulilo and G. Schofield, 'Convergent Validity of a Piezoelectric Pedometer and an Omnidirectional Accelerometer for Measuring Children's Physical Activity', *PEDIATRIC EXERCISE SCIENCE*, vol. 23, no. 3, 399–410, AUG 2011, ISSN: 0899-8493. DOI: {10.1123/pes.23.3.399}.

[29] G. S. A. Trapp, B. Giles-Corti, M. Bulsara, H. E. Christian, A. F. Timperio, G. R. McCormack and K. Villanueva, 'Measurement of children's physical activity using a pedometer with a built-in memory', *JOURNAL OF SCIENCE AND MEDICINE IN SPORT*, vol. 16, no. 3, 222–226, MAY 2013, ISSN: 1440-2440. DOI: {10.1016/j.jsams.2012.06.011}.

[30] L. Louie, R. Eston, A. Rowlands, K. Tong, D. Ingledew and F. Fu, 'Validity of heart rate, pedometry, and accelerometry for estimating the energy cost of activity in Hong Kong Chinese boys', *PEDIATRIC EXERCISE SCIENCE*, vol. 11, no. 3, 229–239, AUG 1999, ISSN: 0899-8493. DOI: {10.1123/pes.11.3.229}.

[31] C. Kilanowski, A. Consalvi and L. Epstein, 'Validation of an electronic pedometer for measurement of physical activity in children', *PEDIATRIC EXERCISE SCIENCE*, vol. 11, no. 1, 63–68, FEB 1999, ISSN: 0899-8493. DOI: {10.1123/pes.11.1.63}.

[32] Y. W. Ye and M. Chia, 'VALIDITY AND RELIABILITY OF OMRON HJ-005 PEDOMETER IN QUANTIFYING FIELD-BASED PHYSICAL ACTIVITY AMONG SINGAPOREAN CHILDREN', in *SPORT SCI-*

ENCE AND STUDIES IN ASIA: ISSUES, REFLECTIONS AND EMERGENT SOLUTIONS, Chia, M and Chiang, J, Ed., 2010, 23–42, ISBN: 978-981-4304-08-5. DOI: {10.1142/9789814304092_0002}.

[33] A. Ott, R. Pate, S. Trost, D. Ward and R. Saunders, 'The use of uniaxial and triaxial accelerometers to measure children's "free-play" physical activity', *PEDIATRIC EXERCISE SCIENCE*, vol. 12, no. 4, 360–370, NOV 2000, ISSN: 0899-8493. DOI: {10.1123/pes.12.4.360}.

[34] A. Nilsson, U. Ekelund, A. Yngve and M. Sjostrom, 'Assessing physical activity among children with accelerometers using different time sampling intervals and placements', *PEDIATRIC EXERCISE SCIENCE*, vol. 14, no. 1, 87–96, FEB 2002, ISSN: 0899-8493. DOI: {10.1123/pes.14.1.87}.

[35] M. Hoos, G. Plasqui, W. Gerver and K. Westerterp, 'Physical activity level measured by doubly labeled water and accelerometry in children', *EUROPEAN JOURNAL OF APPLIED PHYSIOLOGY*, vol. 89, no. 6, 624–626, AUG 2003, ISSN: 1439-6319. DOI: {10.1007/s00421-003-0891-6}.

[36] C. Tanaka, S. Tanaka, J. Kawahara and T. Midorikawa, 'Triaxial accelerometry for assessment of physical activity in young children', *OBESITY*, vol. 15, no. 5, 1233–1241, MAY 2007, ISSN: 1930-7381. DOI: {10.1038/oby.2007.145}.

[37] J. Hussey, K. Bennett, J. O. Dwyer, S. Langford, C. Bell and J. Gormley, 'Validation of the RT3 in the measurement of physical activity in children', *JOURNAL OF SCIENCE AND MEDICINE IN SPORT*, vol. 12, no. 1, 130–133, JAN 2009, ISSN: 1440-2440. DOI: {10.1016/j.jsams.2007.09.010}.

[38] G. V. Krishnaveni, S. R. Veena, R. Kuriyan, R. P. Kishore, A. K. Wills, M. Nalinakshi, S. Kehoe, C. H. D. Fall and A. V. Kurpad, 'Relationship between physical activity measured using accelerometers and energy expenditure measured using doubly labelled water in Indian children', *EUROPEAN JOURNAL OF CLINICAL NUTRITION*, vol. 63, no. 11, 1313–1319, NOV 2009, ISSN: 0954-3007. DOI: {10.1038/ejcn.2009.95}.

[39] R. M. Pulsford, M. Cortina-Borja, C. Rich, F.-E. Kinnafick, C. Dezateux and L. J. Griffiths, 'Actigraph Accelerometer-Defined Boundaries for Sedentary Behaviour and Physical Activity Intensities in 7 Year Old Children', *PLOS ONE*, vol. 6, no. 8, AUG 11 2011, ISSN: 1932-6203. DOI: {10.1371/journal.pone.0021822}.

[40] J. M. Neugebauer, D. A. Hawkins and L. Beckett, 'Estimating Youth Locomotion Ground Reaction Forces Using an Accelerometer-Based Activity Monitor', *PLOS ONE*, vol. 7, no. 10, OCT 25 2012, ISSN: 1932-6203. DOI: {10.1371/journal.pone.0048182}.

[41] R. Ojiambo, K. Konstabel, T. Veidebaum, J. Reilly, V. Verbestel, I. Huybrechts, I. Sioen, J. A. Casajus, L. A. Moreno, G. Vicente-Rodriguez, K. Bammann, B. M. Tubic, S. Marild, K. Westerterp, Y. P. Pitsiladis and I. Consortium, 'Validity of hip-mounted uniaxial accelerometry with heart-rate monitoring vs. triaxial accelerometry in the assessment of free-living energy expenditure in young children: the IDEFICS Validation Study', *JOURNAL OF APPLIED PHYSIOLOGY*, vol. 113, no. 10, 1530–1536, NOV 2012, ISSN: 8750-7587. DOI: {10.1152/japplphysiol.01290.2011}.

[42] A. C. Routen, D. Upton, M. G. Edwards and D. M. Peters, 'Discrepancies in accelerometer-measured physical activity in children due to cut-point non-equivalence and placement site', *JOURNAL OF SPORTS SCIENCES*, vol. 30, no. 12, 1303–1310, 2012, ISSN: 0264-0414. DOI: {10.1080/02640414.2012.709266}.

[43] L. R. S. Phillips, G. Parfitt and A. V. Rowlands, 'Calibration of the GENEA accelerometer for assessment of physical activity intensity in children', *JOURNAL OF SCIENCE AND MEDICINE IN SPORT*, vol. 16, no. 2, 124–128, MAR 2013, ISSN: 1440-2440. DOI: {10.1016/j.jsams.2012.05.013}.

[44] S. A. Reading and K. Prickett, 'Evaluation of Children Playing a New-Generation Motion-Sensitive Active Videogame by Accelerometry and Indirect Calorimetry', *GAMES FOR HEALTH JOURNAL*, vol. 2, no. 3, SI, 166–173, JUN 2013, ISSN: 2161-783X. DOI: {10.1089/g4h.2013.0021}.

[45] A. V. Rowlands, K. Rennie, R. Kozarski, R. M. Stanley, R. G. Eston, G. C. Parfitt and T. S. Olds, 'Children's Physical Activity Assessed with Wrist- and Hip-Worn Accelerometers', *MEDICINE AND SCIENCE IN SPORTS AND EXERCISE*, vol. 46, no. 12, 2308–2316, DEC 2014, ISSN: 0195-9131. DOI: {10.1249/MSS.0000000000000365}.

[46] Y. Jang, Y. Song, H. W. Noh and S. Kim, 'A Basic Study of Activity Type Detection and Energy Expenditure Estimation for Children and Youth in Daily Life Using 3-axis Accelerometer and 3-Stage Cascaded Artificial Neural Network', in *2015 37TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC)*, ser. IEEE Engineering in Medi-

cine and Biology Society Conference Proceedings, 37th Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC), Milan, ITALY, AUG 25-29, 2015, IEEE Engn Med & Biol Soc, 2015, 2860–2863, ISBN: 978-1-4244-9270-1.

[47] U. Meyer, D. Ernst, S. Schott, C. Riera, J. Hattendorf, J. Romkes, U. Granacher, B. Goepfert and S. Kriemler, 'Validation of two accelerometers to determine mechanical loading of physical activities in children', *JOURNAL OF SPORTS SCIENCES*, vol. 33, no. 16, 1702–1709, OCT 2 2015, ISSN: 0264-0414. DOI: {10.1080/02640414.2015.1004638}.

[48] J. L. Chandler, K. Brazendale, M. W. Beets and B. A. Mealing, 'Classification of physical activity intensities using a wrist-worn accelerometer in 8-12-year-old children', *PEDIATRIC OBESITY*, vol. 11, no. 2, 120–127, APR 2016, ISSN: 2047-6310. DOI: {10.1111/ijpo.12033}.

[49] X. Ren, W. Ding, S. E. Crouter, Y. Mu and R. Xie, 'Activity recognition and intensity estimation in youth from accelerometer data aided by machine learning', *APPLIED INTELLIGENCE*, vol. 45, no. 2, 512–529, SEP 2016, ISSN: 0924-669X. DOI: {10.1007/s10489-016-0773-3}.

[50] M. C. Bisi, G. P. Panebianco, R. Polman and R. Stagni, 'Objective assessment of movement competence in children using wearable sensors: An instrumented version of the TGMD-2 locomotor subtest', *GAIT & POSTURE*, vol. 56, 42–48, JUL 2017, ISSN: 0966-6362. DOI: {10.1016/j.gaitpost.2017.04.025}.

[51] Y. Gao, M. Melin, K. Makarainen, T. Rantalainen, A. J. Pesola, A. Laukkanen, A. Saakslahti and T. Finni, 'Children's physical activity and sedentary time compared using assessments of accelerometry counts and muscle activity level', *PEERJ*, vol. 6, AUG 21 2018, ISSN: 2167-8359. DOI: {10.7717/peerj.5437}.

[52] Y. Kim and M. Lochbaum, 'Comparison of Polar Active Watch and Waist- and Wrist-Worn ActiGraph Accelerometers for Measuring Children's Physical Activity Levels during Unstructured Afterschool Programs', *INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH*, vol. 15, no. 10, OCT 2018, ISSN: 1660-4601. DOI: {10.3390/ijerph15102268}.

[53] K. Mooses, M. Oja, S. Reisberg, J. Vilo and M. Kull, 'Validating Fitbit Zip for monitoring physical activity of children in school: a cross-sectional study', *BMC PUBLIC HEALTH*, vol. 18, JUL 11 2018, ISSN: 1471-2458. DOI: {10.1186/s12889-018-5752-7}.

[54] S. Kang, Y. Kim, W. Byun, J. Suk and J.-M. Lee, 'Comparison of a Wearable Tracker with Actigraph for Classifying Physical Activity Intensity and Heart Rate in Children', *INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH*, vol. 16, no. 15, AUG 1 2019. DOI: {10.3390/ijerph16152663}.

[55] R. C. Bailey, J. Olson, S. L. Pepper, J. Porszasz, T. J. Barstow, D. M. Cooper *et al.*, 'The level and tempo of children's physical activities: An observational study', *Medicine and science in sports and exercise*, vol. 27, pp. 1033–1033, 1995.

[56] A. Aibar and J. Chanal, 'Physical education: The effect of epoch lengths on children's physical activity in a structured context', *PloS one*, vol. 10, no. 4, e0121238, 2015.

[57] A. International, 'Standard consumer safety specification for toy safety', 2003.

[58] S. Kulak and R. E. Stein, 'Toy age-labeling: An overview for pediatricians of how toys receive their age safety and developmental designations', *Pediatrics*, vol. 138, no. 1, e20151803, 2016.

Appendix A: Systematic search string

((TI=("quality" OR "quantity" OR "variety" OR "characteristics" OR "features" OR divers* OR "variation" OR "amount" OR "capacity" OR "frequency" OR intensit* OR measur* OR determin* OR assess* OR evaluat* OR quantify* OR profil* OR estimat* OR monitor*) AND TI=(movement* OR sport* OR exercis* OR activit* OR "physical activity" OR motion* OR "locomotion" OR “energy expenditure”) AND (TI=("child" OR "children" OR "childhood" OR "schoolchild" OR "schoolchildren" OR "boy" OR "boys" OR "girl" OR "girls" OR "schoolage" OR "schoolboy" OR "schoolboys" OR "schoolgirl" OR "schoolgirls" OR "Pediatrics" OR "Pediatric" OR "Pediatrics" OR "Paediatric" OR "Paediatrics" OR "schoolchild*" OR "pediat*" OR "paediat*" OR "youth" OR "youths" OR prepubert* OR "six year old" OR "seven year old" OR "eight year old" OR "nine year old" OR "ten year old "OR "eleven year old" OR "twelve year old" OR "6 year old" OR "7 year old" OR "8 year old" OR "9 year old" OR "10 year old" OR "11 year old" OR "12 year old" OR "six years old" OR "seven years old" OR "eight years old" OR "nine years old" OR "ten years old" OR "eleven years old" OR "twelve years old" OR "6 years old" OR "7 years old" OR "8 years old" OR "9 years old" OR "10 years old" OR "11 years old" OR "12 years old") NOT AU=Child) AND TI=("sensor" OR "sensors" OR "imu" OR "imus" OR "inertial measurements" OR "inertial measurement" OR "accelerometer" OR "accelerometers" OR acceleromet* OR "gyroscope" OR "gyroscopes" OR gyroscop* OR "magnetometer" OR "magnetometers" OR magnetomet* OR “angular velocity” OR “angular velocities” OR inclino* OR "pedometer" OR "pedometers" OR "Wearable Electronic Device" OR "Wearable Electronic Devices" OR "Fitness Tracker" OR "Fitness Trackers" OR "Wearable" OR "Wearables" OR "Activity Tracker" OR "Activity Trackers" OR "fitbit" OR "fitbits") NOT TI=(Obese OR obesity OR overweight OR Autism OR "cerebral Palsy" OR CF OR "Cystic Fibrosis" OR disease OR diabetes OR diabetic OR questionnaire OR "observational")) OR (WC=(sport) AND TI=("quality" OR "quantity" OR "variety" OR "characteristics" OR "features" OR divers* OR "variation" OR "amount" OR "capacity" OR "frequency" OR intensit* OR measur* OR determin* OR assess* OR evaluat* OR quantify* OR profil* OR estimat* OR monitor*) AND TI=(movement* OR sport* OR exercis* OR activit* OR "physical activity" OR motion* OR "locomotion" OR “energy expenditure”)AND (TI=("child" OR "children" OR "childhood" OR "schoolchild" OR "schoolchildren" OR "boy" OR "boys" OR "girl" OR "girls" OR "schoolage" OR "schoolboy" OR "schoolboys" OR "schoolgirl" OR "schoolgirls" OR "Pediatrics" OR "Pediatric" OR "Pediatrics" OR "Paediatric" OR "Paediatrics" OR "schoolchild*" OR "pediat*" OR "paediat*" OR "youth" OR "youths" OR prepubert* OR "six year old" OR "seven year old" OR "eight year old" OR "nine year old" OR "ten year old "OR "eleven year old" OR "twelve year old" OR "6 year old" OR "7 year old" OR "8 year old" OR "9 year old" OR "10 year old" OR "11 year old" OR "12 year old" OR "six years old" OR "seven years old" OR "eight years old" OR "nine years old" OR "ten years old" OR "eleven years old" OR "twelve years old" OR "6 years old" OR "7 years old" OR "8 years old" OR "9 years old" OR "10 years old" OR "11 years old" OR "12 years old") NOT AU=Child) AND TS=(("sensor" OR "sensors" OR "imu" OR "imus" OR "inertial measurements" OR "inertial measurement" OR "accelerometer" OR "accelerometers" OR acceleromet* OR "gyroscope" OR "gyroscopes" OR gyroscop* OR "magnetometer" OR "magnetometers" OR magnetomet* OR “angular velocity” OR “angular velocities” OR inclino* OR "pedometer" OR "pedometers" OR "Wearable Electronic Device" OR "Wearable Electronic Devices" OR "Fitness Tracker" OR "Fitness Trackers" OR "Wearable" OR "Wearables" OR "Activity Tracker" OR "Activity Trackers" OR "fitbit" OR "fitbits") NEAR/5(measur* OR method*)) AND TI=(characteristic* OR assess* OR measur*) NOT TI=(Obese OR obesity OR overweight OR Autism OR "cerebral Palsy" OR CF OR "Cystic Fibrosis" OR disease OR diabetes OR diabetic OR questionnaire OR "observational"))) AND la=english

B
Scripts

B.1. Reading data

```
% Karen Rijnders
% Master Thesis Biomedical Engineering - Sports Engineering
% February - September 2020
close all

%% Manually Load Data
saveName = 'IMU_P...'; %Name file after participant

%% Get Data of Left Wrist
% Left wrist Uncalibrated Low Noise Accelerometer data
LW_acc_x_ln = leftWrist_Accel_LN_X_UNCAL;
LW_acc_y_ln = leftWrist_Accel_LN_Y_UNCAL;
LW_acc_z_ln = leftWrist_Accel_LN_Z_UNCAL;

% Left wrist Uncalibrated Wide Range Accelerometer data
LW_acc_x_wr = leftWrist_Accel_WR_X_UNCAL;
LW_acc_y_wr = leftWrist_Accel_WR_Y_UNCAL;
LW_acc_z_wr = leftWrist_Accel_WR_Z_UNCAL;

% Left wrist Uncalibrated Gyroscope data
LW_gyr_x = leftWrist_Gyro_X_UNCAL;
LW_gyr_y = leftWrist_Gyro_Y_UNCAL;
LW_gyr_z = leftWrist_Gyro_Z_UNCAL;

% Left wrist Uncalibrated Magnetometer data
LW_mag_x = leftWrist_Mag_X_UNCAL;
LW_mag_y = leftWrist_Mag_Y_UNCAL;
LW_mag_z = leftWrist_Mag_Z_UNCAL;

% Left Wrist rtc Data
LW_time = leftWrist_Timestamp_Unix_CAL;

%% Get Data of Right Wrist
% Right wrist Uncalibrated Low Noise Accelerometer data
RW_acc_x_ln = rightWrist_Accel_LN_X_UNCAL;
RW_acc_y_ln = rightWrist_Accel_LN_Y_UNCAL;
RW_acc_z_ln = rightWrist_Accel_LN_Z_UNCAL;

% Right wrist Uncalibrated Wide Range Accelerometer data
RW_acc_x_wr = rightWrist_Accel_WR_X_UNCAL;
RW_acc_y_wr = rightWrist_Accel_WR_Y_UNCAL;
RW_acc_z_wr = rightWrist_Accel_WR_Z_UNCAL;

% Right wrist Uncalibrated Gyroscope data
RW_gyr_x = rightWrist_Gyro_X_UNCAL;
```

```
RW_gyr_y = rightWrist_Gyro_Y_UNCAL;
RW_gyr_z = rightWrist_Gyro_Z_UNCAL;

% Right wrist Uncalibrated Magnetometer data
RW_mag_x = rightWrist_Mag_X_UNCAL;
RW_mag_y = rightWrist_Mag_Y_UNCAL;
RW_mag_z = rightWrist_Mag_Z_UNCAL;

% Left Wrist rtc Data
RW_time = rightWrist_Timestamp_Unix_CAL;

%% Get Data of Hip
% Hip Uncalibrated Low Noise Accelerometer data
H_acc_x_ln = Hip_Accel_LN_X_UNCAL;
H_acc_y_ln = Hip_Accel_LN_Y_UNCAL;
H_acc_z_ln = Hip_Accel_LN_Z_UNCAL;

% Hip Uncalibrated Wide Range Accelerometer data
H_acc_x_wr = Hip_Accel_WR_X_UNCAL;
H_acc_y_wr = Hip_Accel_WR_Y_UNCAL;
H_acc_z_wr = Hip_Accel_WR_Z_UNCAL;

% Hip Uncalibrated Gyroscope data
H_gyr_x = Hip_Gyro_X_UNCAL;
H_gyr_y = Hip_Gyro_Y_UNCAL;
H_gyr_z = Hip_Gyro_Z_UNCAL;

% Hip Uncalibrated Magnetometer data
H_mag_x = Hip_Mag_X_UNCAL;
H_mag_y = Hip_Mag_Y_UNCAL;
H_mag_z = Hip_Mag_Z_UNCAL;

% Hip rtc Data
H_time = Hip_Timestamp_Unix_CAL;

%% Get Data of Right or Left Ankle

if exist(rightAnkle_Accel_LN_X_UNCAL) == 1
    % Right ankle Uncalibrated Low Noise Accelerometer data
    RA_acc_x_ln = rightAnkle_Accel_LN_X_UNCAL;
    RA_acc_y_ln = rightAnkle_Accel_LN_Y_UNCAL;
    RA_acc_z_ln = rightAnkle_Accel_LN_Z_UNCAL;

    % Right ankle Uncalibrated Wide Range Accelerometer data
    RA_acc_x_wr = rightAnkle_Accel_WR_X_UNCAL;
    RA_acc_y_wr = rightAnkle_Accel_WR_Y_UNCAL;
    RA_acc_z_wr = rightAnkle_Accel_WR_Z_UNCAL;

    % Right ankle Uncalibrated Gyroscope data
    RA_gyr_x = rightAnkle_Gyro_X_UNCAL;
    RA_gyr_y = rightAnkle_Gyro_Y_UNCAL;
    RA_gyr_z = rightAnkle_Gyro_Z_UNCAL;

    % Right ankle Uncalibrated Magnetometer data
    RA_mag_x = rightAnkle_Mag_X_UNCAL;
    RA_mag_y = rightAnkle_Mag_Y_UNCAL;
    RA_mag_z = rightAnkle_Mag_Z_UNCAL;
```

```
    % Right ankle rtc Data
    RA_time = rightAnkle_Timestamp_Unix_CAL;

elseif exist(leftAnkle_Accel_LN_X_UNCAL) == 1
    % Left ankle Uncalibrated Low Noise Accelerometer data
    LA_acc_x_ln = leftAnkle_Accel_LN_X_UNCAL;
    LA_acc_y_ln = leftAnkle_Accel_LN_Y_UNCAL;
    LA_acc_z_ln = leftAnkle_Accel_LN_Z_UNCAL;

    % Left ankle Uncalibrated Wide Range Accelerometer data
    LA_acc_x_wr = leftAnkle_Accel_WR_X_UNCAL;
    LA_acc_y_wr = leftAnkle_Accel_WR_Y_UNCAL;
    LA_acc_z_wr = leftAnkle_Accel_WR_Z_UNCAL;

    % Left ankle Uncalibrated Gyroscope data
    LA_gyr_x = leftAnkle_Gyro_X_UNCAL;
    LA_gyr_y = leftAnkle_Gyro_Y_UNCAL;
    LA_gyr_z = leftAnkle_Gyro_Z_UNCAL;

    % Left ankle Uncalibrated Magnetometer data
    LA_mag_x = leftAnkle_Mag_X_UNCAL;
    LA_mag_y = leftAnkle_Mag_Y_UNCAL;
    LA_mag_z = leftAnkle_Mag_Z_UNCAL;

    % Left ankle rtc Data
    LA_time = leftAnkle_Timestamp_Unix_CAL;
end

%% Calculate Real rtc for each sensor in Seconds
% rtc = datestr([1]*double(a)./86400+datenum(1970,1,1,0,0,0),'yyyymmddTHH:MM:SS
   ')
% rtc = unixtime(double(a))
LW_rtc = datetime(uint64(LW_time/1000),'ConvertFrom','posixtime','TimeZone','
   Europe/Amsterdam','Format','dd-MMM-yyyy HH:mm:ss');

RW_rtc = datetime(uint64(RW_time/1000),'ConvertFrom','posixtime','TimeZone','
   Europe/Amsterdam','Format','dd-MMM-yyyy HH:mm:ss');

H_rtc = datetime(uint64(H_time/1000),'ConvertFrom','posixtime','TimeZone','
   Europe/Amsterdam','Format','dd-MMM-yyyy HH:mm:ss');

if exist(RA_time) == 1
    RA_rtc = datetime(uint64(RA_time/1000),'ConvertFrom','posixtime','TimeZone'
       ,'Europe/Amsterdam','Format','dd-MMM-yyyy HH:mm:ss');
elseif exist(LA_time) == 1
    LA_rtc = datetime(uint64(RA_time/1000),'ConvertFrom','posixtime','TimeZone'
       ,'Europe/Amsterdam','Format','dd-MMM-yyyy HH:mm:ss');
end

%% Save everything in a structure
% Left wrist data
IMU.LW.Acc_LN = [LW_acc_x_ln, LW_acc_y_ln, LW_acc_z_ln];
IMU.LW.Acc_WR = [LW_acc_x_wr, LW_acc_y_wr, LW_acc_z_wr];
IMU.LW.Gyr = [LW_gyr_x, LW_gyr_y, LW_gyr_z];
IMU.LW.Mag = [LW_mag_x, LW_mag_y, LW_mag_z];
IMU.LW.rtc = LW_rtc;
```

```
IMU.LW.TimeStamp = LW_time;

% Right wrist data
IMU.RW.Acc_LN = [RW_acc_x_ln, RW_acc_y_ln, RW_acc_z_ln];
IMU.RW.Acc_WR = [RW_acc_x_wr, RW_acc_y_wr, RW_acc_z_wr];
IMU.RW.Gyr = [RW_gyr_x, RW_gyr_y, RW_gyr_z];
IMU.RW.Mag = [RW_mag_x, RW_mag_y, RW_mag_z];
IMU.RW.rtc = RW_rtc;
IMU.RW.TimeStamp = RW_time;

% Hip data
IMU.H.Acc_LN = [H_acc_x_ln, H_acc_y_ln, H_acc_z_ln];
IMU.H.Acc_WR = [H_acc_x_wr, H_acc_y_wr, H_acc_z_wr];
IMU.H.Gyr = [H_gyr_x, H_gyr_y, H_gyr_z];
IMU.H.Mag = [H_mag_x, H_mag_y, H_mag_z];
IMU.H.rtc = H_rtc;
IMU.H.TimeStamp = H_time;

if exist(RA_rtc) == 1
    % Right ankle data
    IMU.RA.Acc_LN = [RA_acc_x_ln, RA_acc_y_ln, RA_acc_z_ln];
    IMU.RA.Acc_WR = [RA_acc_x_wr, RA_acc_y_wr, RA_acc_z_wr];
    IMU.RA.Gyr = [RA_gyr_x, RA_gyr_y, RA_gyr_z];
    IMU.RA.Mag = [RA_mag_x, RA_mag_y, RA_mag_z];
    IMU.RA.rtc = RA_rtc;
    IMU.RA.TimeStamp = RA_time;

elseif exist(LA_rtc) == 1
    % Left ankle data
    IMU.LA.Acc_LN = [LA_acc_x_ln, LA_acc_y_ln, LA_acc_z_ln];
    IMU.LA.Acc_WR = [LA_acc_x_wr, LA_acc_y_wr, LA_acc_z_wr];
    IMU.LA.Gyr = [LA_gyr_x, LA_gyr_y, LA_gyr_z];
    IMU.LA.Mag = [LA_mag_x, LA_mag_y, LA_mag_z];
    IMU.LA.rtc = LA_rtc;
    IMU.LA.TimeStamp = LA_time;
end

%% Save data
save(saveName);
clear all
```

B.2. Segmenting data

```
%% Raw data per activity can be found in Act.RAW, Epoch data per activity can
    be found in Act.EPOCH
% Karen Rijnders
% Master Thesis Biomedical Engineering - Sports Engineering
% February - September 2020
clear all

%% Segmenting data based on real time
for w = 1:8 % Loop over participants

    IMU_data = sprintf('IMU_P00%d', w);
    Time_data = sprintf('times_P00%d',w);

    load(IMU_data)
    load(Time_data)

    % Create names for placements, activities and sensortypes
    placeName = fieldnames(IMU)';
    sensorType = ["Acc_LN", "Acc_WR", "Gyr", "Mag", "rtc"];
    actName = string(who('-file', Time_data))';
    Tries = ["First", "Second", "Third", "Fourth", "Fifth"];
    Epochs = ["OneSecond", "HalfSecond", "ThirdSecond", "FourthSecond"]; % "
        TwoSecond" can be added, also add in loop.

    % Define start and stop of individual activities
    for r = 1: size(actName,2)
        act(r) = {eval(actName(r))};

        actStart(r) = {act{1,r}(:,1)};
        actEnd(r) = {act{1,r}(:,2)};
    end
    % Find indices of relevant sensor values
    % Initiate empty vector
    actInd = [];

    if isfield(IMU,'LA') ==1
        RTC = {LW_rtc, RW_rtc, H_rtc, LA_rtc};
    end

    if isfield(IMU,'RA') ==1
        RTC = {LW_rtc, RW_rtc, H_rtc, RA_rtc};
    end

    for i=1:(length(actName)) % find indices of start en end of activities per
        sensor
        for k = 1:size(act{:,i},1)
            for j = 1:length(RTC)
                Start = min(find(RTC{j} == actStart{i}(k))); % Start time
                Stop = max(find(RTC{j} == actEnd{i}(k))); % End time

                actInd{k,i,j} = [Start ; Stop];
            end
        end
    end

    % Get Raw data per activity
```

```
for ii = 1: size(actInd,2)                        % loop over activities
    fieldname = actName(ii);
    for hh = 1:size(act{:,ii},1)
        tries = Tries(hh);
        for jj = 1: size(actInd,3)                    % loop over placements
            placement = placeName{jj};
            for kk = 1: length(sensorType)            % loop over sensortypes
                sensorname = sensorType{kk};

                Act.RAW.(fieldname).(tries).(placement).(sensorname) = IMU
                    .(placement).(sensorname)((actInd{hh,ii,jj}(1):actInd{hh
                    ,ii,jj}(2)),:);

            end
        end
    end
end

% 1, 0.5, 0.33 and 0.25 s epochs
for ii = 1: size(actName,2)                        % loop over activities
    fieldname = actName(ii);
    for hh = 1:size(act{:,ii},1)
        tries = Tries(hh);
        for jj = 1: size(actInd,3)                    % loop over placements
            placement = placeName{jj};
            for kk = 1: length(sensorType)            % loop over sensortypes
                sensorname = sensorType{kk};

                RTC =Act.RAW.(fieldname).(tries).(placement).rtc;  % Find
                    at which time a certain activity occured for a specific
                    wearsite

                % Make matrix with first column is 1 second epoch etc.
                groupEpoch = [];
                groupEpoch(:,1) = findgroups(RTC);                    % 1 s
                     epoch: Group all data that was measured within the same
                     second
                groupEpoch = [groupEpoch(:,1), zeros(size(groupEpoch,1),3)
                    ];

                totalTime = max(groupEpoch(:,1));
                for rr = 1: totalTime
                    indices = find(groupEpoch(:,1) == rr);

                    % 0.5 s epoch
                    one = min(indices):floor((max(indices)+min(indices))/2)
                        ;
                    two = ceil((max(indices)+min(indices))/2):max(indices);
                    groupEpoch(one,2) = (groupEpoch(one,1)*2)-1;
                    groupEpoch(two,2) = groupEpoch(two,1)*2;

                    % 0.33 s epoch
                    one = min(indices):floor((max(indices)+2*min(indices))
                        /3);
                    two = ceil((max(indices)+2*min(indices))/3): floor((2*
                        max(indices)+min(indices))/3);
                    three = ceil((2*max(indices)+min(indices))/3):max(
```

```
                    indices);
                groupEpoch(one,3) = (groupEpoch(one,1)*3)-2;
                groupEpoch(two,3) = (groupEpoch(two,1)*3)-1;
                groupEpoch(three,3) = groupEpoch(three,1)*3;

                % 0.25 s epoch
                one = min(indices):floor((max(indices)+3*min(indices))
                    /4);
                two = ceil((max(indices)+3*min(indices))/4): floor((2*
                    max(indices)+2*min(indices))/4);
                three = ceil((2*max(indices)+2*min(indices))/4):floor
                    ((3*max(indices)+min(indices))/4);
                four = ceil((3*max(indices)+min(indices))/4):max(
                    indices);
                groupEpoch(one,4) = (groupEpoch(one,1)*4)-3;
                groupEpoch(two,4) = (groupEpoch(two,1)*4)-2;
                groupEpoch(three,4) = (groupEpoch(three,1)*4)-1;
                groupEpoch(four,4) = groupEpoch(four,1)*4;
            end

            for n = 1:size(groupEpoch,2)
                epochlength = Epochs(n);
                Act.EPOCH.(epochlength).(fieldname).(tries).(placement)
                    .(sensorname) = ...
                    splitapply(@(x)mean(x,1), Act.RAW.(fieldname).(
                        tries).(placement).(sensorname), groupEpoch(:,n)
                        );
            end
        end
      end
    end
  end

  % Save Data
  Name_Data= sprintf('Data_P00%d', w);
  save(Name_Data, 'Act')

  clear all
end
```

B.3. Preparing data

```
% Karen Rijnders
% Master Thesis Biomedical Engineering - Sports Engineering
% February - September 2020
clear all

%% Preparing Epoch data
for z = 1:8 % Loop over participants
    for w = 1:4 % Loop over epochs
        dataset = sprintf('Data_P00%d', z);
        load(dataset)
        epochName = string(fieldnames(Act.EPOCH));

        epochname = epochName(w);
        sensorType = ["Acc_LN", "Acc_WR", "Gyr"];
        Tries = string(fieldnames(Act.EPOCH.(epochname).kick_3))';
        actName = string(fieldnames(Act.EPOCH.(epochname)))';

        for i = 1: size(actName,2)                      % loop over activities
            fieldname = actName(i);

            for j = 1:length(fieldnames(Act.EPOCH.(epochname).(fieldname)))
                                      % loop over trials
                tries = Tries(j);

                LW_AccLN{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).LW.
                    Acc_LN;
                LW_AccLN{i,2} = actName(i);

                RW_AccLN{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).RW.
                    Acc_LN;
                RW_AccLN{i,2} = actName(i);

                H_AccLN{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).H.
                    Acc_LN;
                H_AccLN{i,2} = actName(i);

                if isfield(Act.EPOCH.(epochname).(fieldname).(tries), 'RA') ==1
                    RA_AccLN{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).
                        RA.Acc_LN;
                    RA_AccLN{i,2} = actName(i);
                end

                if isfield(Act.EPOCH.(epochname).(fieldname).(tries),'LA') ==1
                    LA_AccLN{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).
                        LA.Acc_LN;
                    LA_AccLN{i,2} = actName(i);
                end

                LW_AccWR{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).LW.
                    Acc_WR;
                LW_AccWR{i,2} = actName(i);

                RW_AccWR{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).RW.
                    Acc_WR;
                RW_AccWR{i,2} = actName(i);
```

```
        H_AccWR{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).H.
            Acc_WR;
        H_AccWR{i,2} = actName(i);

        if isfield(Act.EPOCH.(epochname).(fieldname).(tries), 'RA') ==1
            RA_AccWR{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).
                RA.Acc_WR;
            RA_AccWR{i,2} = actName(i);
        end

        if isfield(Act.EPOCH.(epochname).(fieldname).(tries),'LA') ==1
            LA_AccWR{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).
                LA.Acc_WR;
            LA_AccWR{i,2} = actName(i);
        end

        LW_Gyr{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).LW.Gyr;
        LW_Gyr{i,2} = actName(i);

        RW_Gyr{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).RW.Gyr;
        RW_Gyr{i,2} = actName(i);

        H_Gyr{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).H.Gyr;
        H_Gyr{i,2} = actName(i);

        if isfield(Act.EPOCH.(epochname).(fieldname).(tries), 'RA') ==1
            RA_Gyr{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).RA.
                Gyr;
            RA_Gyr{i,2} = actName(i);
        end

        if isfield(Act.EPOCH.(epochname).(fieldname).(tries),'LA') ==1
            LA_Gyr{i,1} = Act.EPOCH.(epochname).(fieldname).(tries).LA.
                Gyr;
            LA_Gyr{i,2} = actName(i);
        end
    end
end

LeftWrist = sprintf('LW_P00%d',z);
RightWrist = sprintf('RW_P00%d',z);
Hip = sprintf('H_P00%d',z);
LeftAnkle = sprintf('LA_P00%d',z);
RightAnkle = sprintf('RA_P00%d',z);

save(fullfile('...', epochname, LeftWrist), 'LW_AccWR', 'LW_AccLN', '
    LW_Gyr')
save(fullfile('...', epochname, RightWrist), 'RW_AccWR', 'RW_AccLN', '
    RW_Gyr')
save(fullfile('...', epochname, Hip), 'H_AccWR', 'H_AccLN', 'H_Gyr')

if exist('LA_AccWR','var') ==1
    save(fullfile('...', epochname, LeftAnkle), 'LA_AccWR', 'LA_AccLN',
        'LA_Gyr')
end

if exist('RA_AccWR','var') ==1
```

```
            save(fullfile('...', epochname, RightAnkle), 'RA_AccWR', 'RA_AccLN'
                , 'RA_Gyr')
        end
    end

    clear all
end
```

B.4. Reshape, Randomize, Train and Test data with overlap through bootstrapping

```
% Karen Rijnders
% Master Thesis Biomedical Engineering - Sports Engineering
% February - September 2020
clear all

%% Reshape, Randomize, Train and Test Data
for w = 1:4 % Loop over all epochs
    D = dir('...'); % Add direction of folders names after epoch length
    epochName = D(w).name;

    namePlace = {'H', 'LA', 'LW', 'RA', 'RW'};
    nameSensor = {'LN', 'WR', 'Gyr'};
    fullnamePlace = {'Hip', 'LeftAnkle', 'LeftWrist', 'RightAnkle', 'RightWrist
       '};
    numParts = 8; % Number of participants

    for i = 1:size(namePlace,2) % Loop over all placements
        for s = 1: size(nameSensor,2) % Loop over the sensor types
            v = 1;
            for j = 1:numParts % Loop over eight participants
                placeFolder = fullnamePlace{i};
                if exist(fullfile('...')) == 2 % File containing data from all
                   sensor types for 1 placement and 1 participant
                    load(fullfile('...')) % This same file

                    Acc = sprintf('%s_%s', namePlace{i},nameSensor{s});
                    Acc = eval(Acc); % Fill in values from file in variable Acc
                    for q = 1: size(Acc,1)
                        Acc{q,3} = LabelActSequences(Acc{q,2}); % Give
                           categorical activity labels to the sensor data
                    end

                    % Merge data from same activity but at different
                    % intensity
                    m=1;
                    Acc_new = [0, {"str"}];
                    for q = 1: size(Acc,1)
                        index = find(string(Acc(:,3))== string(Acc{q,3}));
                        if isempty(find(string(Acc_new(:,2)) == string(Acc{q
                           ,3})))
                            Acc_new{m,1} = cat(1,Acc{index,1});
                            Acc_new{m,2} = Acc{q,3};
                            m=m+1;
                        end
                    end
                    classes = categories(categorical(string(Acc_new(:,2))));
                    numClasses = numel(classes);

                    % Sort in Alphabetical order
                    [~, ix] = sort(string(Acc_new(:,2)));
                    Acc_new = Acc_new(ix,:);

                    % Make non alphabetical order for all types and epochs,
                    % P003 did not have rope skipping data (10 activities
                    % total in stead of 11)
                    if w == 1 && j==1 && s ==1 && i==1
```

```
                ix11 = randperm(11);
            elseif w == 1 && j==3 && s ==1 && i==1
                ind11 = find(ix11 == 11);
                ix10 = ix11;
                ix10(ind11) = [];
                ix10 = squeeze(ix10);
            end

            if j == 3
                Acc_new = Acc_new(ix10,:);
            else
                Acc_new = Acc_new(ix11,:);
            end

            % Devide into groups of 4 activities, with one
            % activity also present in the next group
            numActs = 4;
            numGroups = ceil((size(Acc_new,1)+numActs)/numActs);
            for b = 1:13
                for r = 1:numGroups
                    k = r*numActs;
                    if r ==1
                        temp = Acc_new(k-numActs+1:k,:);
                    elseif r ==2
                        temp = Acc_new(k-numActs:k-1,:);
                    elseif r==3
                        temp = Acc_new(k-numActs-1:k-2,:);
                    else
                        temp = [Acc_new(k-numActs-2:end,:); Acc_new
                            (1,:)];
                    end
                    randIdx = randperm(size(temp,1));

                    Seqs{r,1} = temp(randIdx,:);
                end

                % Make the labels into sequences of this label,
                % matching the length of the sequences of sensor
                % data
                for r = 1:size(Seqs,1)
                    m=0;
                    for c = 1: size(Seqs{r,1},1)
                        num = size(Seqs{r,1}{c,1},1);
                        k = j*numGroups;
                        X{v,b}(1:3, m+1:m+num) = Seqs{r,1}{c,1}';
                        Y{v,b}(1, m+1:m+num) = repmat(categorical(Seqs{
                            r,1}{c,2}), [num , 1]);
                        m = m+num;
                    end
                    v = v+1;
                end
            end
        end
    end

    % Remove empty cells
    index = cellfun(@isempty, X) == 0;
```

```
X = X(index);
Y = Y(index);

% Save data
save(fullfile('...'), 'X', 'Y', 'b', 'numGroups', 'numActs', '
    classes') % Save the Reshaped & Randomized data with overlap
clear X
clear Y

% Make scalograms from continuous wavelet transforms and get
    activations
if exist('...') == 0 % If the file with Googlenet activations do
    not yet exist, compute googlenet activations

    if epochName == FourthSecond
        windowLength = 40;
        windowStep = 10;
    elseif epochName == ThirdSecond
        windowLength = 30;
        windowStep = 8;
    elseif epochName == HalfSecond
        windowLength = 20;
        windowStep = 5;
    elseif epochName == OneSecond
        windowLength = 10;
        windowStep = 3;
    end

    % Separate data in 10 sconds window with 7-8 second overlap
    [X_new, Y_new] = SlidingWindow_FT(X, Y, windowLength,
       windowStep);
    [numSeqs, numFrames] = size(X_new);

    % Load google net and compute its activations on the
    % scalogram data
    netCNN = googlenet;
    poolLayer = "pool5-7x7_s1";

    for k = 1:numSeqs % Loop over participants
        for j = 1:numFrames
            if isempty(X_new{k,j}) ==0
                [feats, labs] = CreateCFS_SW_3D(X_new{k,j}, Y_new{k
                    ,j});
                Labels{k,1}(:,j) = labs;
                sequences{k,1}(:,j) = activations(netCNN,feats,
                    poolLayer ,'OutputAs', 'columns');
            end
        end
    end

    % Save file with googlenet activations
    save('...', 'Labels', 'sequences')

else
    % If file with googlenet activations already exists, load
    % this file
    load('...')
```

```
end

% Bootstrap data to create all combinations of Train and Test
% data with ratio 7:1
for p = 1:numParts
    if exist('...') == 2 % if file containing the accuracies
       already exists, skip the rest of the loop
        continue
    else

        % Get indices with data per participant
        temp = [p:numParts:numParts*b];
        iEnd = numGroups*temp;
        iStart = iEnd - numGroups+1;
        i = [];
        for k = 1:b
            i = [i, iStart(k):iEnd(k)];
        end

        % Divide over Train and Test dataset
        sequencesTest = sequences(i,1);
        sequencesTrain = sequences(setdiff(1:end,i),1);
        LabelsTest = Labels(i,1);
        LabelsTrain = Labels(setdiff(1:end,i),1);

        numClasses = numel(classes);
        numFeatures = size(sequencesTrain{1},1);
        miniBatchSize = numParts-1;

        % Set up neural networ layers
        layers = [
            sequenceInputLayer(numFeatures,'Name', 'input')
            bilstmLayer(500,'OutputMode','sequence','Name','
               bilstm_1')
            dropoutLayer(0.5,'Name','drop')
            fullyConnectedLayer(numClasses,'Name','fc')
            softmaxLayer('Name','softmax')
            classificationLayer('Name','classification')];

        options = trainingOptions('adam', ...
            'MiniBatchSize',miniBatchSize, ...
            'InitialLearnRate',1e-4, ...
            'MaxEpochs', 20, ...
            'ExecutionEnvironment', 'cpu', ...
            'GradientThreshold',2, ...
            'Plots','training-progress', ...
            'Shuffle', 'every-epoch',...
            'Verbose',false);

        % Train network
        [netLSTM,info] = trainNetwork(sequencesTrain, LabelsTrain,
           layers,options);
        clear sequencesTrain LabelsTrain

        % Validate network using test datasets
        for k = 1:size(sequencesTest,1)
            [YPred{i(k),1}, scores{i(k),1}] = classify(netLSTM,
```

```
                                sequencesTest{k,1}, 'MiniBatchSize', miniBatchSize);
                            accuracy(k) = mean(YPred{i(k)} == LabelsTest{k,1});
                        end

                        % Save file containing accuracies for one participant with
                        % one sensor type at one placement.
                        save('...', 'YPred', 'scores', 'accuracy', 'LabelsTest', '
                            sequencesTest');
                    end
                end
            end
            clear all
        end
    end
```

B.5. Support functions

B.5.1. CreateCFS_SW_3D

```
function [feats, lab] = CreateCFS_SW_3D(Signals, Labels)

data = Signals;
labels = string(Labels);

% Separate triaxial data into three individual axes
datax = data(1,:);
datay = data(2,:);
dataz = data(3,:);

% Compute continuous wavelet transform and concatenate the data of
% the three axes
signalLength = size(datax,2);
fbx = cwtfilterbank('SignalLength',signalLength,'VoicesPerOctave',12);
fby = cwtfilterbank('SignalLength',signalLength,'VoicesPerOctave',12);
fbz = cwtfilterbank('SignalLength',signalLength,'VoicesPerOctave',12);

cfsx = abs(fbx.wt(datax));
cfsy = abs(fby.wt(datay));
cfsz = abs(fbz.wt(dataz));
cfs = cat(1, cfsx, cfsy, cfsz);

% Make scalogram image of wavelet transform
im = ind2rgb(im2uint8(rescale(cfs)), jet)*255;

% Use mode of activities present in one scalogram (10 seconds of data) to
% label the window
[classes,~,map] = unique(labels);
modeClass = classes(mode(map));

% Make sure image fits the Googlenet input layer requirements
feats = imresize(im,[224 224]);
lab = categorical(modeClass);
end
```

B.5.2. LabelActSequences

```
function Label = LabelActSequences(var)
var = string(var);

    if startsWith(var, "box") == 1
        Label = categorical("box");

    elseif startsWith(var, "walk") == 1
        Label = categorical("walk");

    elseif startsWith(var, "run") == 1
        Label = categorical("run");

    elseif startsWith(var, "kick") == 1
        Label = categorical("kick");

    elseif startsWith(var, "throw") == 1
        Label = categorical("throw");
```

```
    elseif startsWith(var, "ropejump") == 1
        Label = categorical("ropejump");

    elseif startsWith(var, "jump") == 1
        Label = categorical("jump");

    elseif startsWith(var, "cycle") == 1
        Label = categorical("cycle");

    elseif startsWith(var, "lie") == 1
        Label = categorical("lie");

    elseif startsWith(var, "sit") == 1
        Label = categorical("sit");

    elseif startsWith(var, "write") == 1
        Label = categorical("write");
    end
end
```

B.5.3. SlidingWindow_FT

```
function [signalsOut, labelsOut] = SlidingWindow_FT(signal, labels,
   windowLength, windowStep)
signalsOut = {};
labelsOut = {};
j = 1;

for i = 1: size(signal,1)
    records = size(signal{i,1},2);
    windowMax = floor(records/windowStep)*windowStep;

    % Begin at first index of data, remainders that do not complete a
    % window are removed
    for k = 0:windowStep:windowMax
        if k+windowLength < windowMax
        signalsOut{i, j} = signal{i,1}(:,(k+1):(k+windowLength));
        labelsOut{i, j} = labels{i,1}(1,(k+1):(k+windowLength));
        j = j+1;
        end
    end
    j = 1;

end
end
```

C
Confusion matrices

C.1. 1 Second Epoch

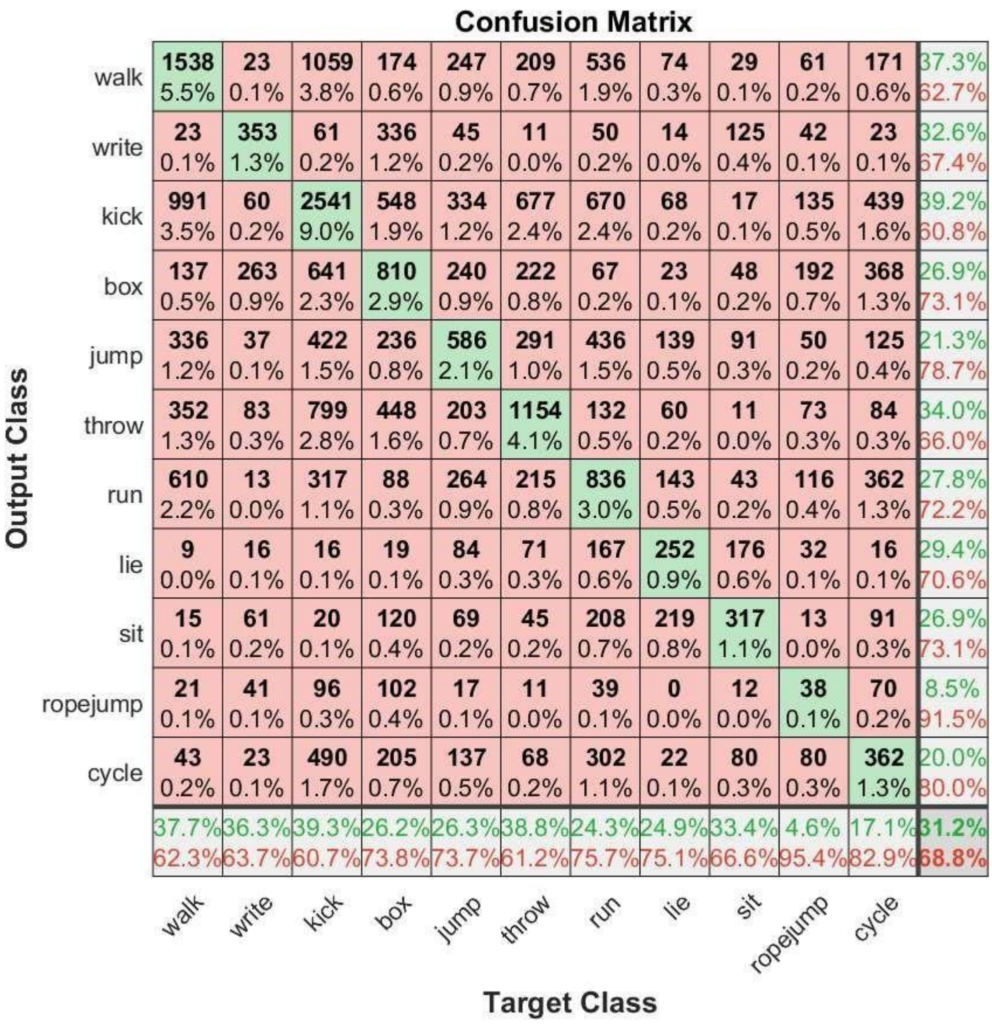


Figure C.1: 1 second epoch, LN accelerometer, Hip

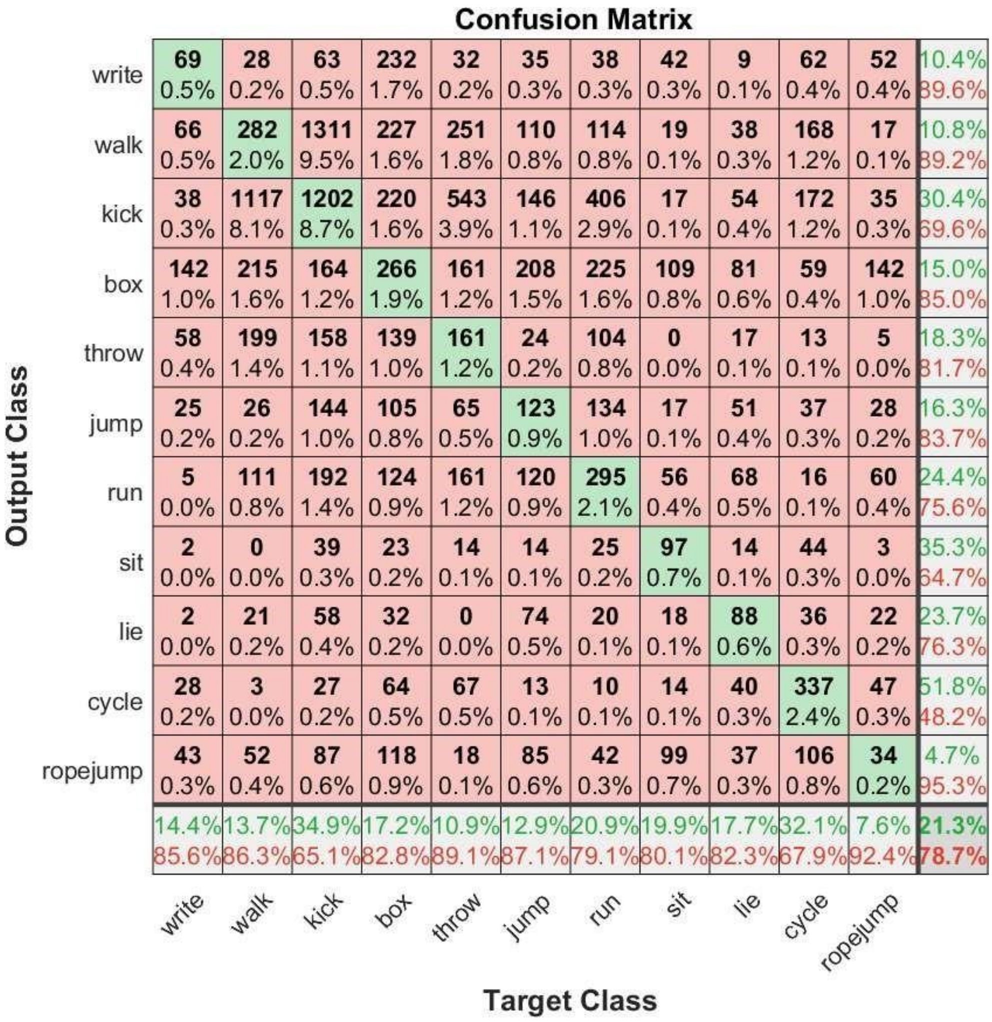


Figure C.2: 1 second epoch, LN accelerometer, Left ankle

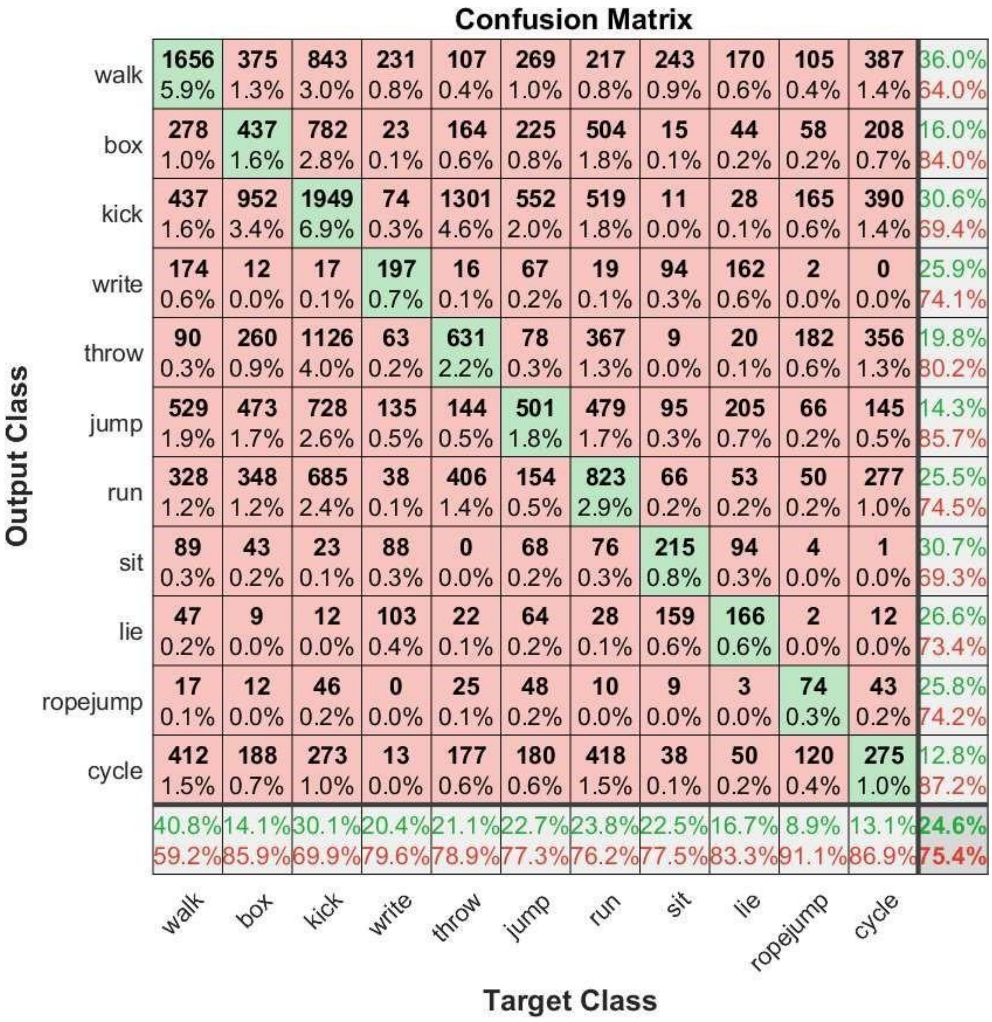


Figure C.3: 1 second epoch, LN accelerometer, Left wrist

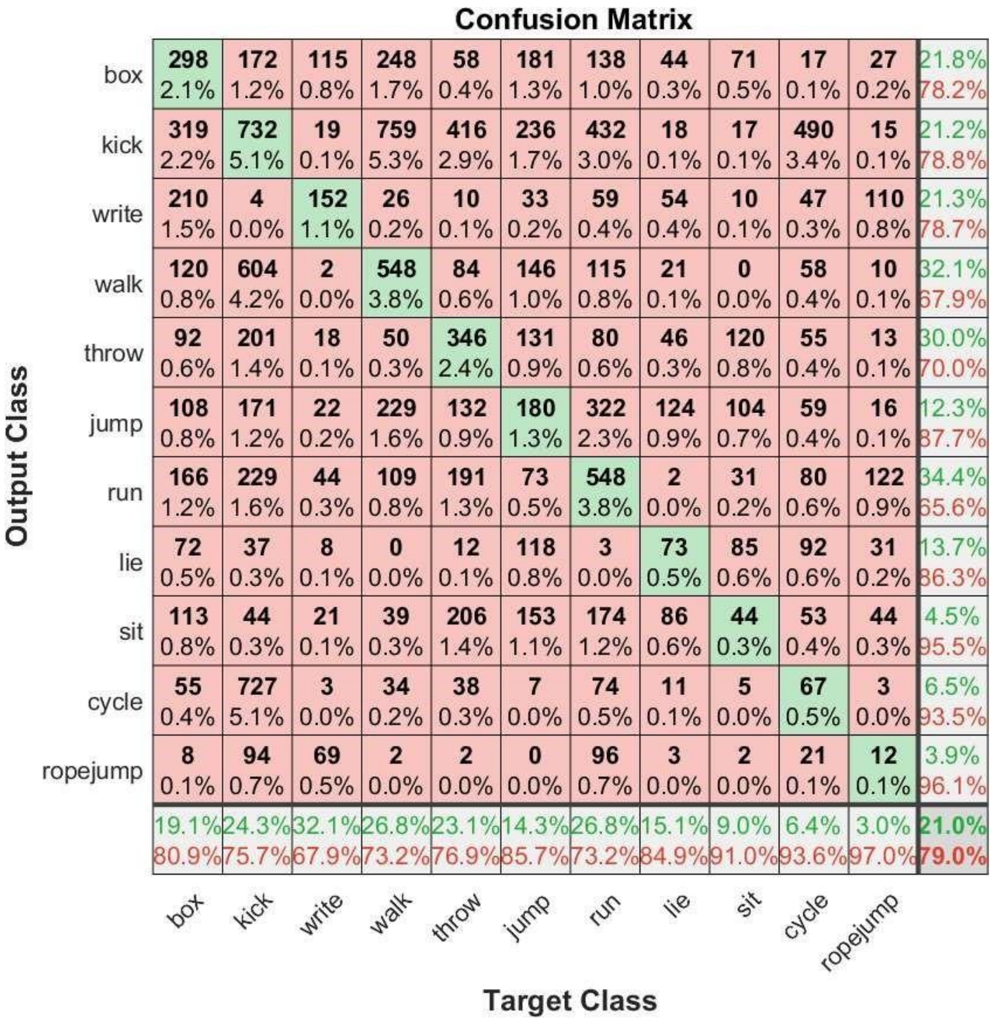


Figure C.4: 1 second epoch, LN accelerometer, Right ankle

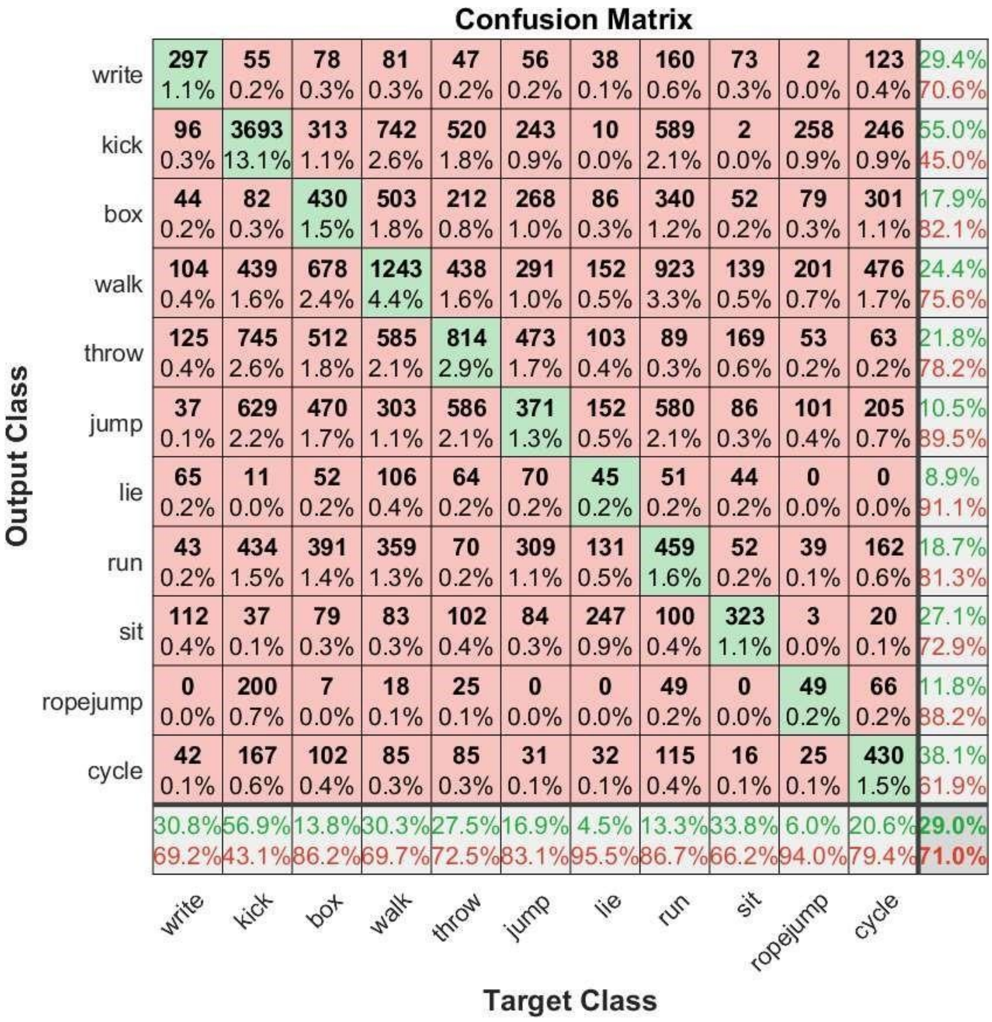


Figure C.5: 1 second epoch, LN accelerometer, Right wrist

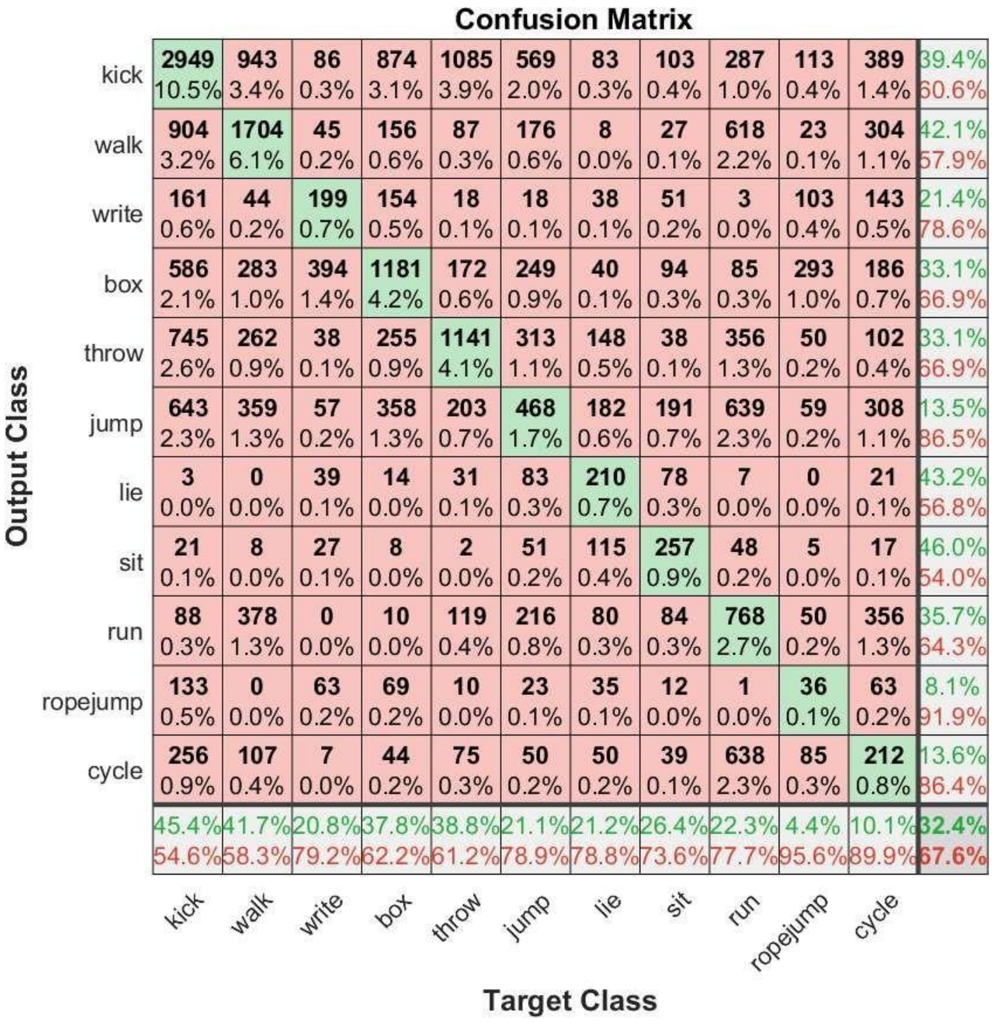


Figure C.6: 1 second epoch, WR accelerometer, Hip

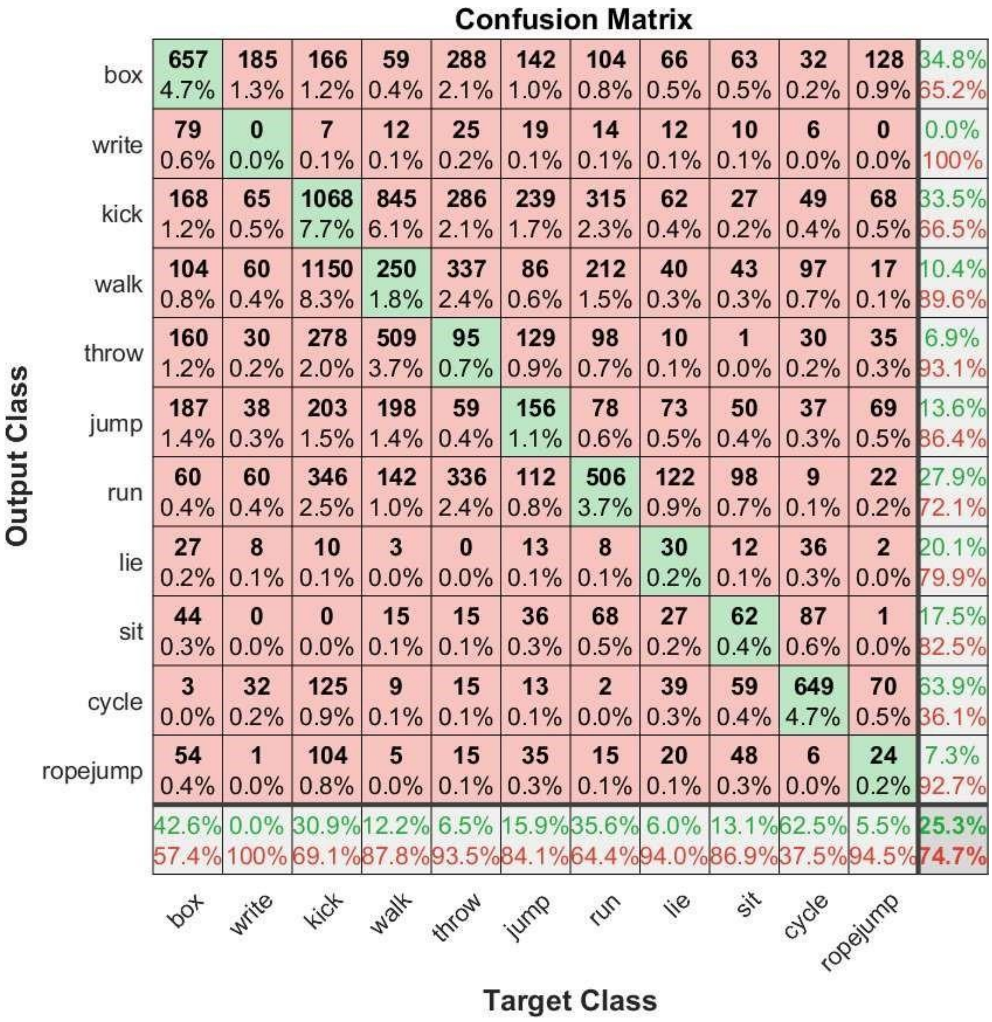


Figure C.7: 1 second epoch, WR accelerometer, Left ankle

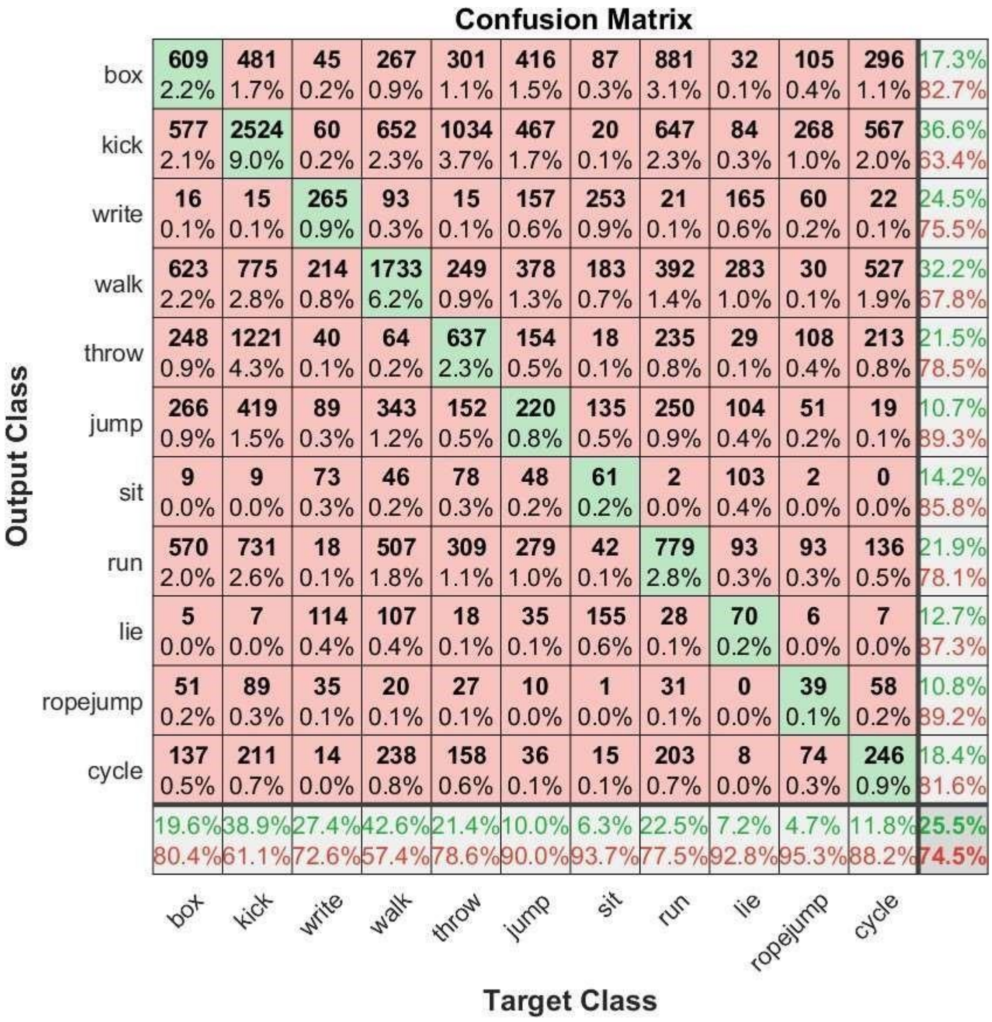


Figure C.8: 1 second epoch, WR accelerometer, Left wrist

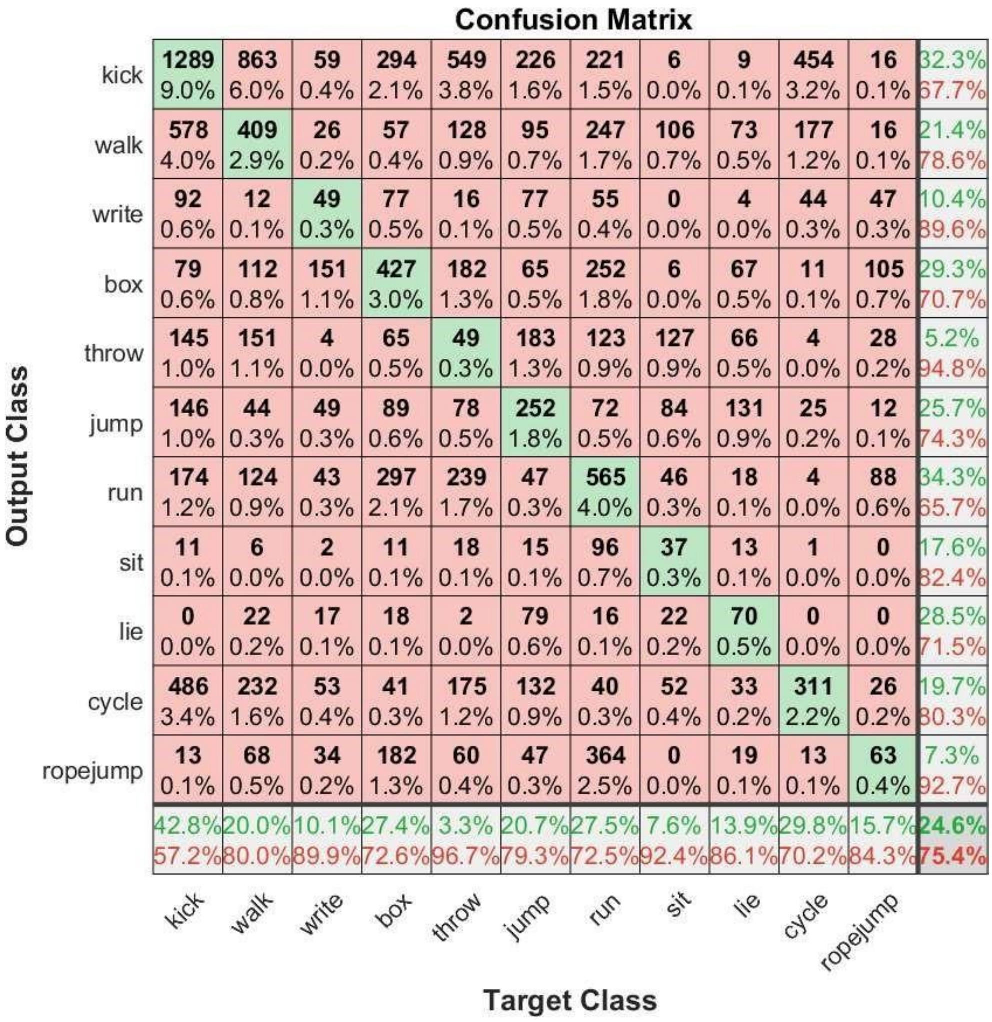


Figure C.9: 1 second epoch, WR accelerometer, Right ankle

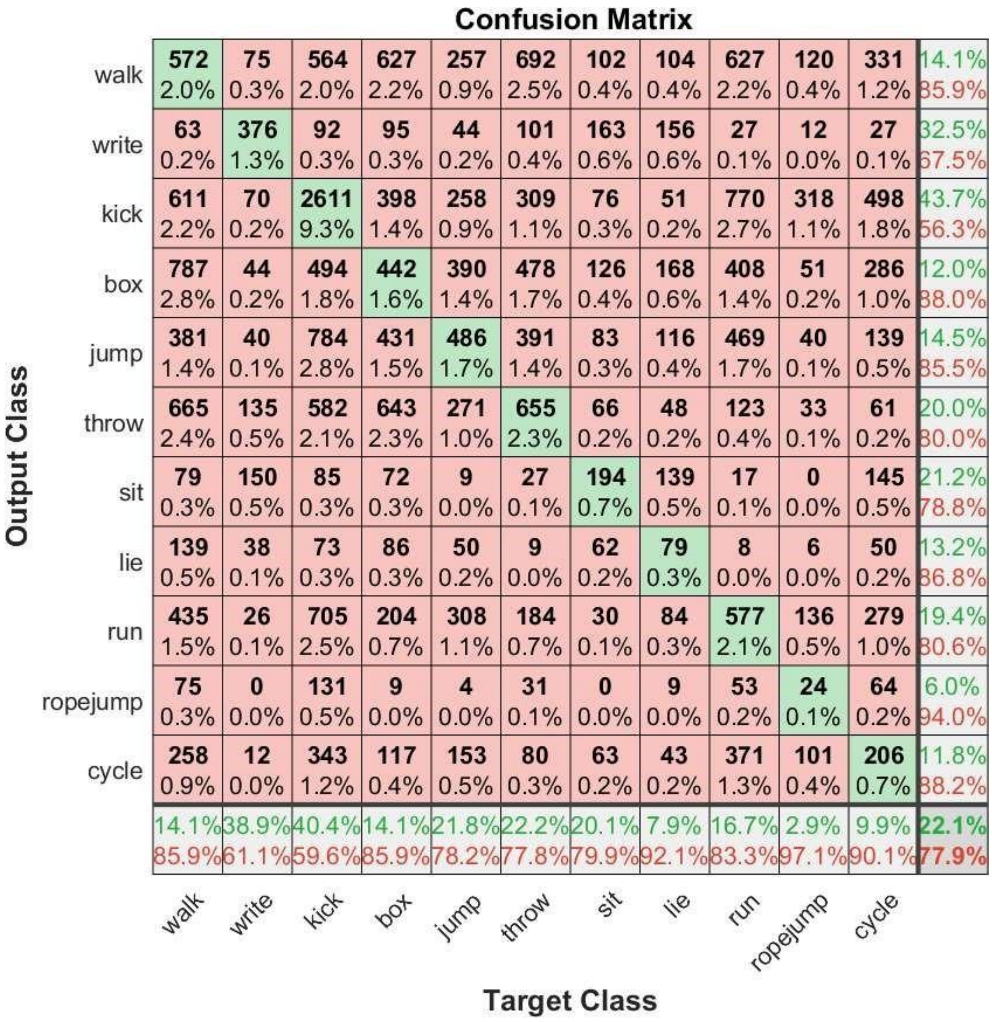


Figure C.10: 1 second epoch, WR accelerometer, Right wrist

C.2. 0.5 Second Epoch

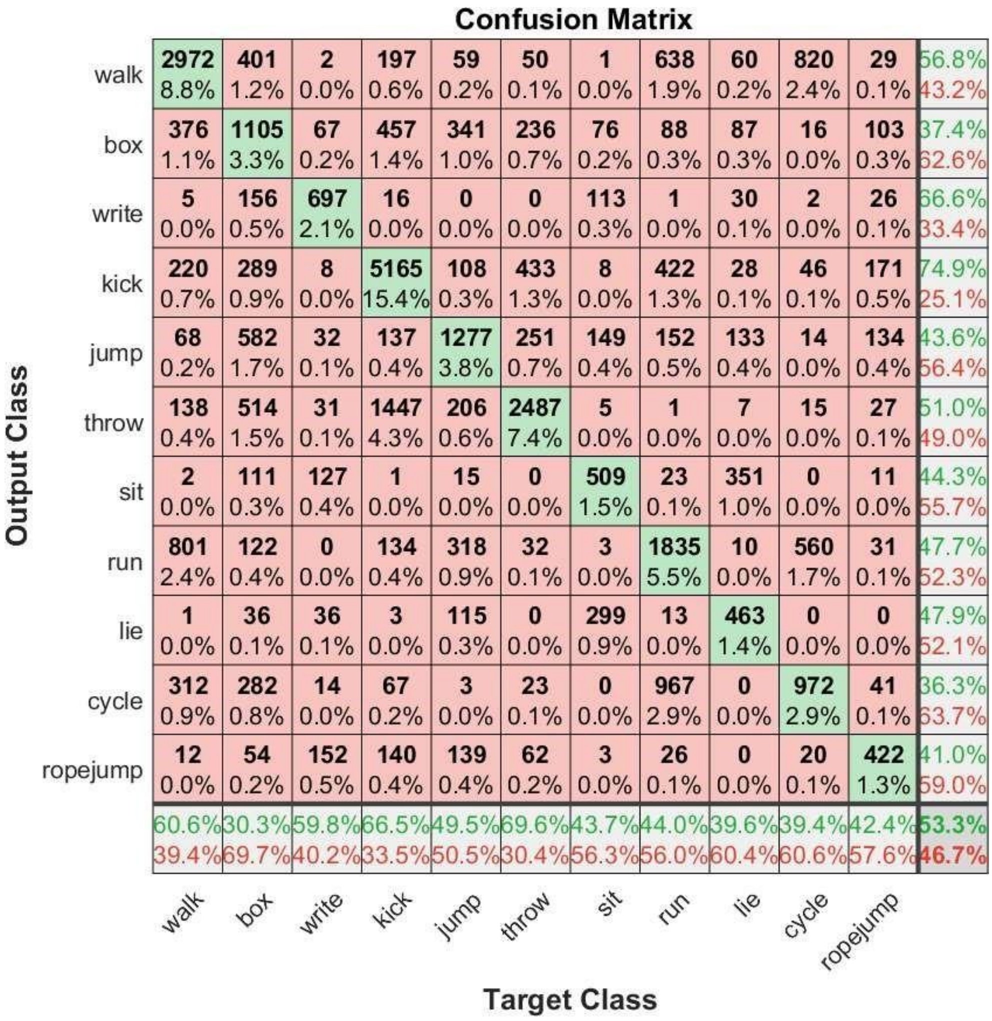


Figure C.11: 0.5 second epoch, LN accelerometer, Hip

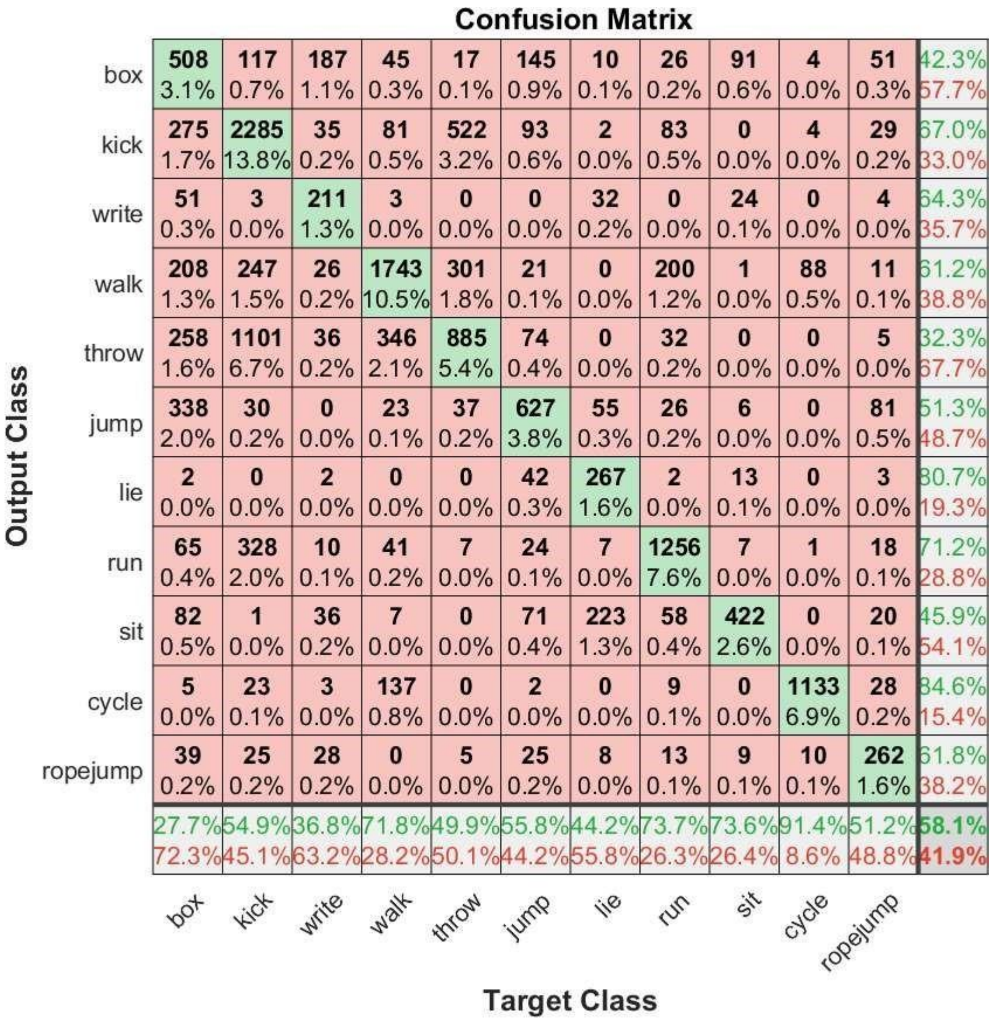


Figure C.12: 0.5 second epoch, LN accelerometer, Left ankle

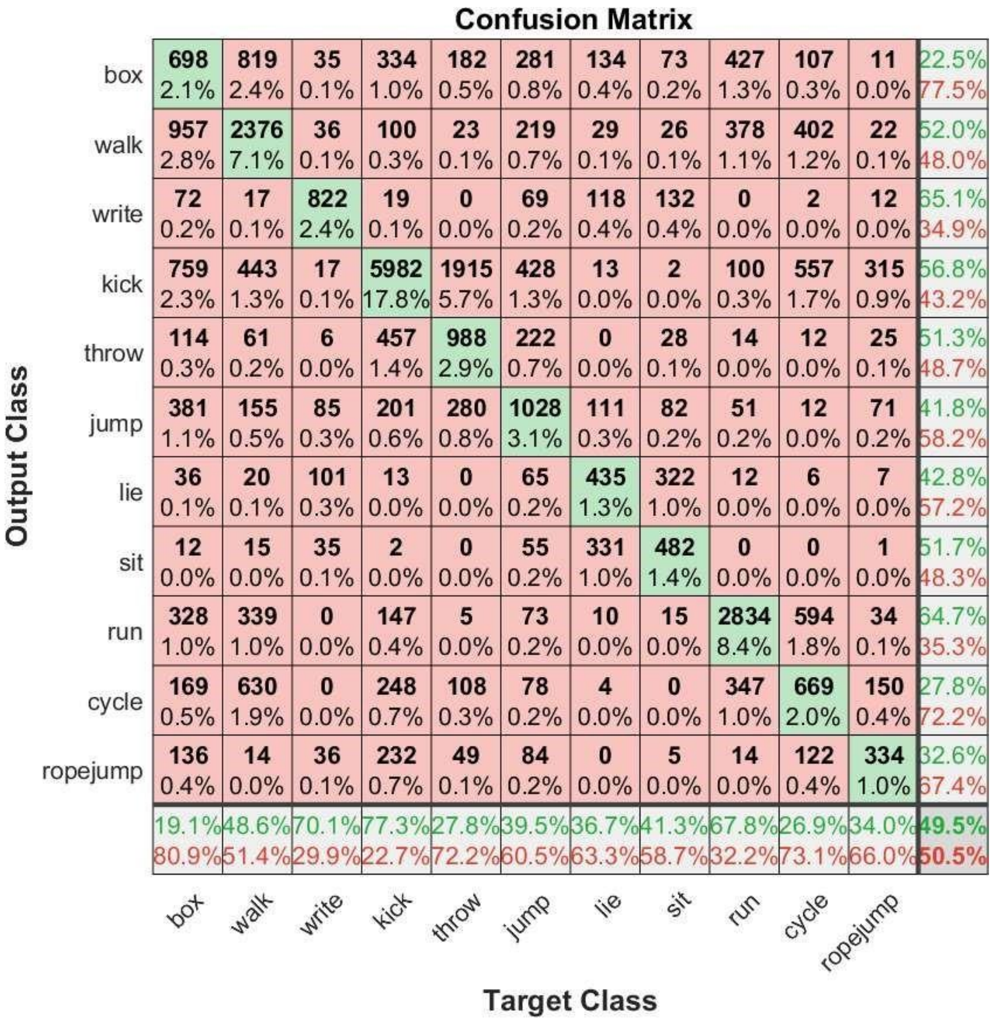


Figure C.13: 0.5 second epoch, LN accelerometer, Left wrist

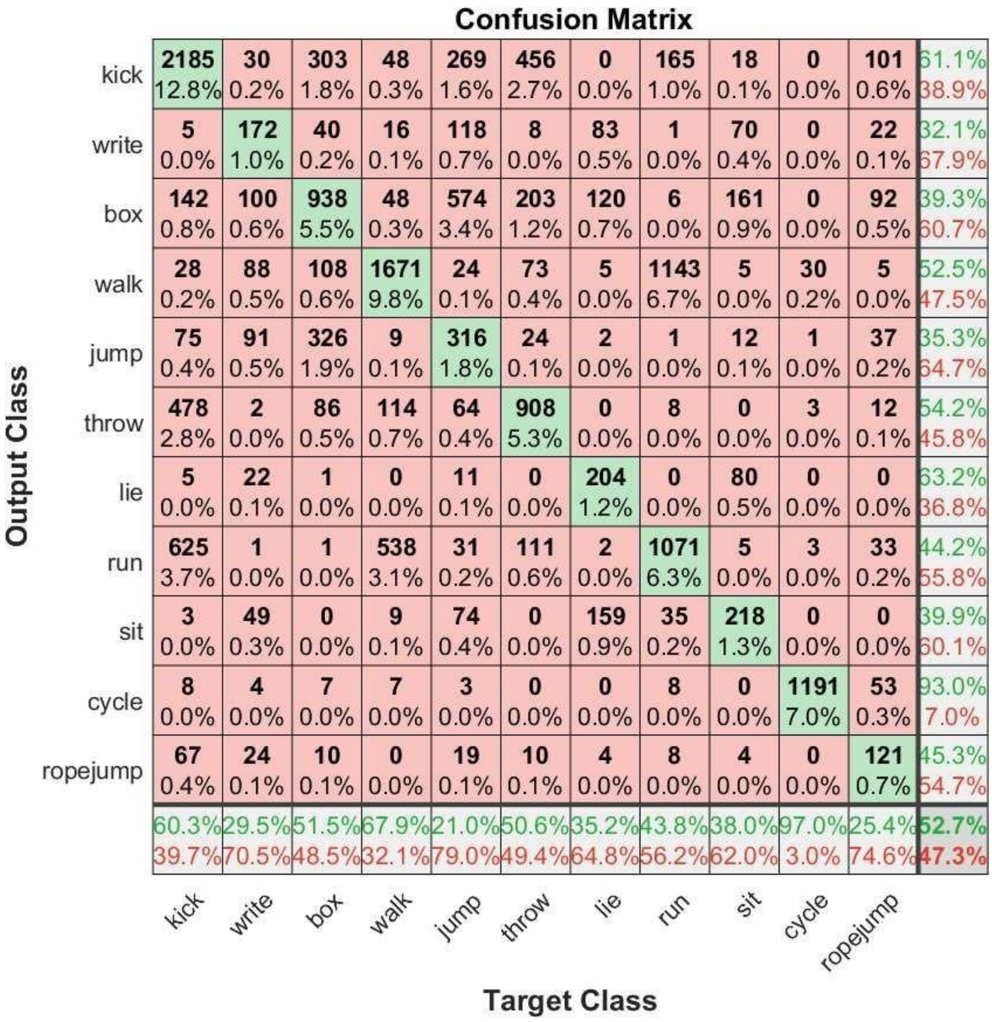


Figure C.14: 0.5 second epoch, LN accelerometer, Right ankle

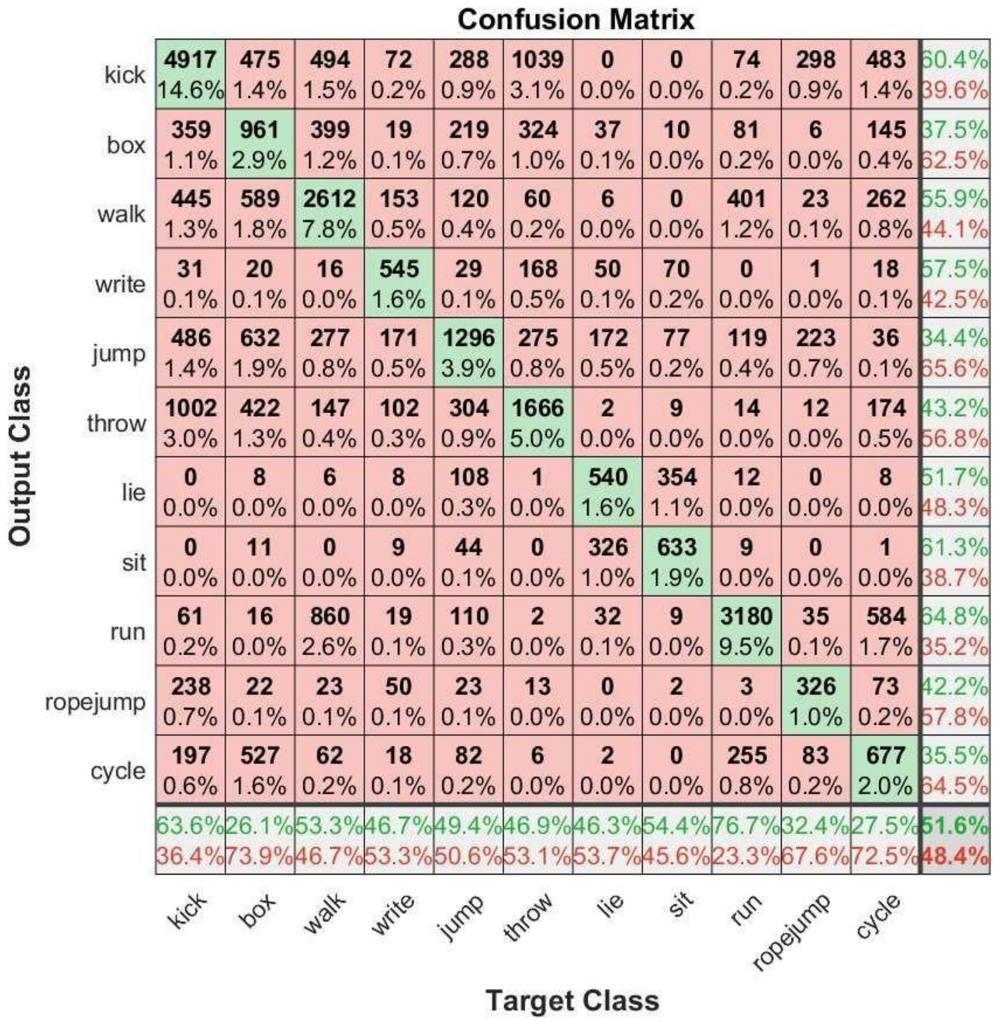


Figure C.15: 0.5 second epoch, LN accelerometer, Right wrist

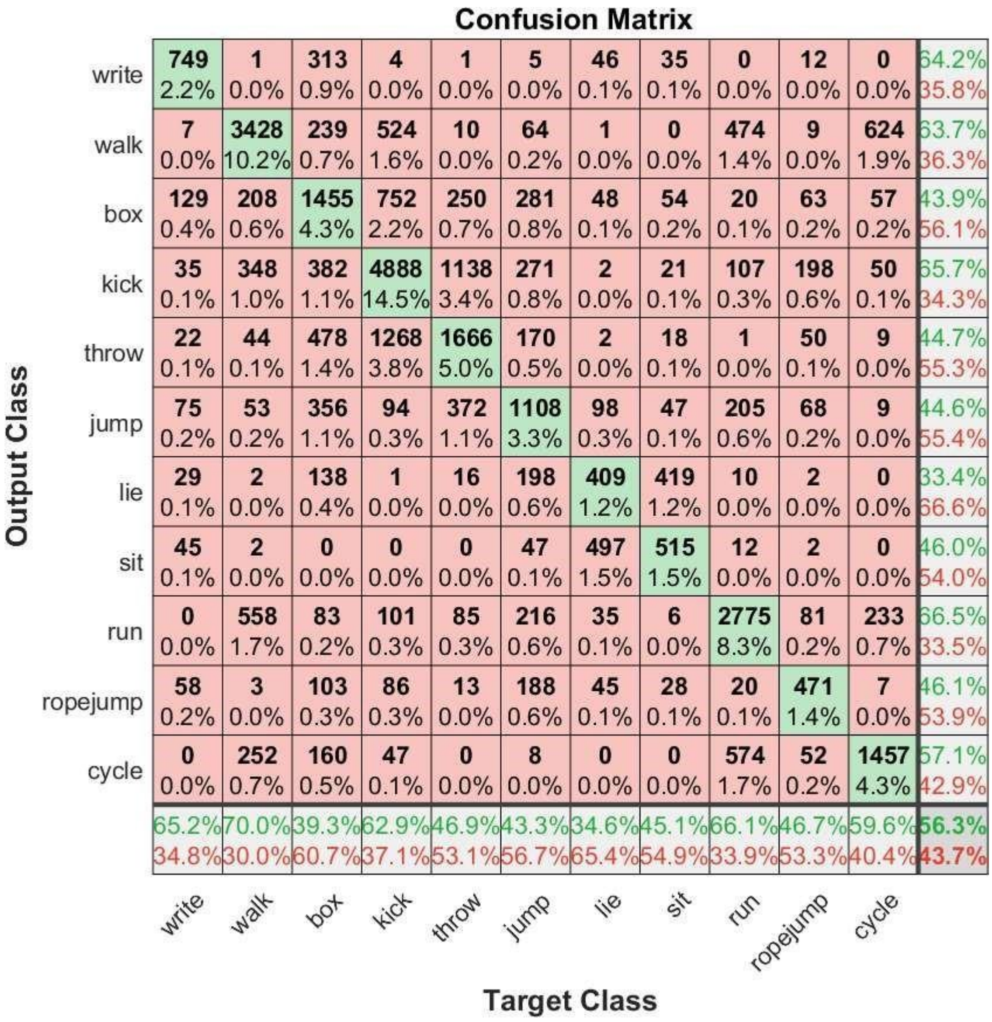


Figure C.16: 0.5 second epoch, WR accelerometer, Hip

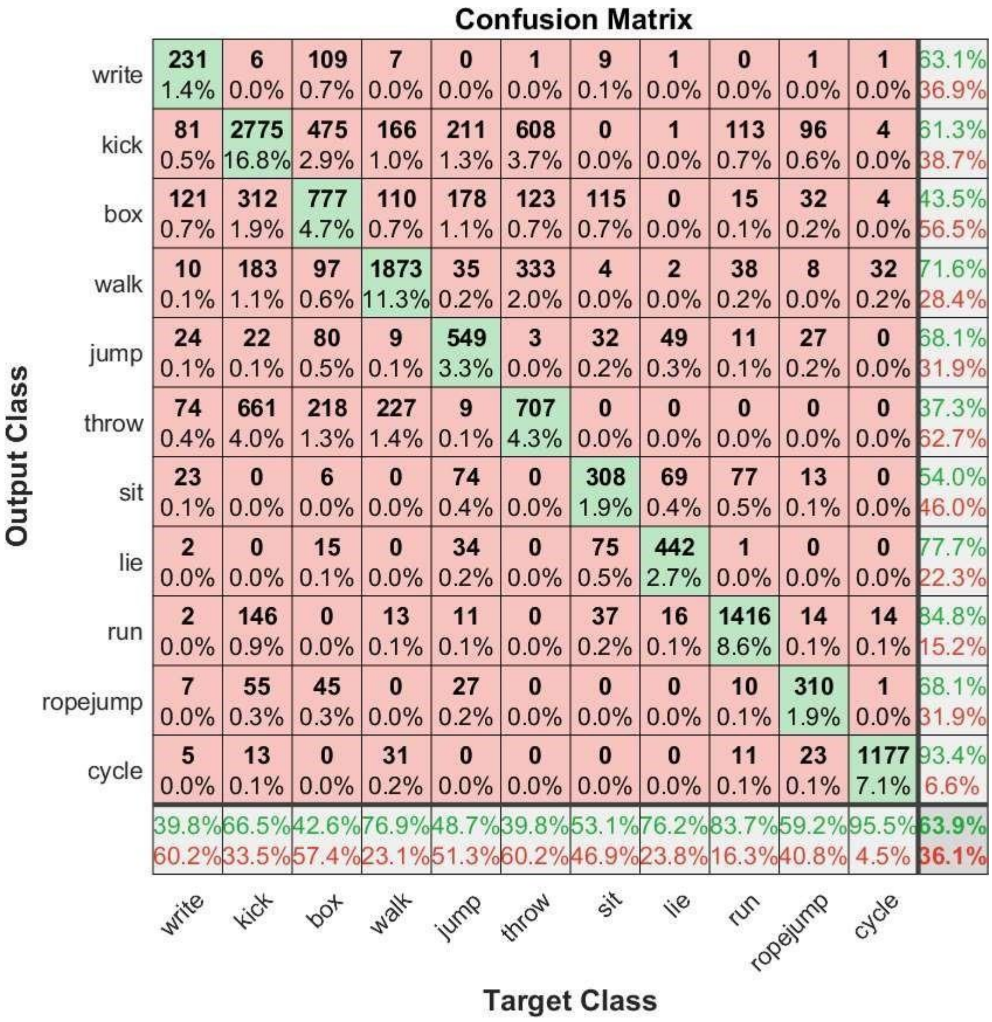


Figure C.17: 0.5 second epoch, WR accelerometer, Left ankle

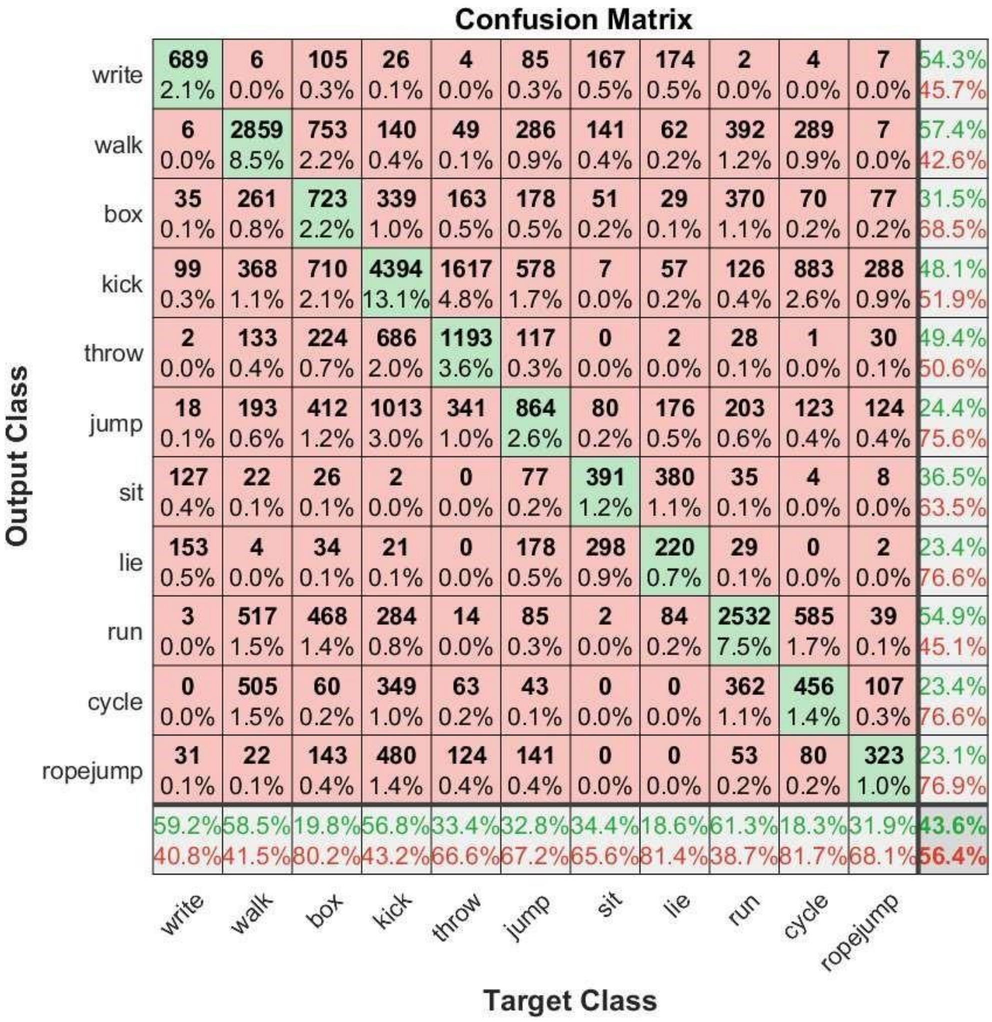


Figure C.18: 0.5 second epoch, WR accelerometer, Left wrist

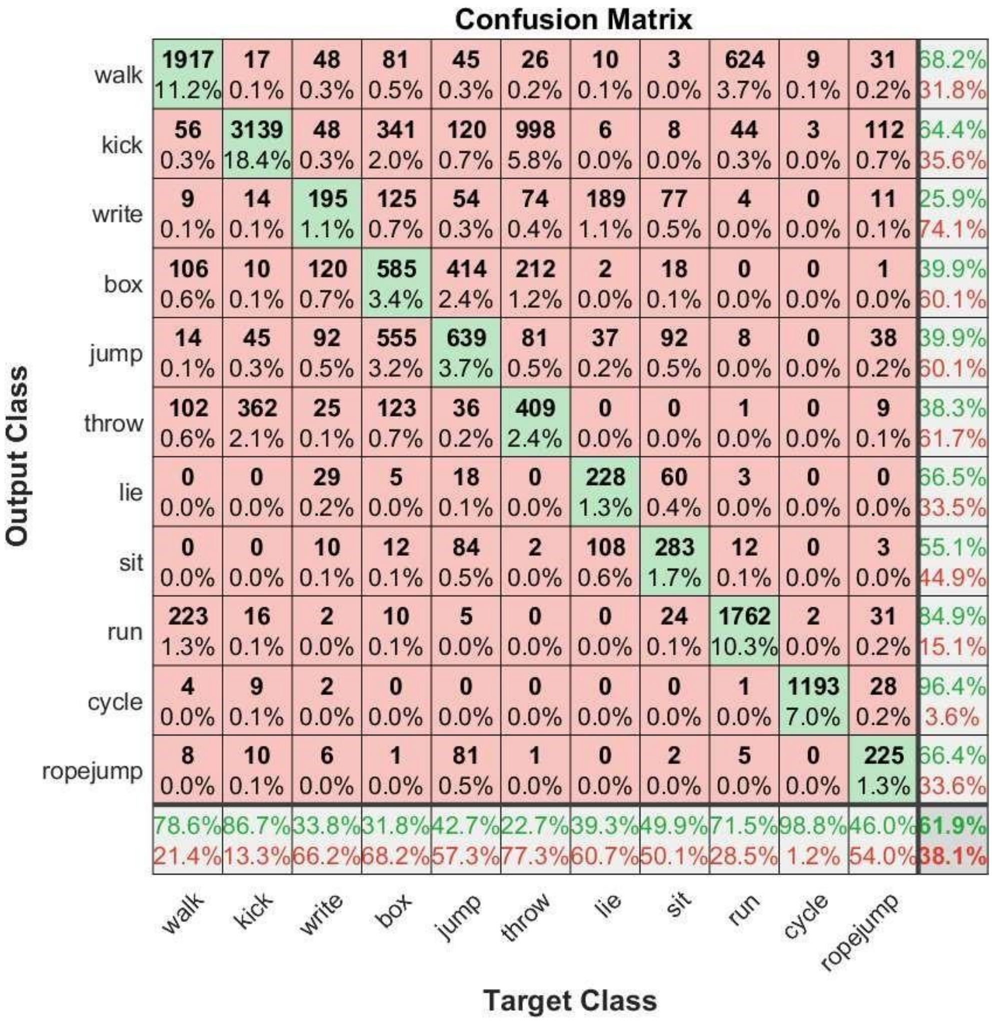


Figure C.19: 0.5 second epoch, WR accelerometer, Right ankle

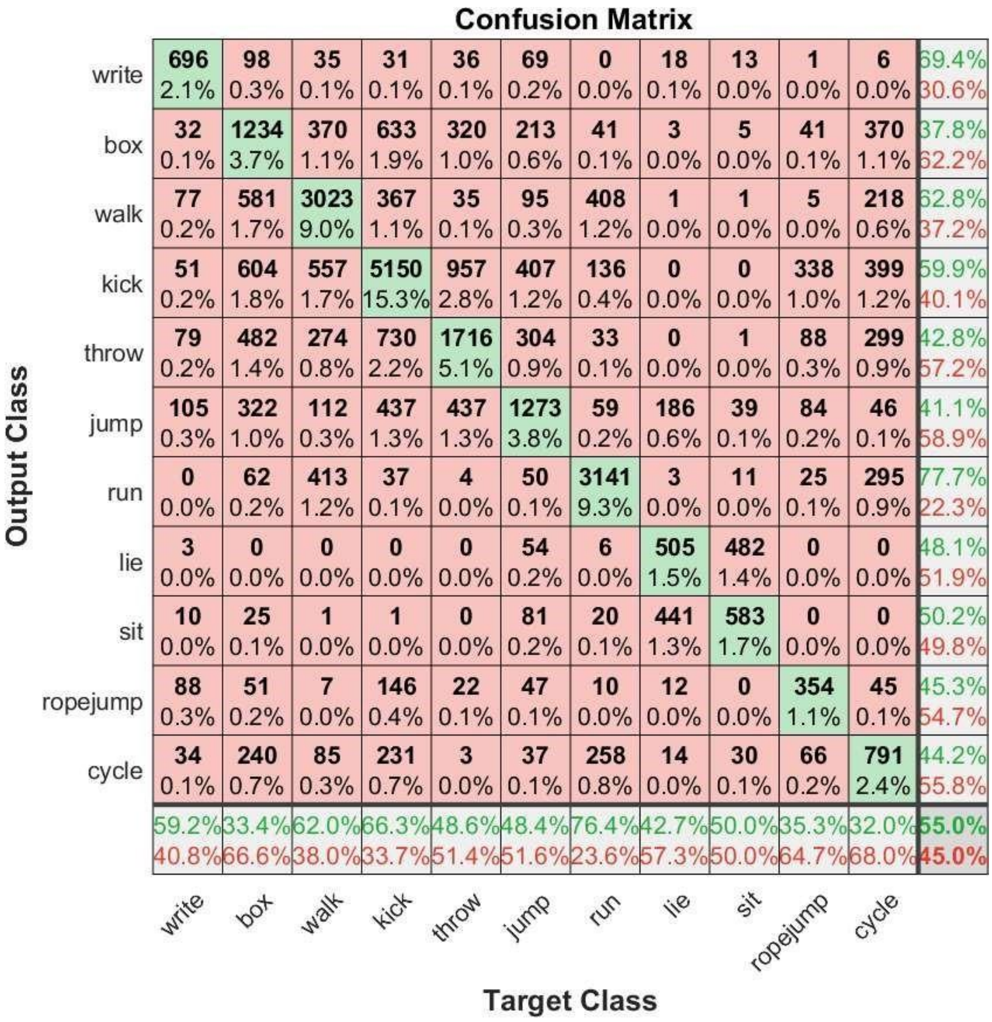


Figure C.20: 0.5 second epoch, WR accelerometer, Right wrist

C.3. 0.33 Second Epoch

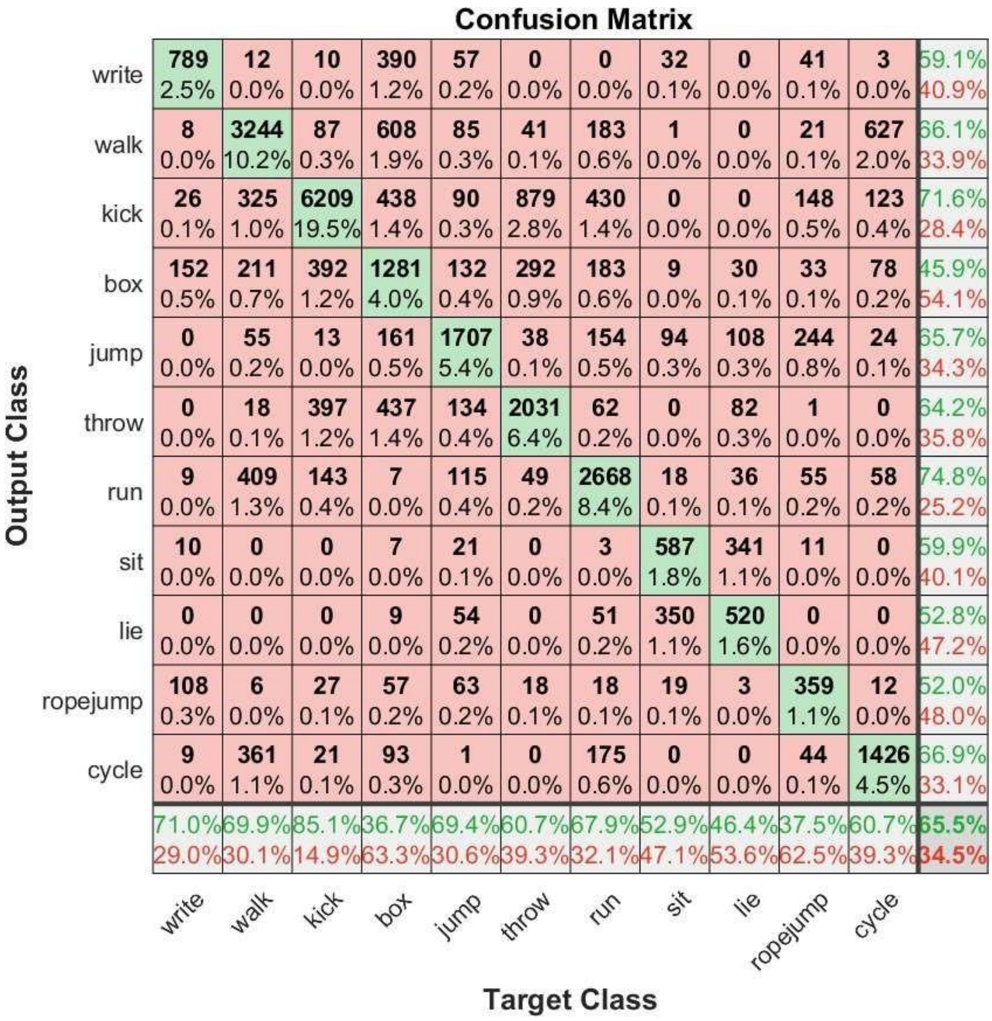


Figure C.21: 0.33 second epoch, LN accelerometer, Hip

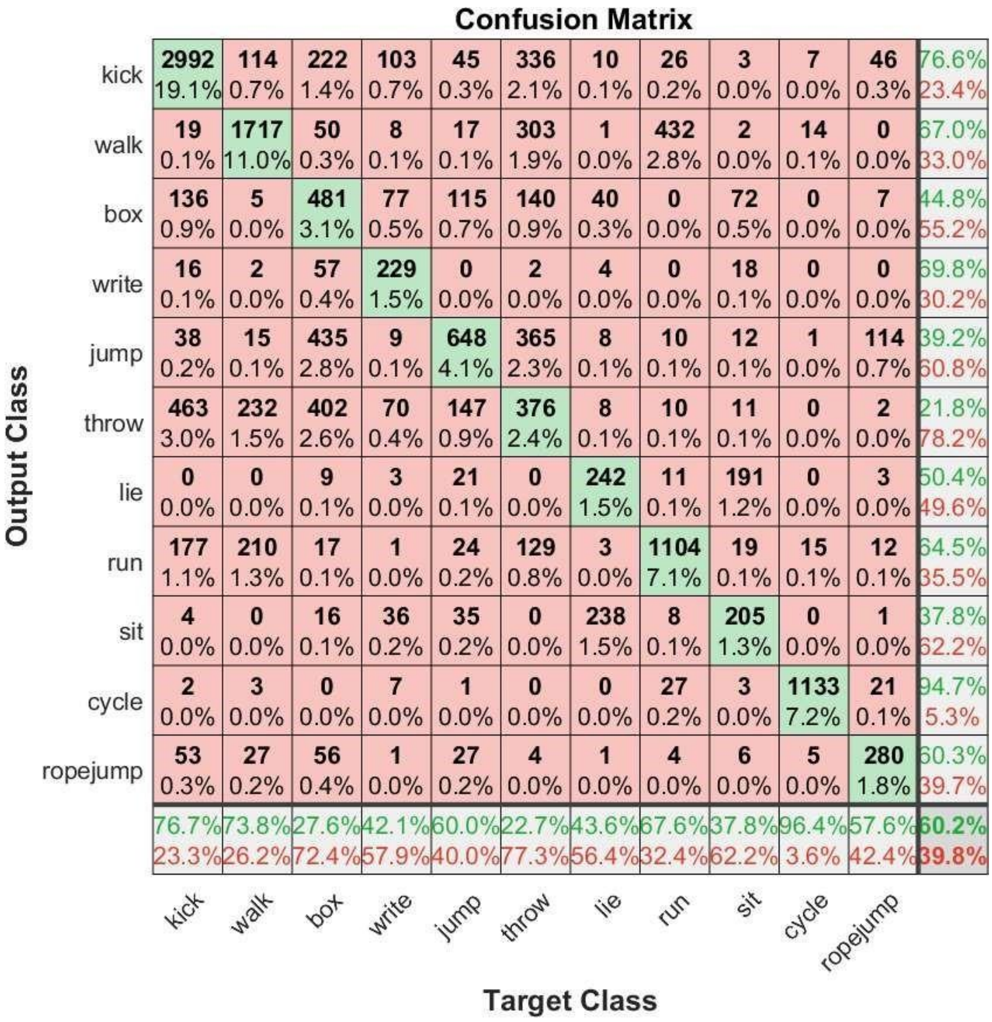


Figure C.22: 0.33 second epoch, LN accelerometer, Left ankle

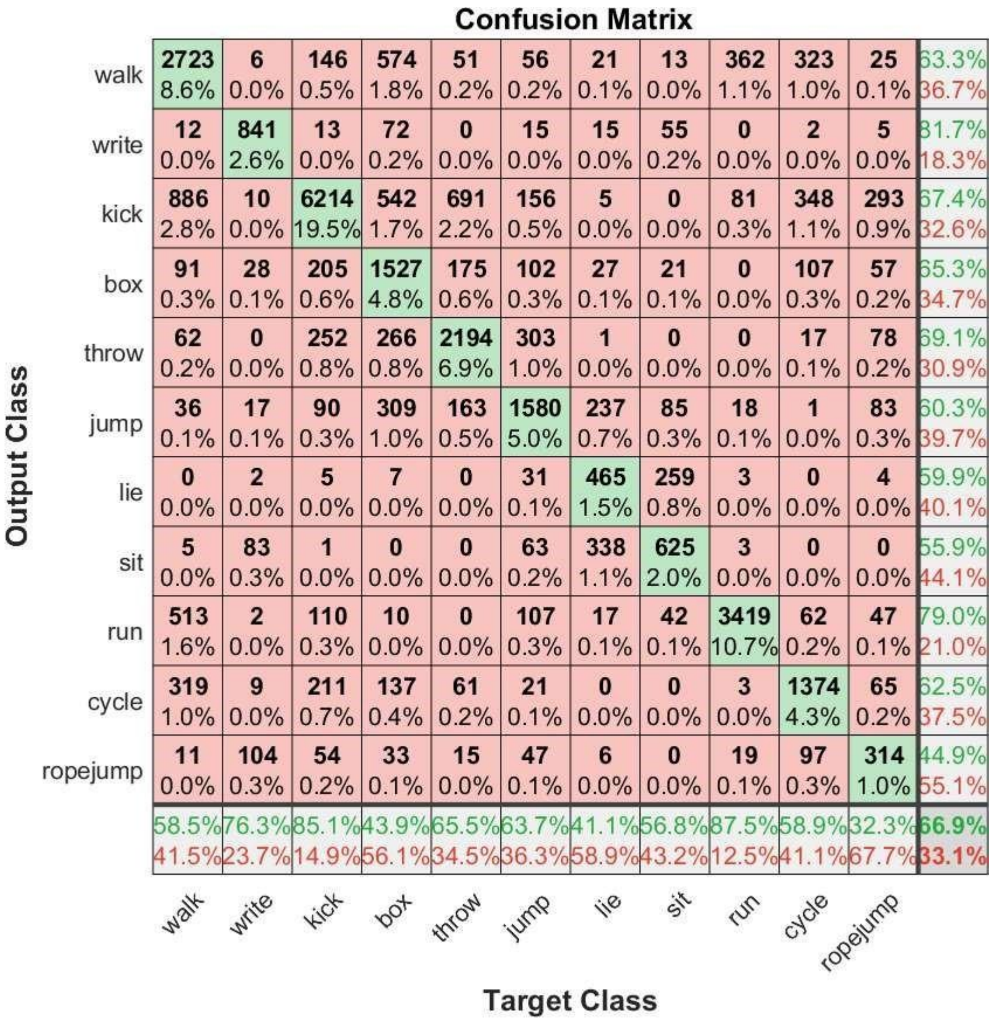


Figure C.23: 0.33 second epoch, LN accelerometer, Left wrist

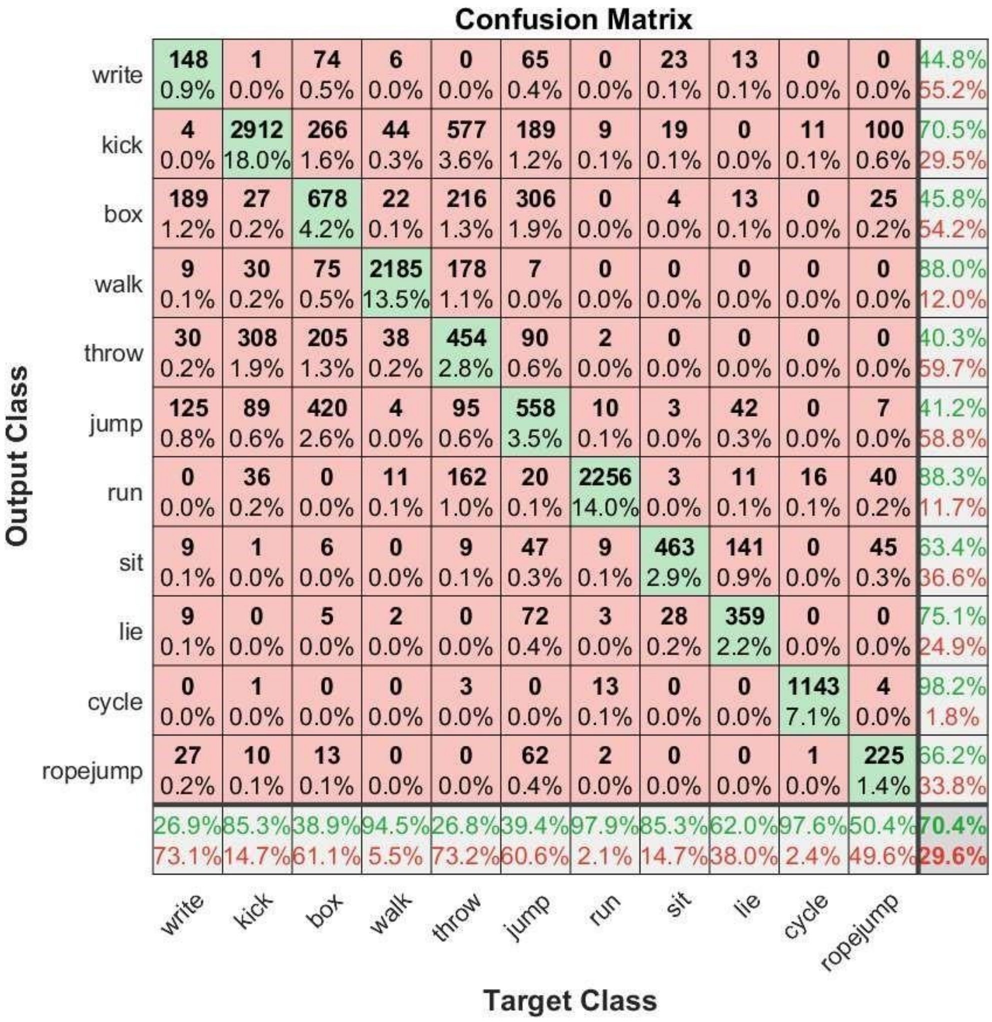


Figure C.24: 0.33 second epoch, LN accelerometer, Right ankle

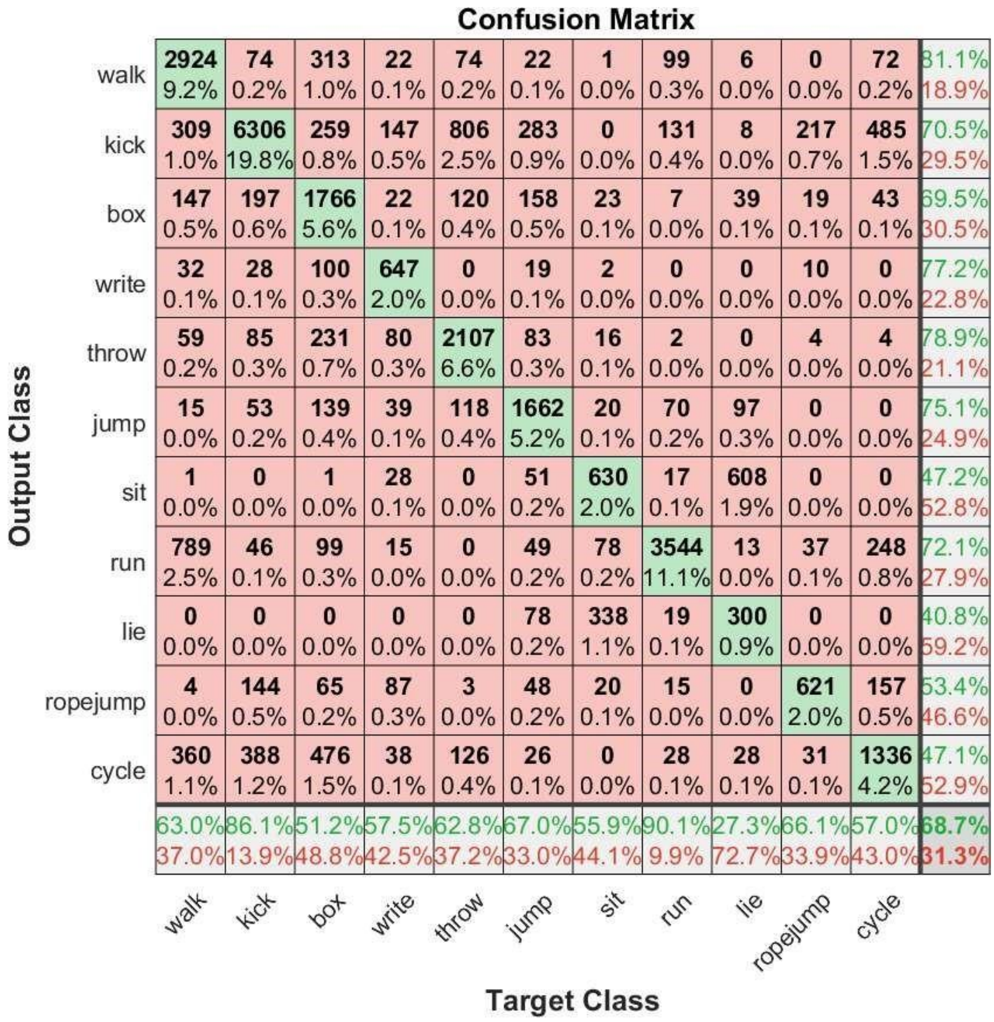


Figure C.25: 0.33 second epoch, LN accelerometer, Right wrist

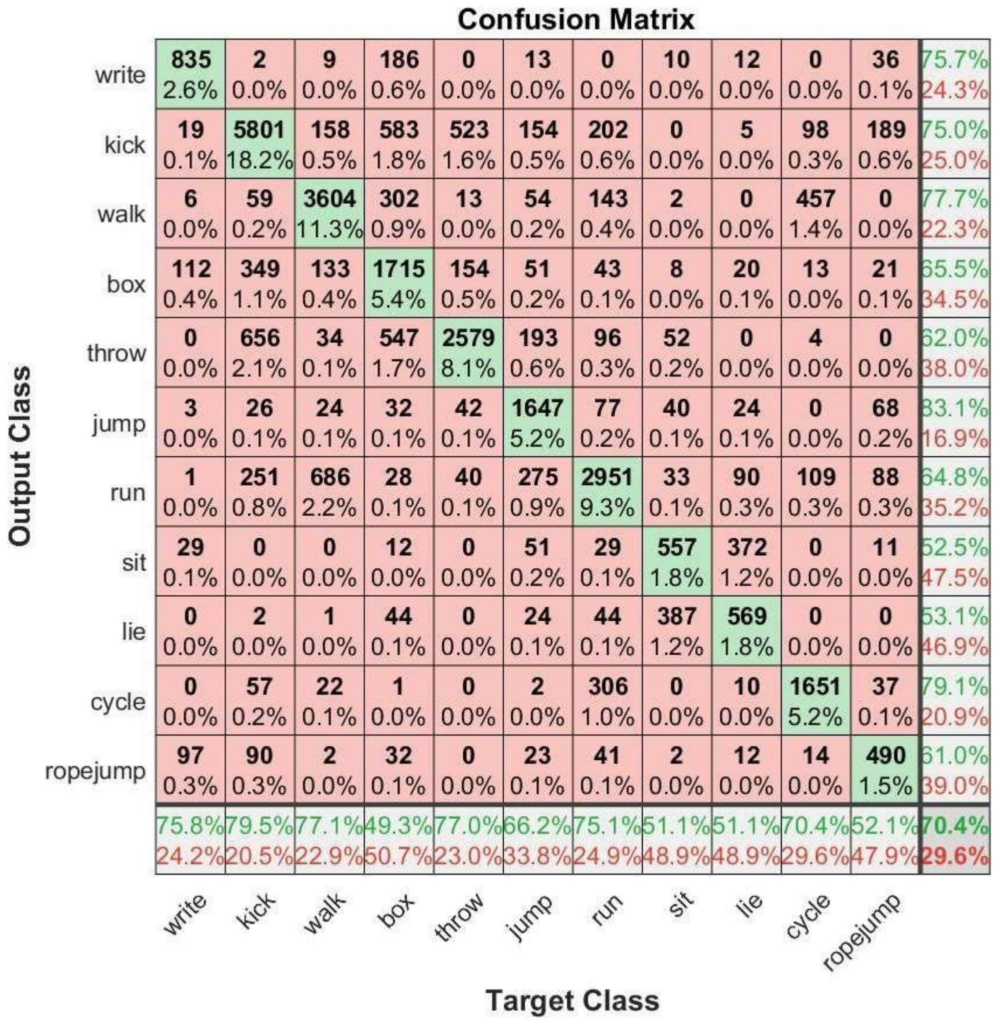


Figure C.26: 0.33 second epoch, WR accelerometer, Hip

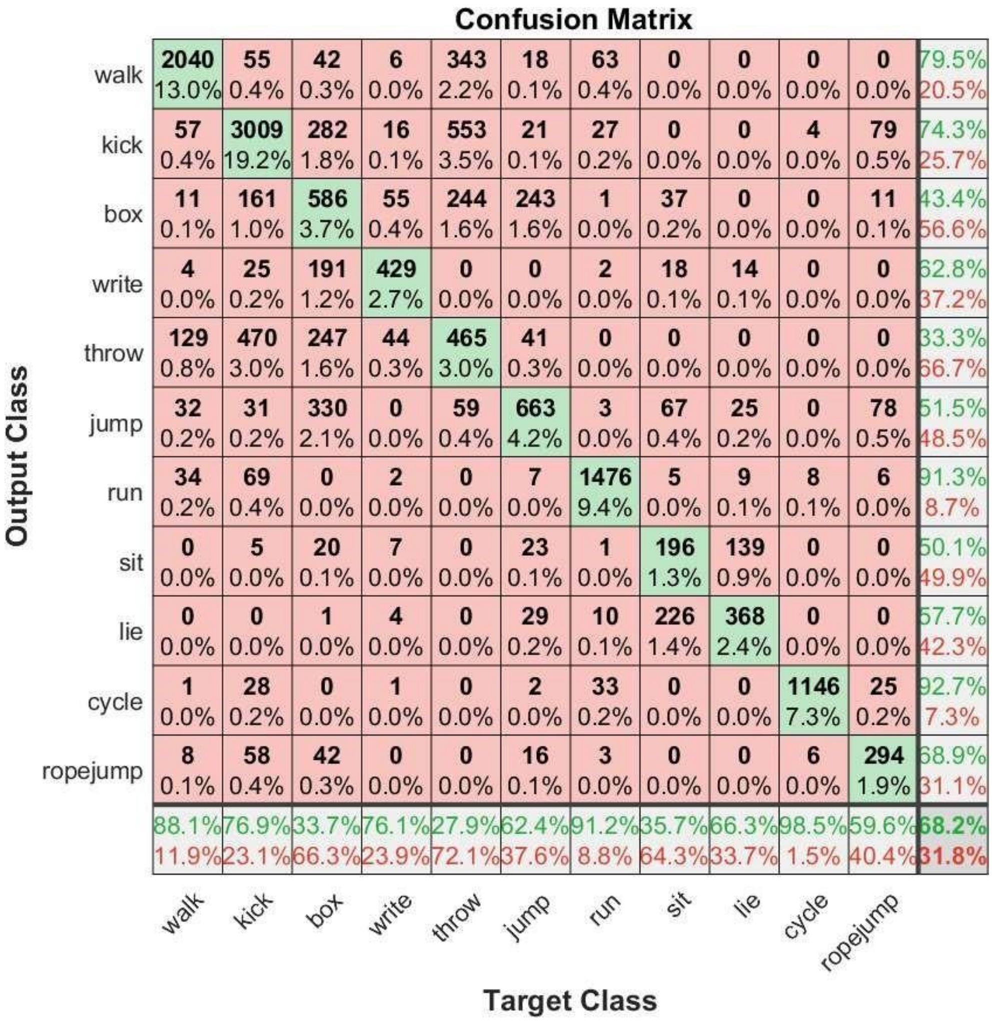


Figure C.27: 0.33 second epoch, WR accelerometer, Left ankle

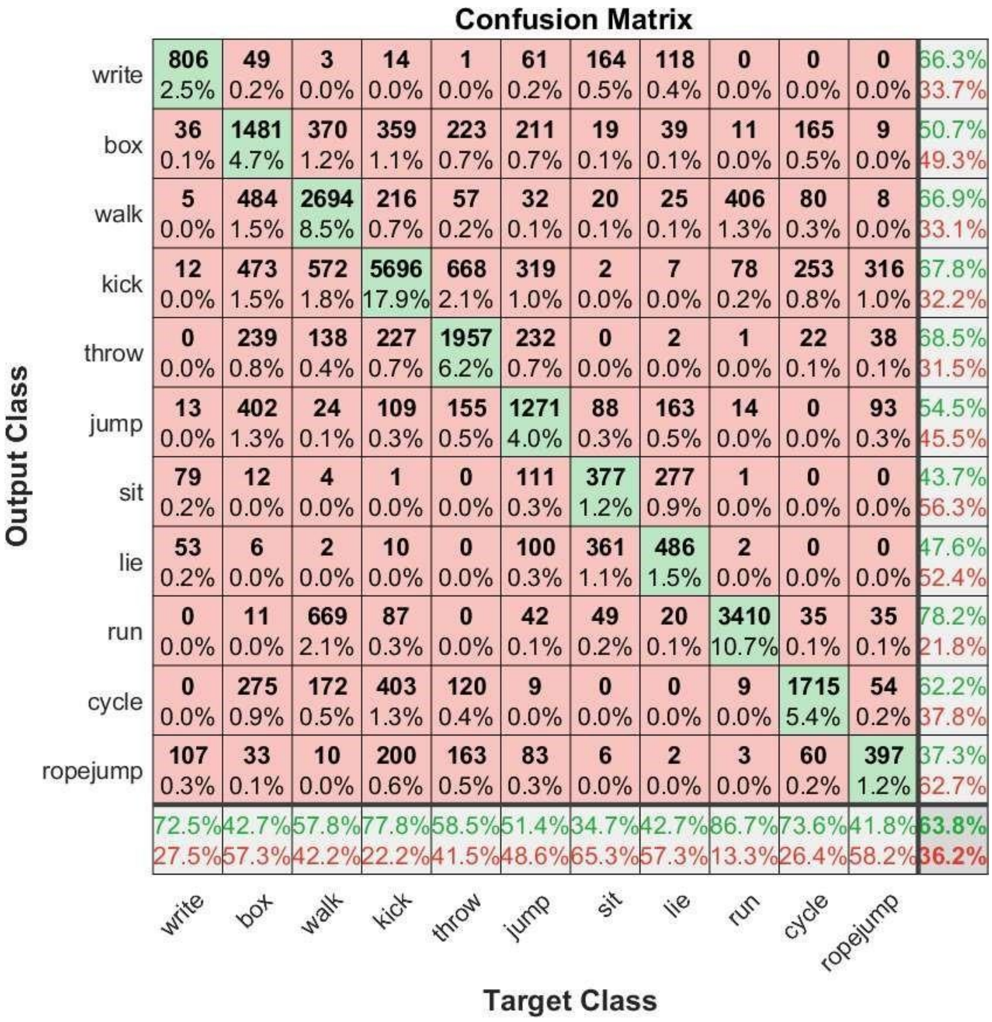


Figure C.28: 0.33 second epoch, WR accelerometer, Left wrist

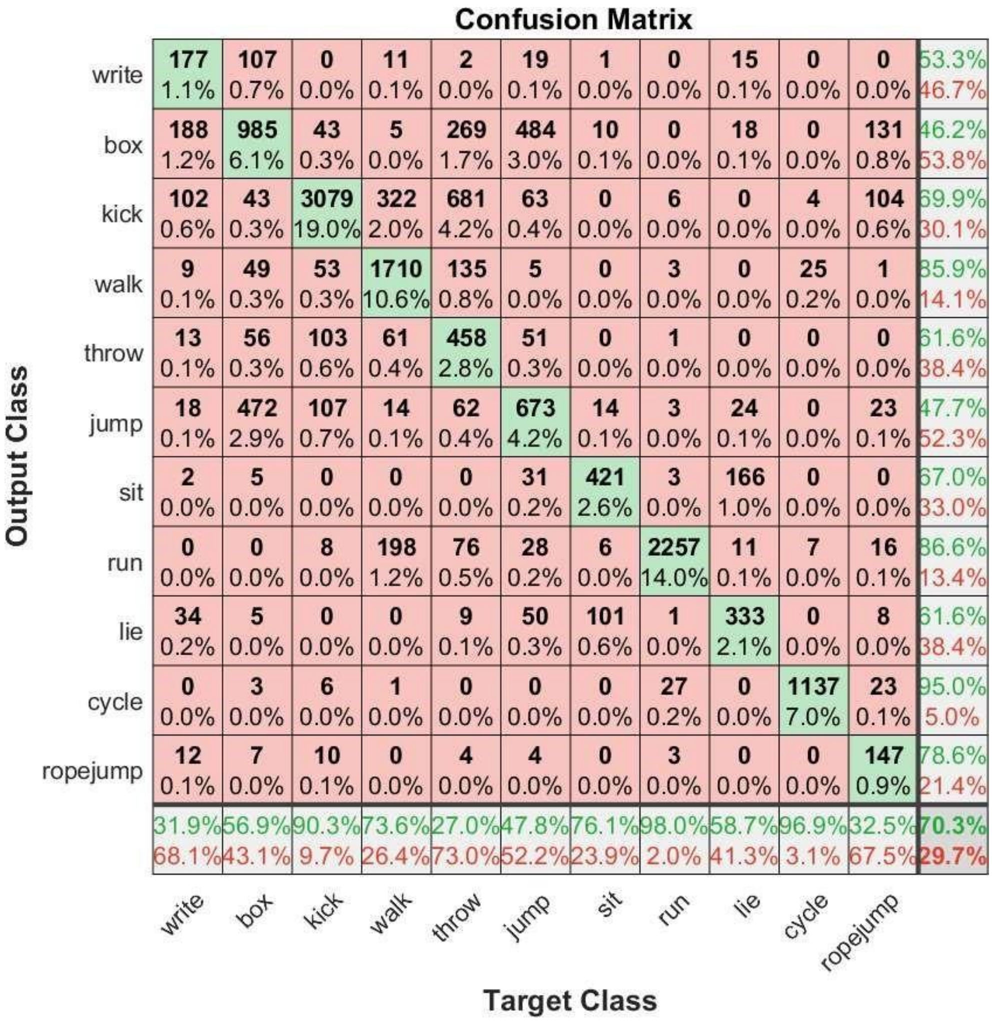


Figure C.29: 0.33 second epoch, WR accelerometer, Right ankle

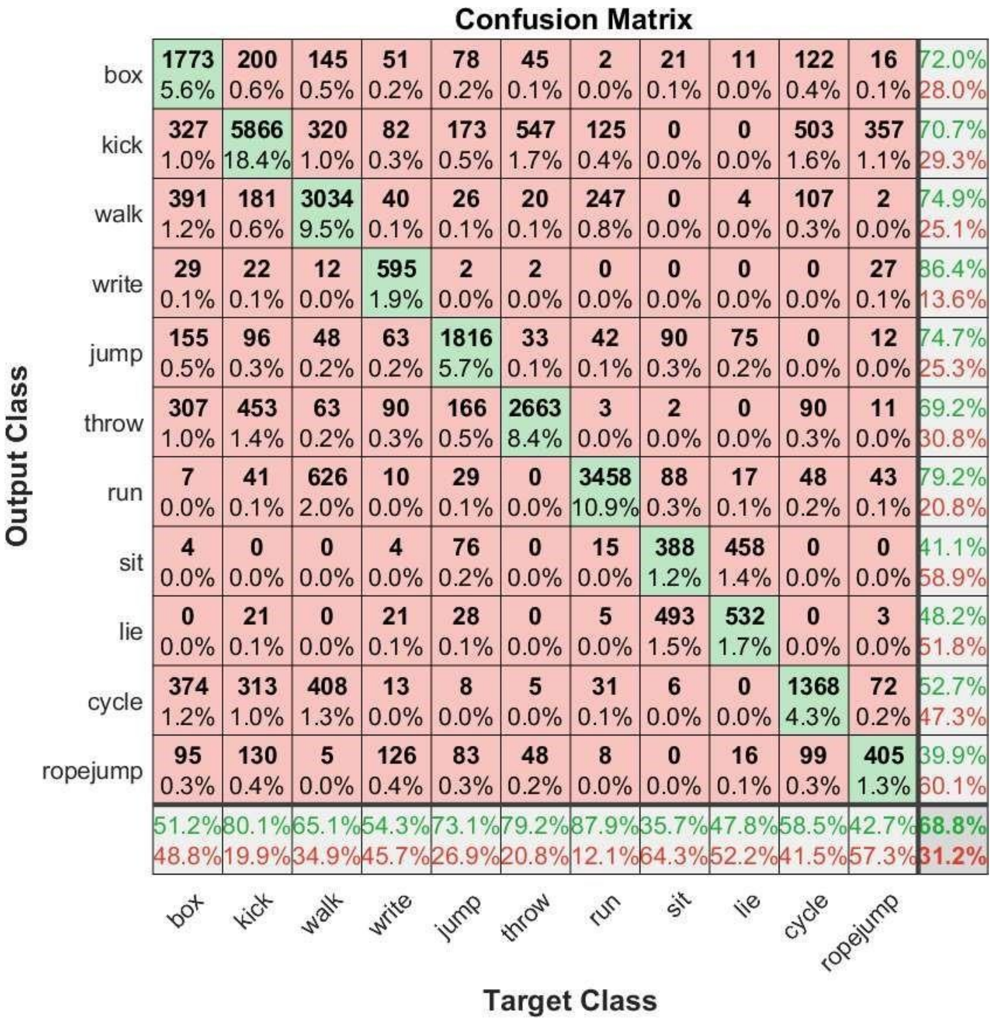


Figure C.30: 0.33 second epoch, WR accelerometer, Right wrist

C.3.1. Hip and right wrist: 2 placements

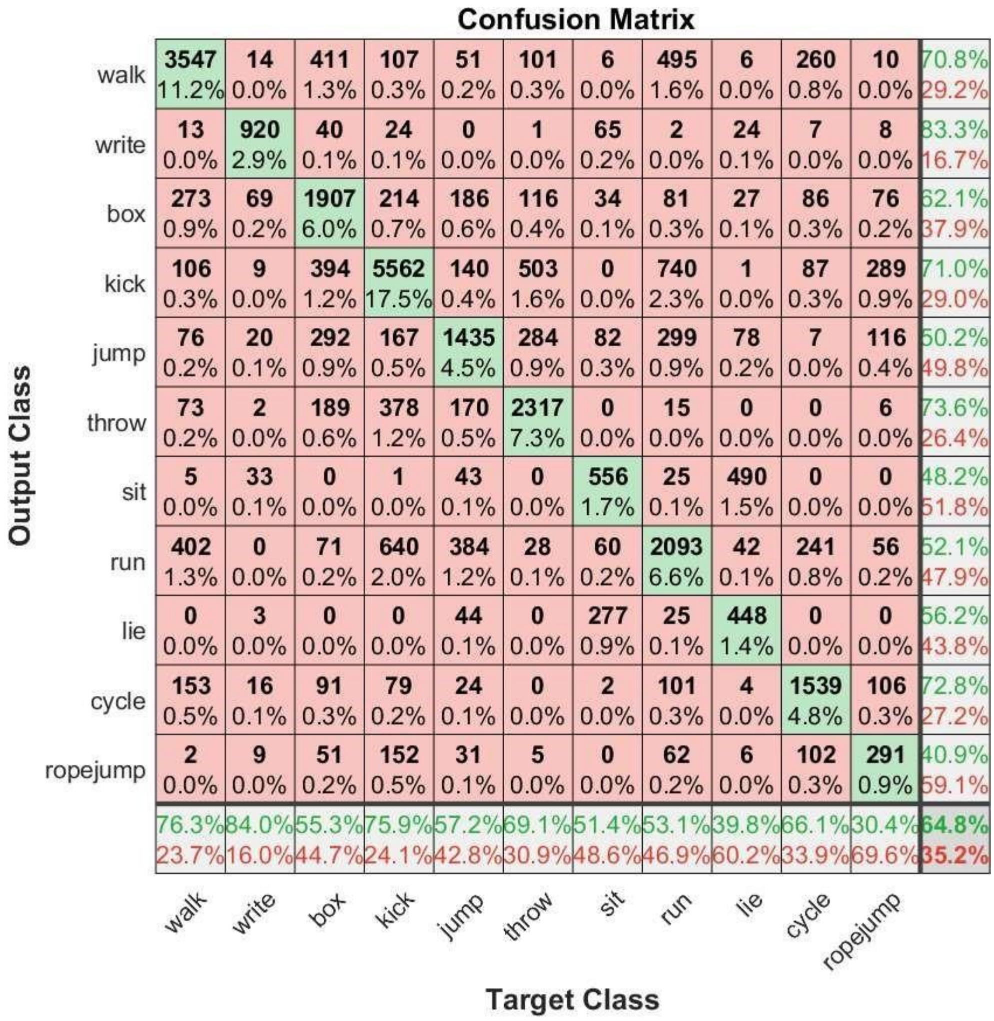


Figure C.31: 0.33 second epoch, LN accelerometer, Hip & Right wrist

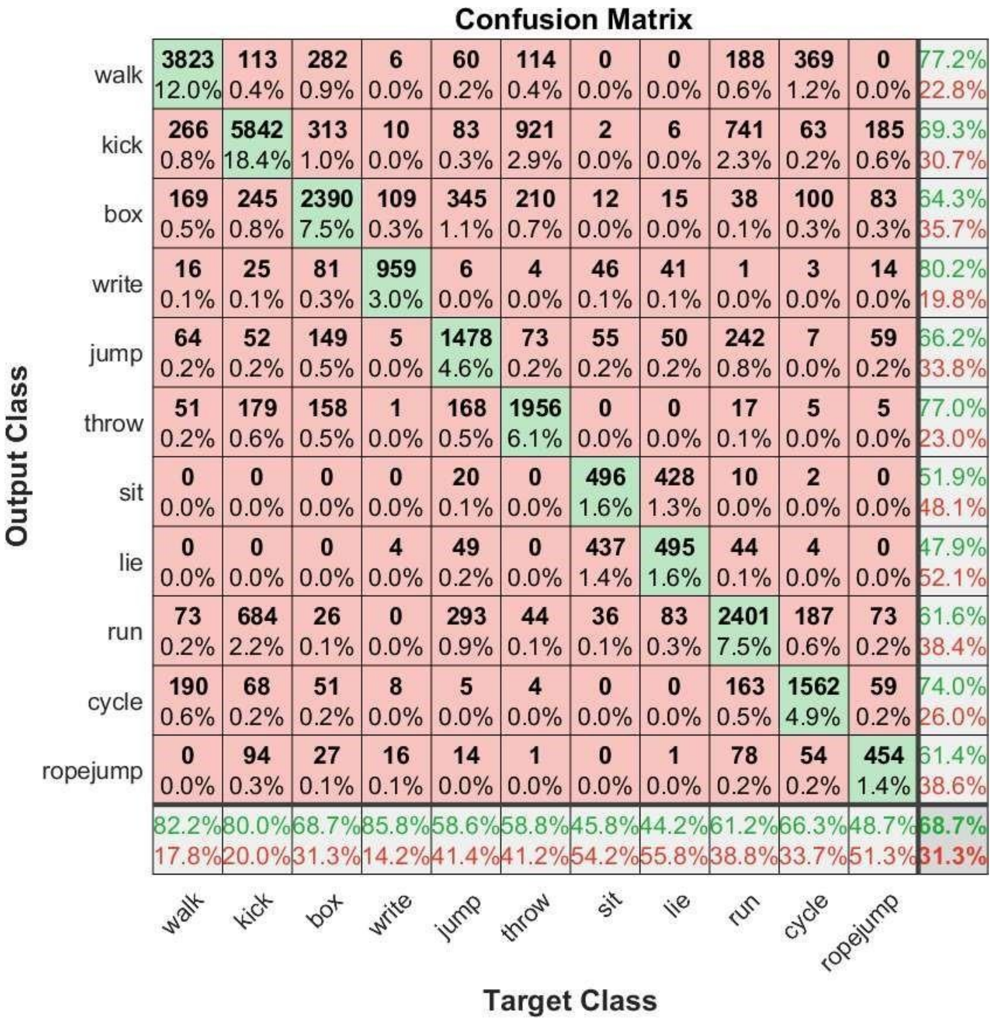


Figure C.32: 0.33 second epoch, WR accelerometer, Hip & Right wrist

C.3.2. Hip and right wrist: Activity - Intensity

Confusion Matrix

Output Class \ Target Class	runVPA	lieSB	cycleMPA	sitSB	cycleVPA	kickMPA	kickVPA	ropejumpVPA	throwMPA	boxVPA	walkLPA	walkMPA	jumpMPA	jumpVPA	throwLPA	writeLPA	
runVPA	2478 8.1%	27 0.1%	75 0.2%	11 0.0%	137 0.5%	102 0.3%	163 0.5%	83 0.3%	68 0.2%	33 0.1%	20 0.1%	309 1.0%	30 0.1%	9 0.0%	0 0.0%	2 0.0%	69.9% 30.1%
lieSB	19 0.1%	538 1.8%	21 0.1%	303 1.0%	48 0.2%	0 0.0%	0 0.0%	0 0.0%	26 0.1%	61 0.2%	0 0.0%	2 0.0%	1 0.0%	24 0.1%	10 0.0%	16 0.1%	50.3% 49.7%
cycleMPA	230 0.8%	17 0.1%	994 3.3%	1 0.0%	638 2.1%	23 0.1%	44 0.1%	9 0.0%	0 0.0%	8 0.0%	243 0.8%	91 0.3%	4 0.0%	7 0.0%	1 0.0%	3 0.0%	43.0% 57.0%
sitSB	7 0.0%	358 1.2%	6 0.0%	526 1.7%	25 0.1%	0 0.0%	0 0.0%	30 0.1%	0 0.0%	38 0.1%	0 0.0%	0 0.0%	1 0.0%	6 0.0%	65 0.2%	103 0.3%	45.2% 54.8%
cycleVPA	170 0.6%	25 0.1%	729 2.4%	3 0.0%	807 2.7%	25 0.1%	130 0.4%	13 0.0%	5 0.0%	65 0.2%	13 0.0%	61 0.2%	12 0.0%	11 0.0%	5 0.0%	3 0.0%	38.9% 61.1%
kickMPA	163 0.5%	0 0.0%	25 0.1%	0 0.0%	43 0.1%	245 0.8%	272 0.9%	80 0.3%	9 0.0%	63 0.2%	21 0.1%	15 0.0%	40 0.1%	1 0.0%	18 0.1%	1 0.0%	24.6% 75.4%
kickVPA	278 0.9%	0 0.0%	78 0.3%	0 0.0%	114 0.4%	311 1.0%	1409 4.6%	137 0.5%	337 1.1%	52 0.2%	5 0.0%	1 0.0%	5 0.0%	1 0.0%	3 0.0%	0 0.0%	51.6% 48.4%
ropejumpVPA	25 0.1%	12 0.0%	8 0.0%	73 0.2%	36 0.1%	20 0.1%	42 0.1%	430 1.4%	12 0.0%	22 0.1%	0 0.0%	7 0.0%	174 0.6%	38 0.1%	7 0.0%	55 0.2%	44.7% 55.3%
throwMPA	55 0.2%	0 0.0%	0 0.0%	0 0.0%	17 0.1%	41 0.1%	239 0.8%	7 0.0%	1648 5.4%	507 1.7%	8 0.0%	51 0.2%	189 0.6%	68 0.2%	164 0.5%	4 0.0%	55.0% 45.0%
boxVPA	51 0.2%	42 0.1%	46 0.2%	0 0.0%	61 0.2%	69 0.2%	282 0.9%	45 0.1%	178 0.6%	1530 5.0%	83 0.3%	181 0.6%	111 0.4%	107 0.4%	80 0.3%	75 0.2%	52.0% 48.0%
walkLPA	50 0.2%	0 0.0%	157 0.5%	1 0.0%	245 0.8%	5 0.0%	9 0.0%	0 0.0%	10 0.0%	179 0.6%	1490 4.9%	131 0.4%	174 0.6%	51 0.2%	3 0.0%	5 0.0%	59.4% 40.6%
walkMPA	169 0.6%	0 0.0%	155 0.5%	0 0.0%	67 0.2%	15 0.0%	37 0.1%	4 0.0%	182 0.6%	217 0.7%	239 0.8%	1296 4.3%	28 0.1%	27 0.1%	12 0.0%	7 0.0%	52.8% 47.2%
jumpMPA	88 0.3%	0 0.0%	7 0.0%	7 0.0%	36 0.1%	32 0.1%	20 0.1%	73 0.2%	56 0.2%	111 0.4%	82 0.3%	100 0.3%	693 2.3%	83 0.3%	89 0.3%	20 0.1%	46.3% 53.7%
jumpVPA	7 0.0%	1 0.0%	8 0.0%	0 0.0%	16 0.1%	1 0.0%	29 0.1%	43 0.1%	83 0.3%	389 1.3%	59 0.2%	22 0.1%	158 0.5%	156 0.5%	27 0.1%	34 0.1%	15.1% 84.9%
throwLPA	18 0.1%	54 0.2%	4 0.0%	42 0.1%	6 0.0%	17 0.1%	52 0.2%	28 0.1%	187 0.6%	152 0.5%	16 0.1%	15 0.0%	143 0.5%	20 0.1%	128 0.4%	32 0.1%	14.0% 86.0%
writeLPA	1 0.0%	57 0.2%	1 0.0%	151 0.5%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	8 0.0%	113 0.4%	4 0.0%	6 0.0%	37 0.1%	31 0.1%	3 0.0%	800 2.6%	66.0% 34.0%
	65.1% 34.9%	47.6% 52.4%	43.0% 57.0%	47.0% 53.0%	35.1% 64.9%	27.0% 73.0%	51.6% 48.4%	43.7% 56.3%	58.7% 41.3%	43.2% 56.8%	65.3% 34.7%	56.6% 43.4%	38.5% 61.5%	24.4% 75.6%	20.8% 79.2%	69.0% 31.0%	49.9% 50.1%

Figure C.33: 0.33 second epoch, LN accelerometer, Hip

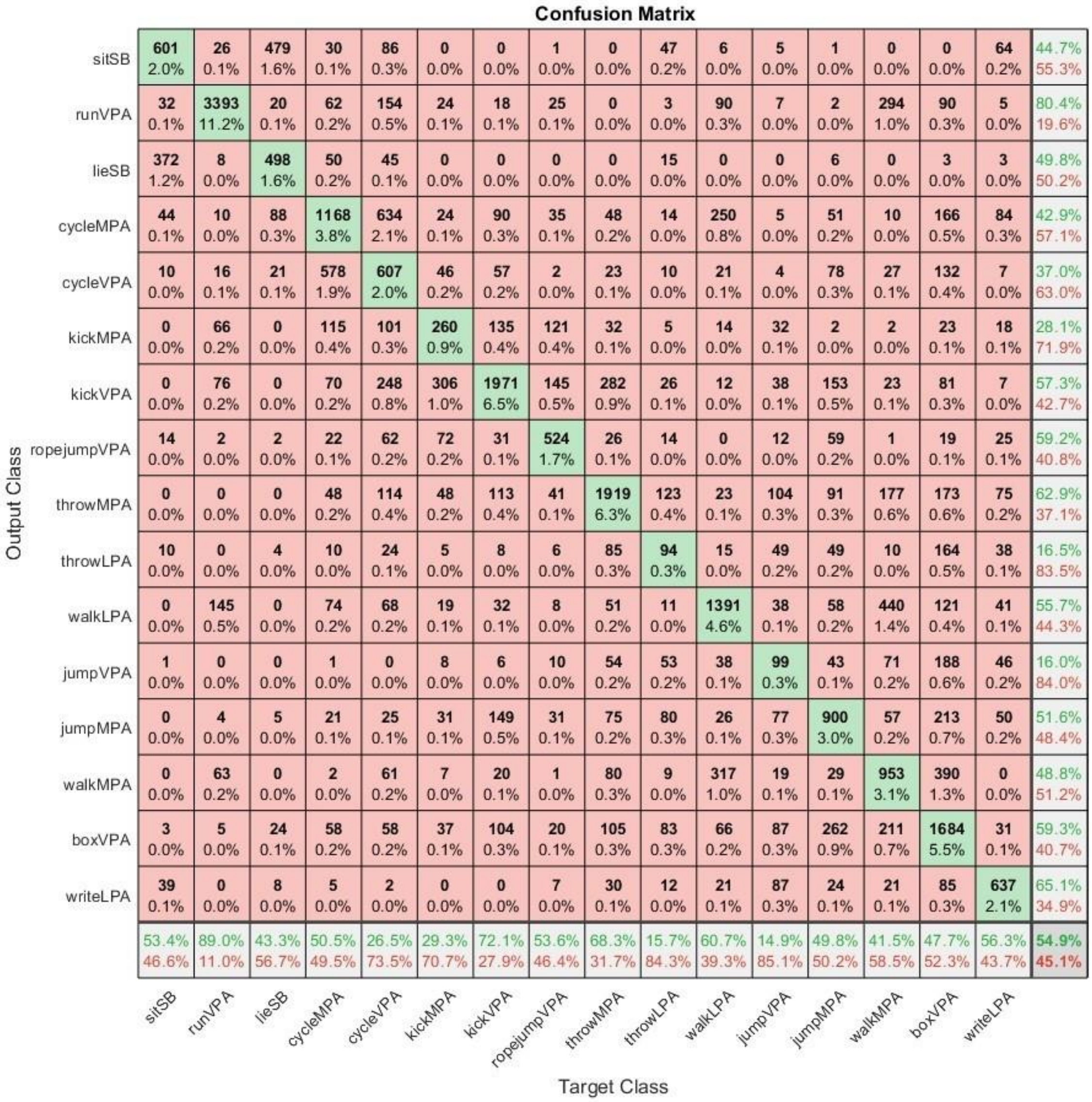


Figure C.34: 0.33 second epoch, LN accelerometer, Right wrist

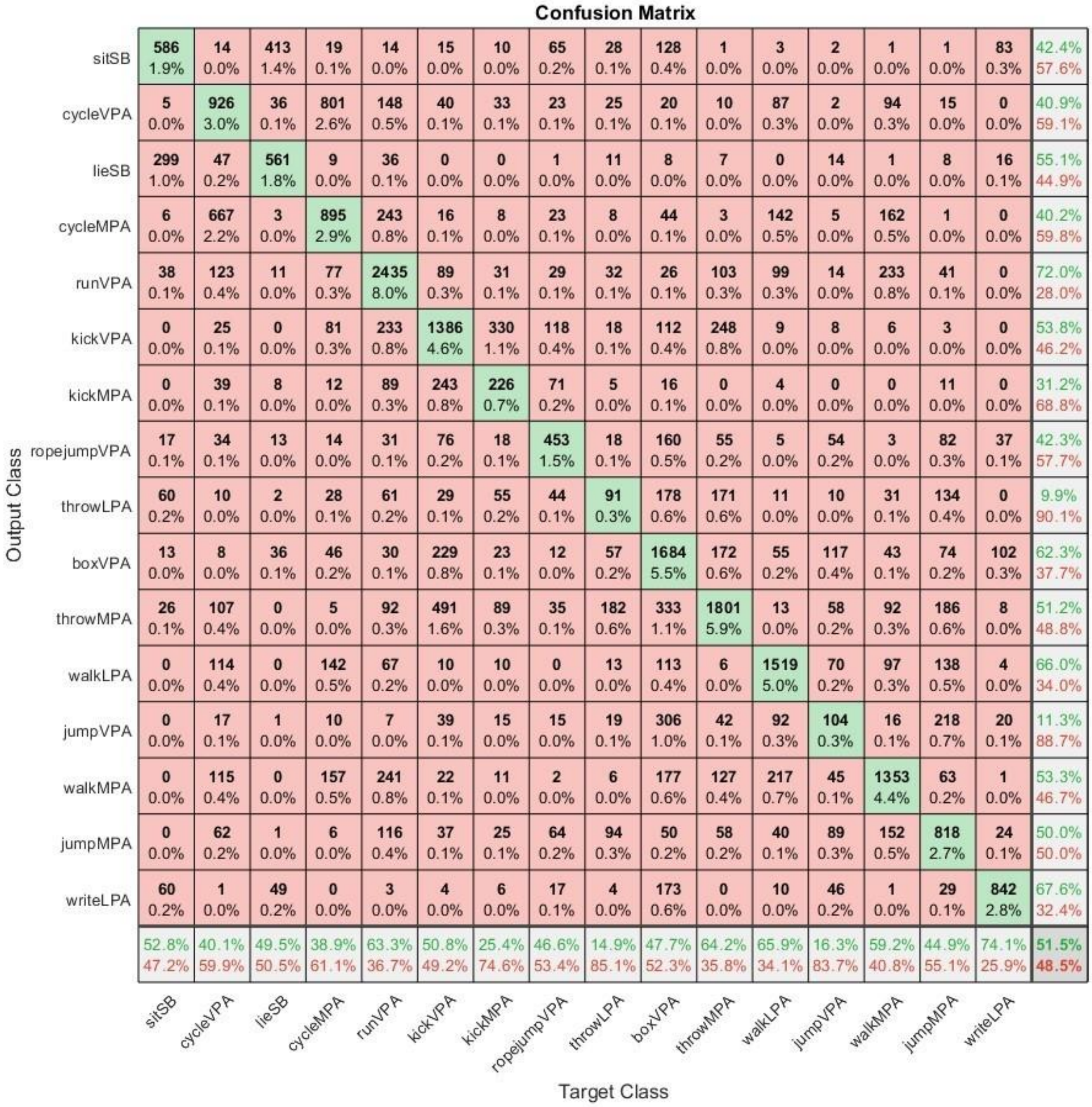


Figure C.35: 0.33 second epoch, WR accelerometer, Hip

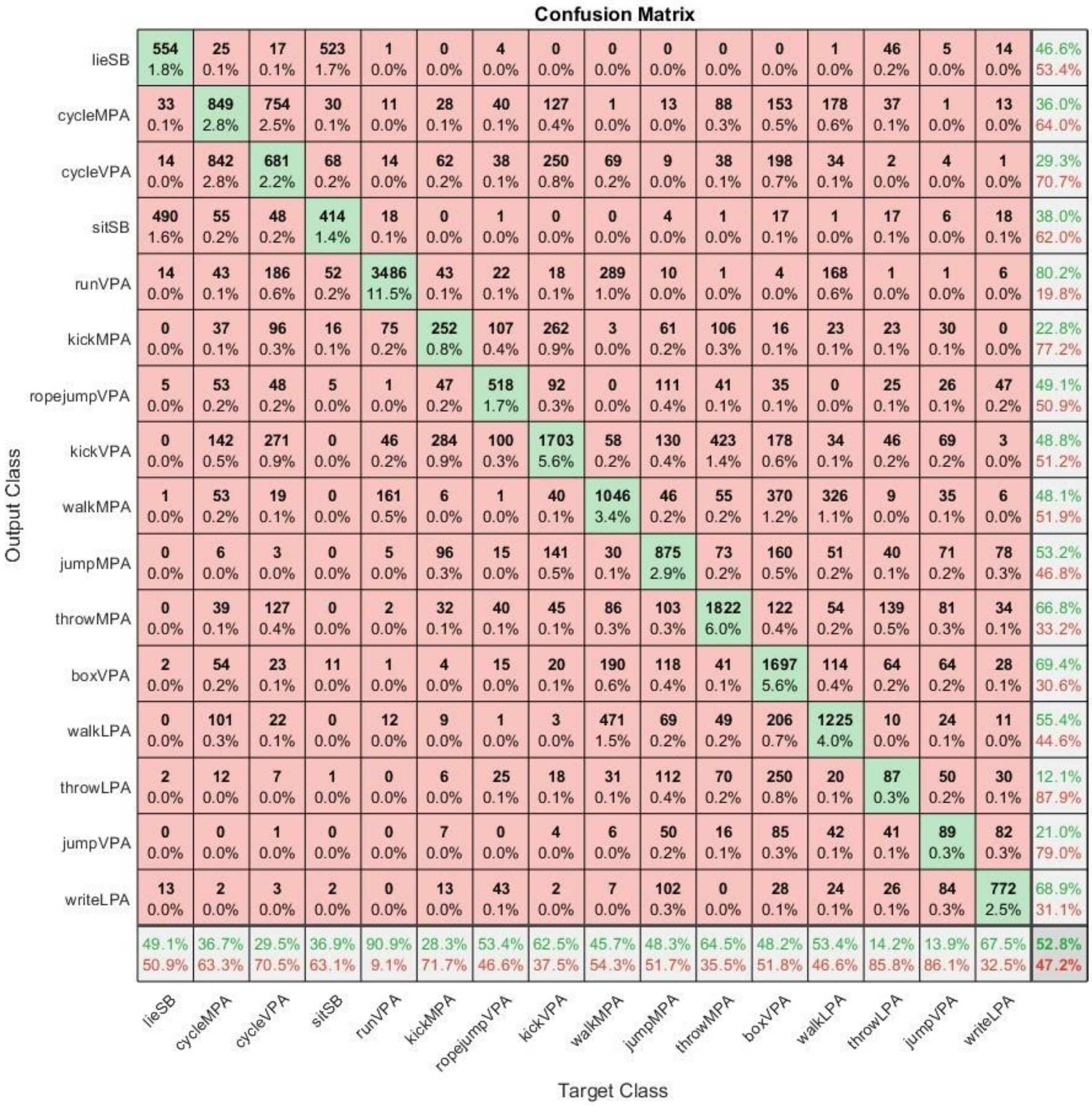


Figure C.36: 0.33 second epoch, WR accelerometer, Right wrist

C.3.3. Hip and right wrist: "Unknown" class with KNN algorithm

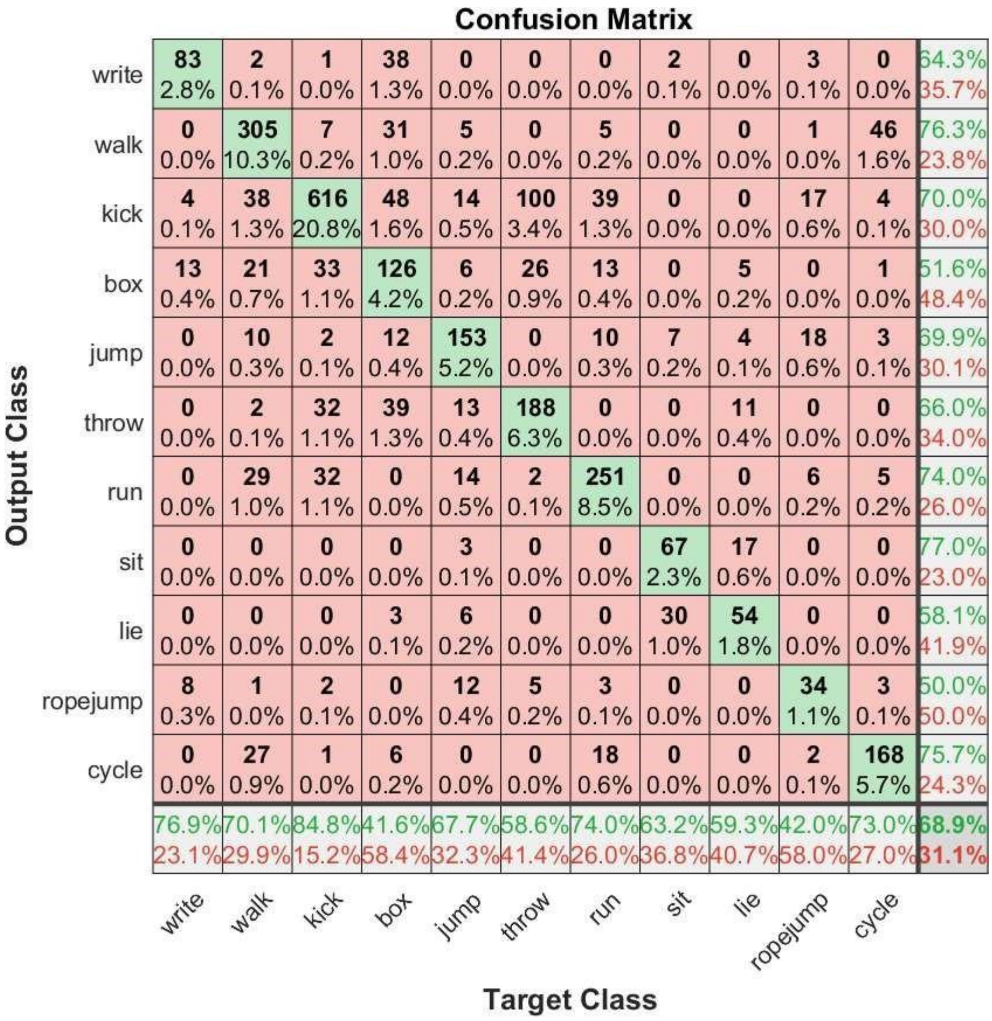


Figure C.37: 0.33 second epoch, LN accelerometer, Hip, Unknown with KNN

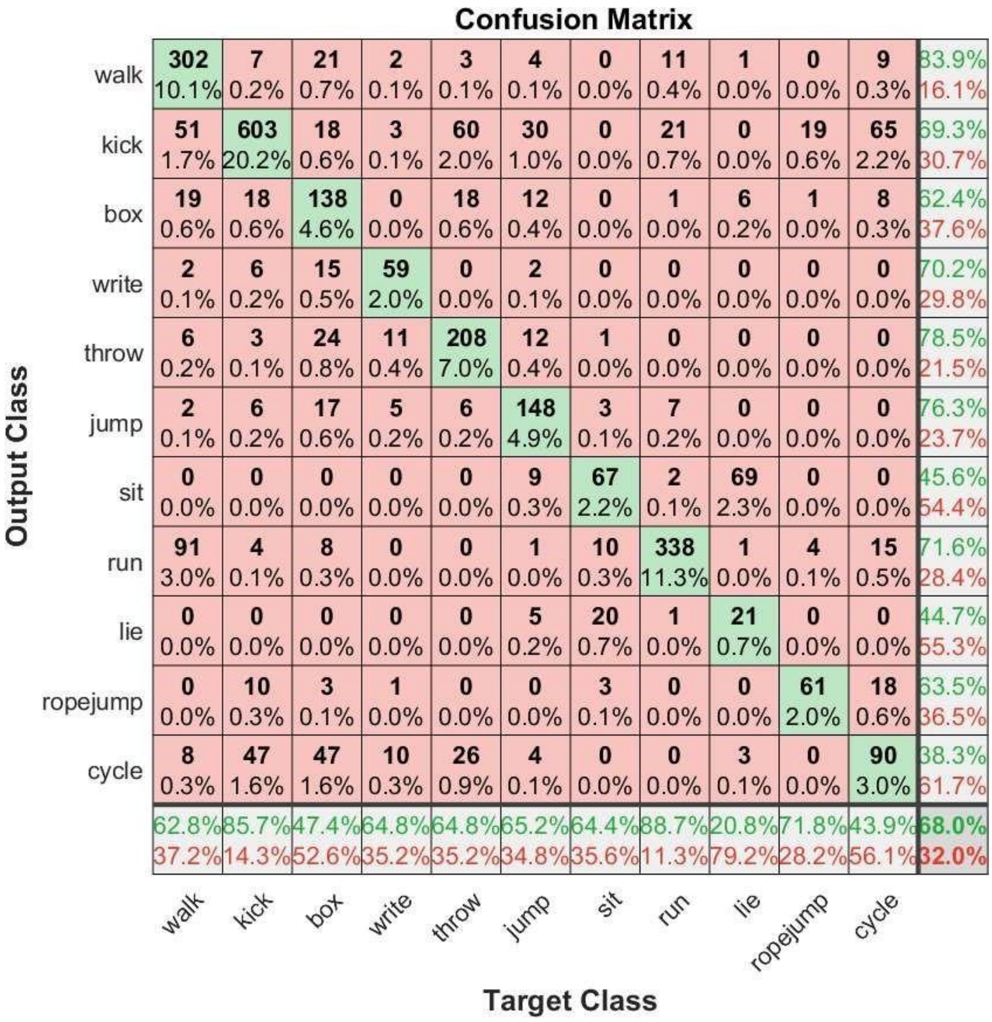


Figure C.38: 0.33 second epoch, LN accelerometer, Right wrist, Unknown with KNN

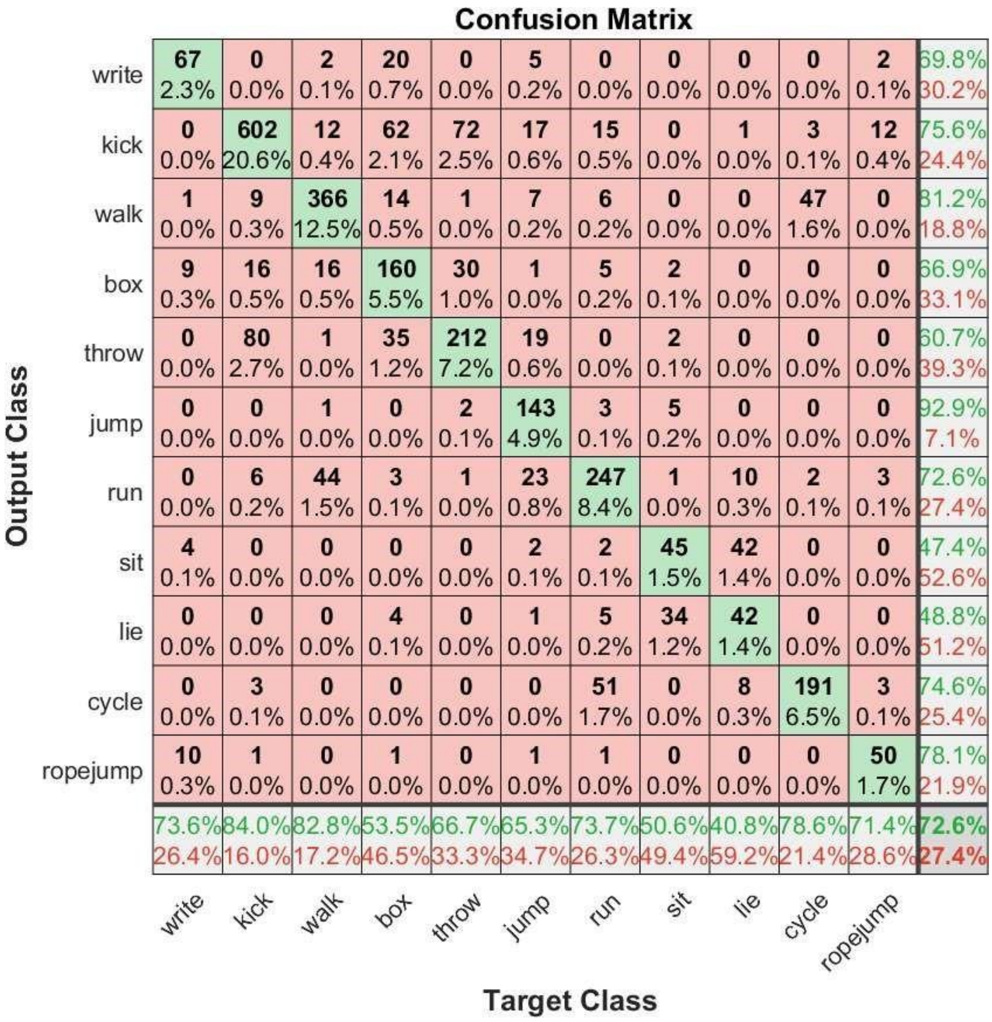


Figure C.39: 0.33 second epoch, WR accelerometer, Hip, Unknown with KNN

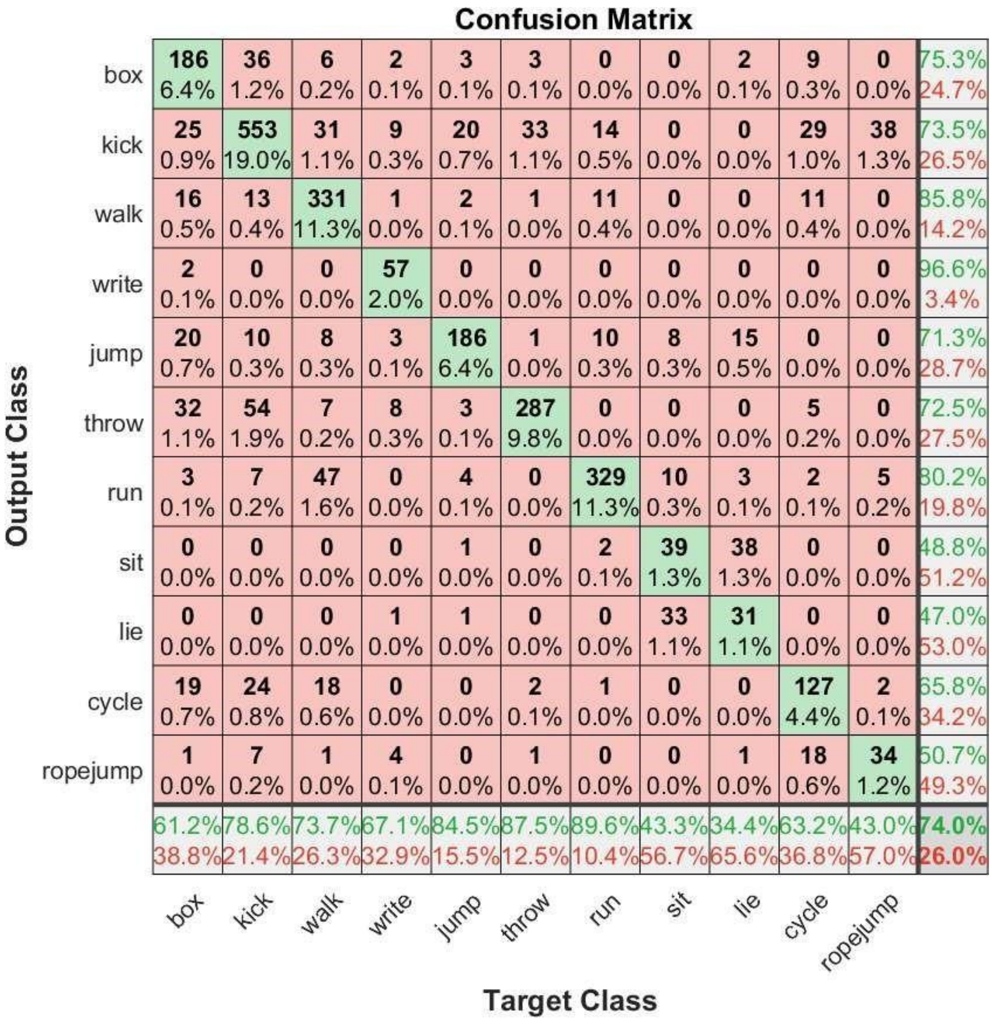


Figure C.40: 0.33 second epoch, WR accelerometer, Right wrist, Unknown with KNN

C.3.4. Hip and right wrist: ”Unknown” class with threshold

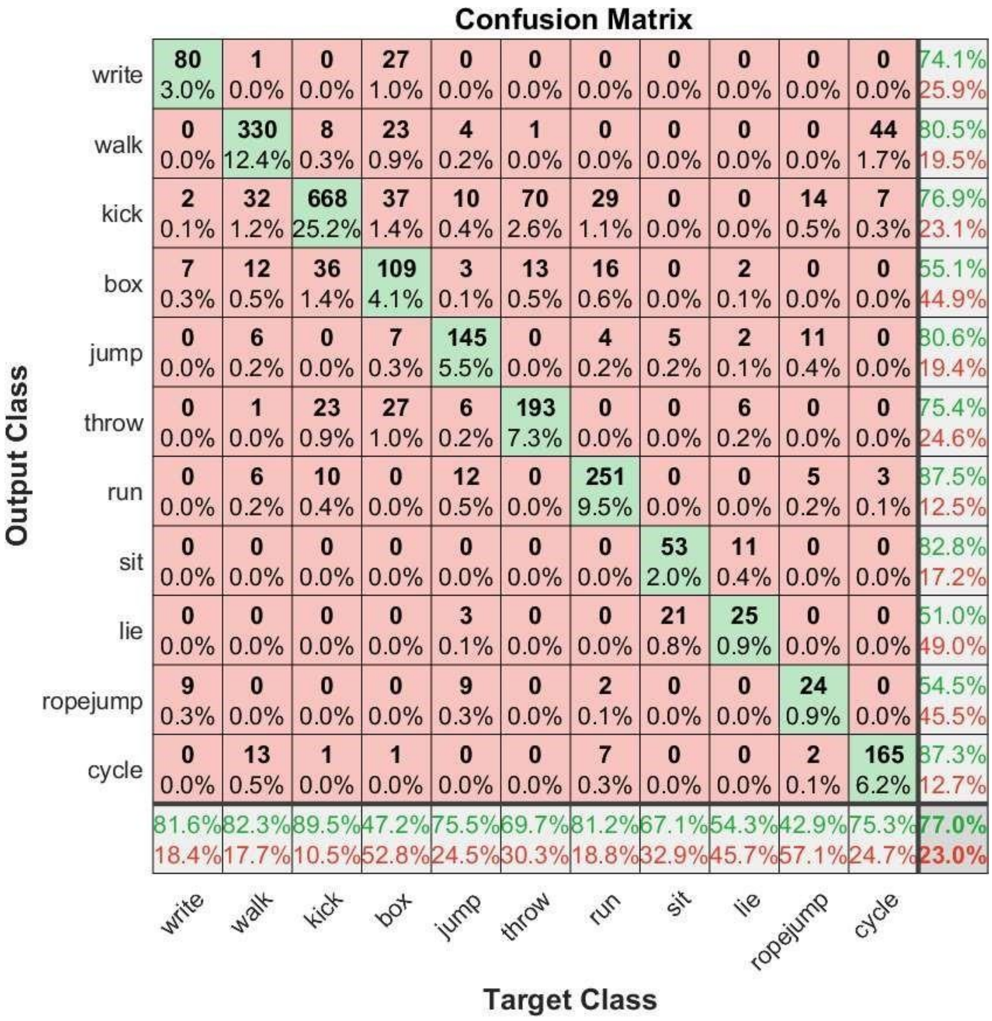


Figure C.41: 0.33 second epoch, LN accelerometer, Hip, Unknown with threshold at 0.775

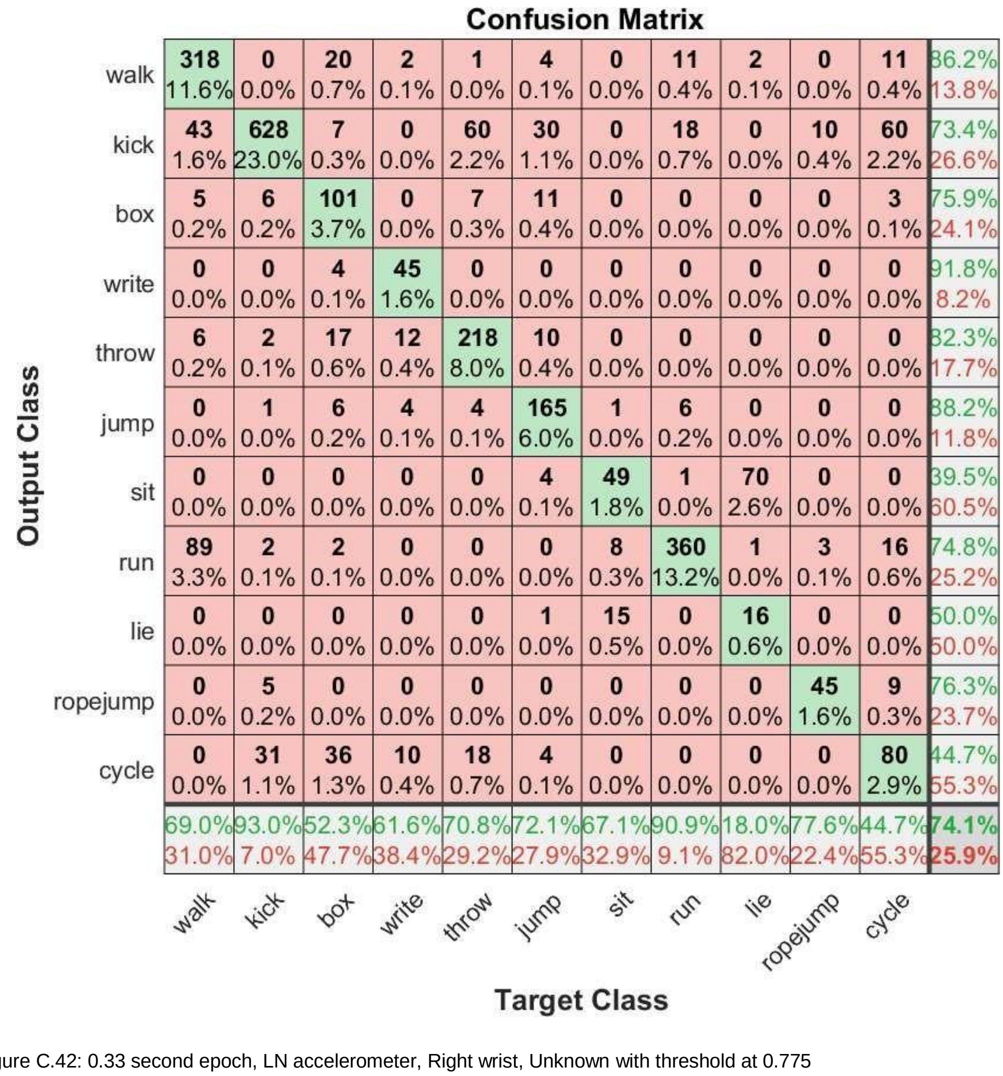


Figure C.42: 0.33 second epoch, LN accelerometer, Right wrist, Unknown with threshold at 0.775

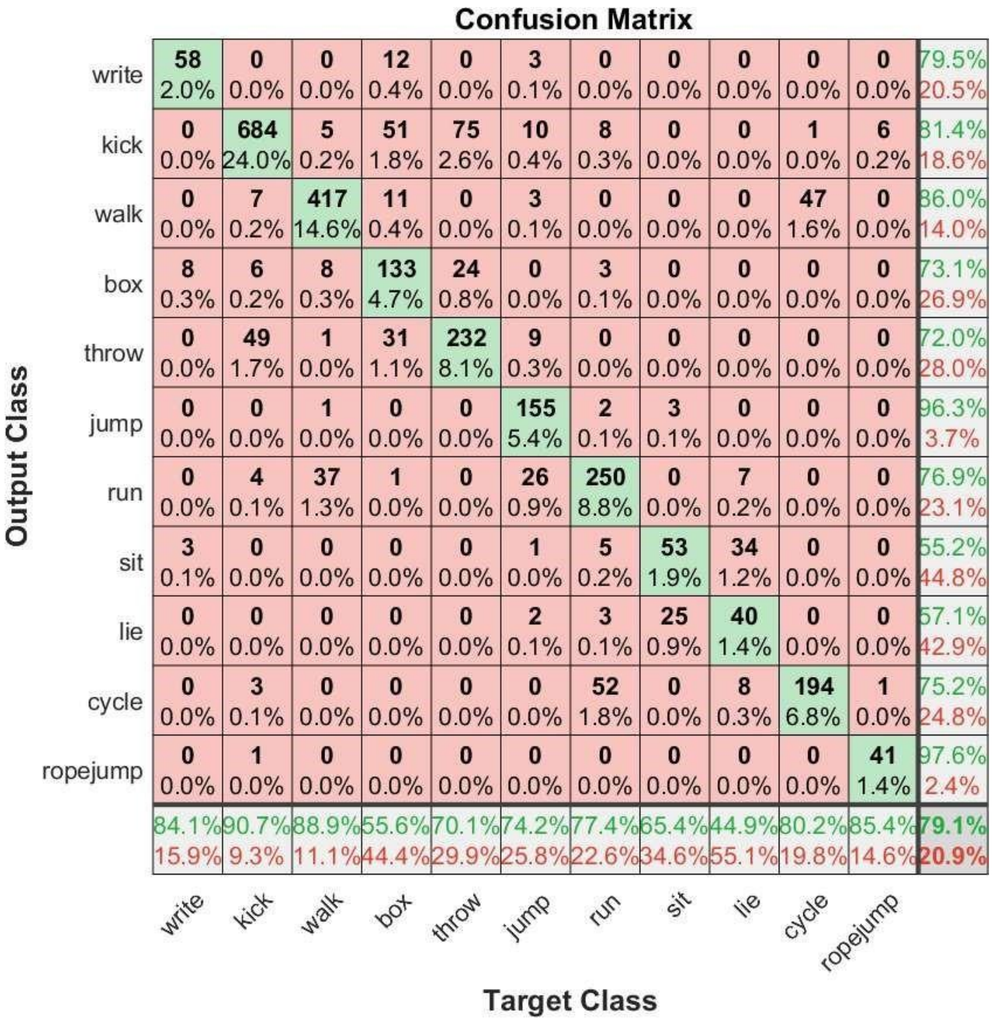


Figure C.43: 0.33 second epoch, WR accelerometer, Hip, Unknown with threshold at 0.775

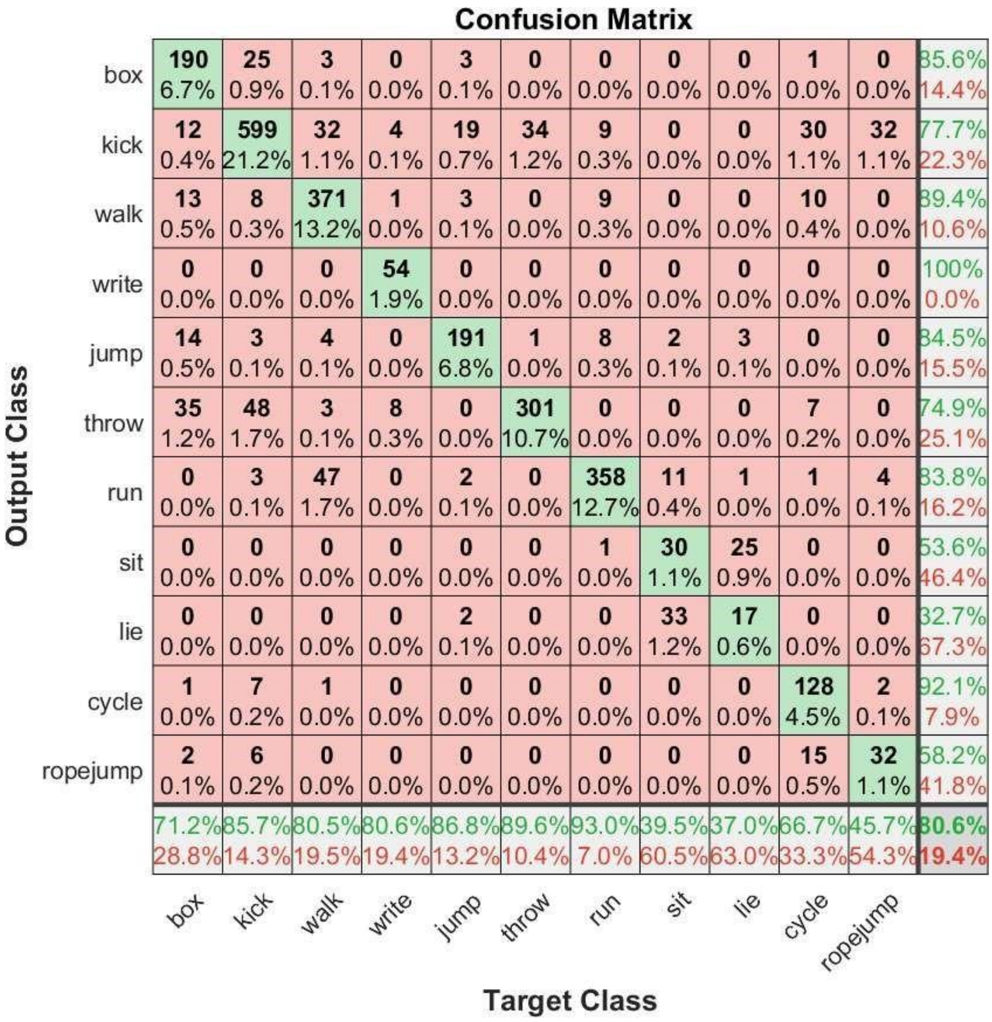


Figure C.44: 0.33 second epoch, WR accelerometer, Right wrist, Unknown with threshold at 0.775

C.4. 0.25 Second Epoch

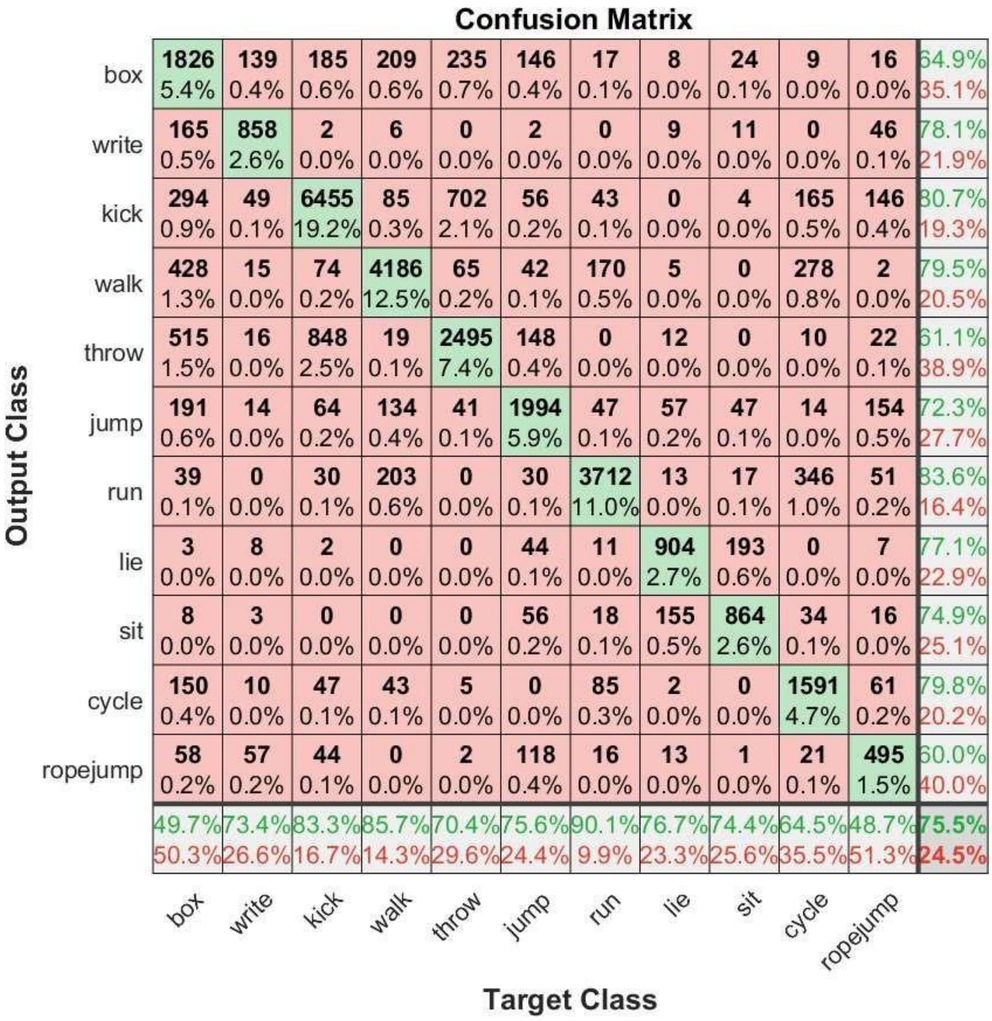


Figure C.45: 0.25 second epoch, LN accelerometer, Hip

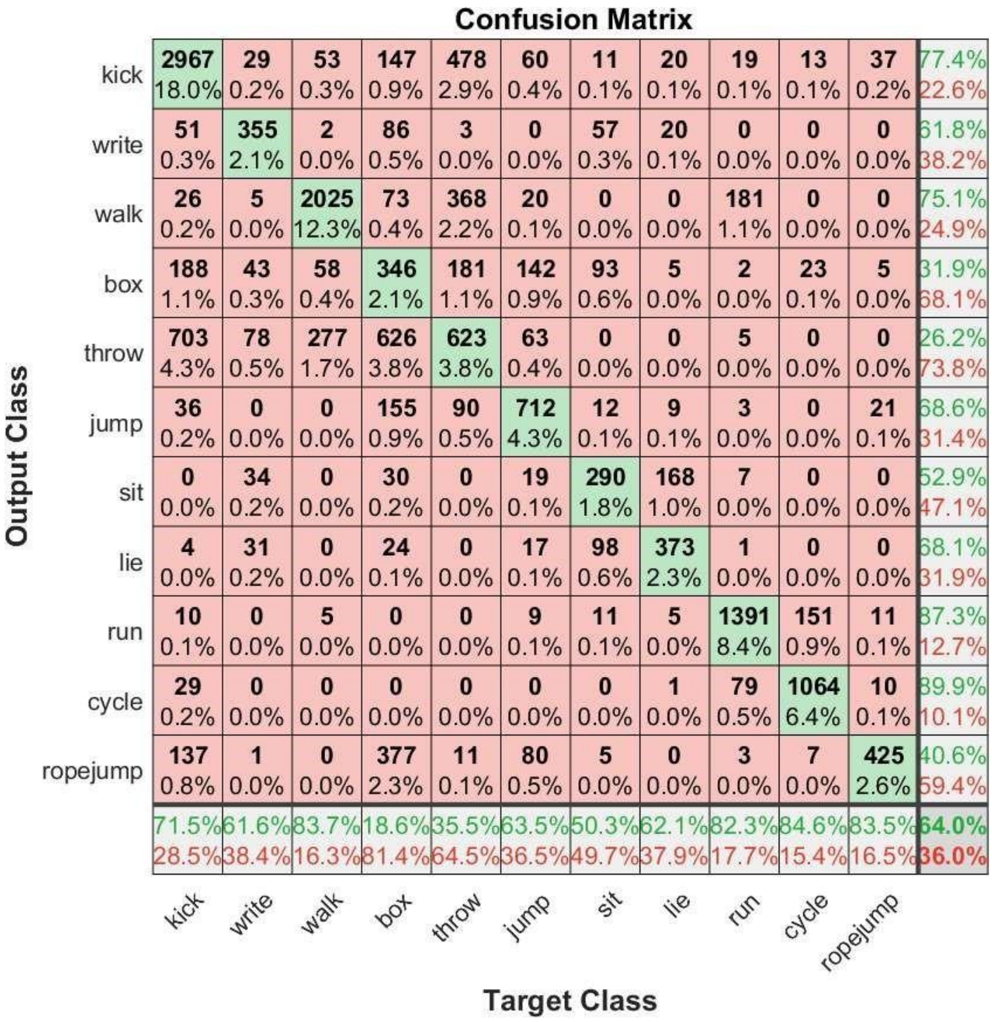


Figure C.46: 0.25 second epoch, LN accelerometer, Left ankle

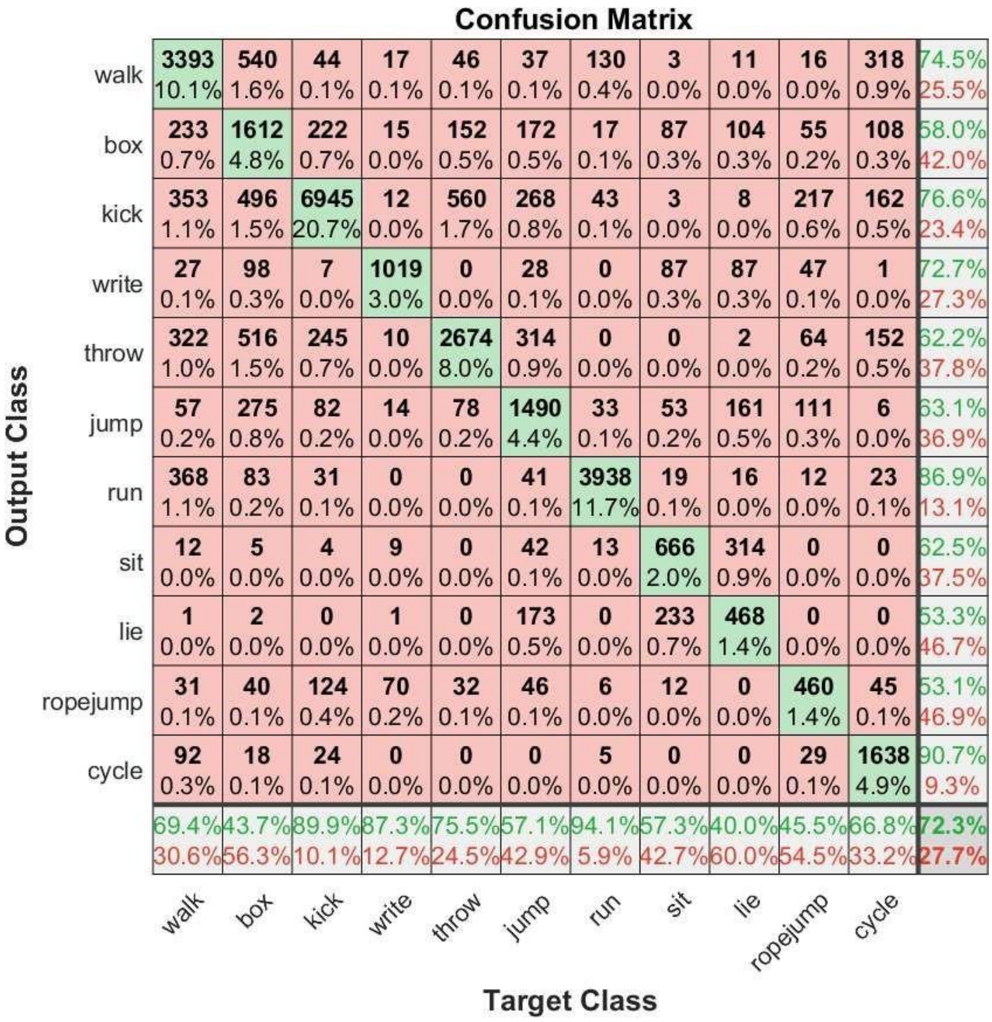


Figure C.47: 0.25 second epoch, LN accelerometer, Left wrist

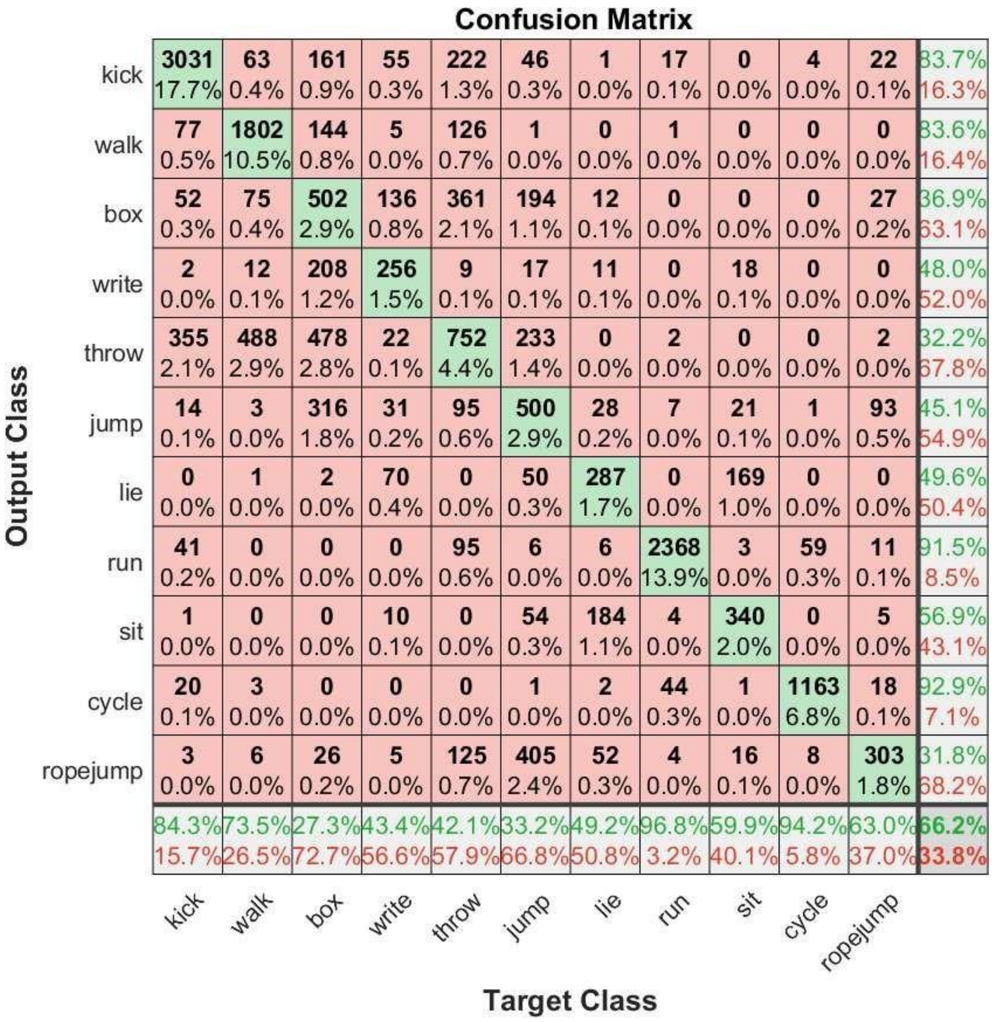


Figure C.48: 0.25 second epoch, LN accelerometer, Right ankle

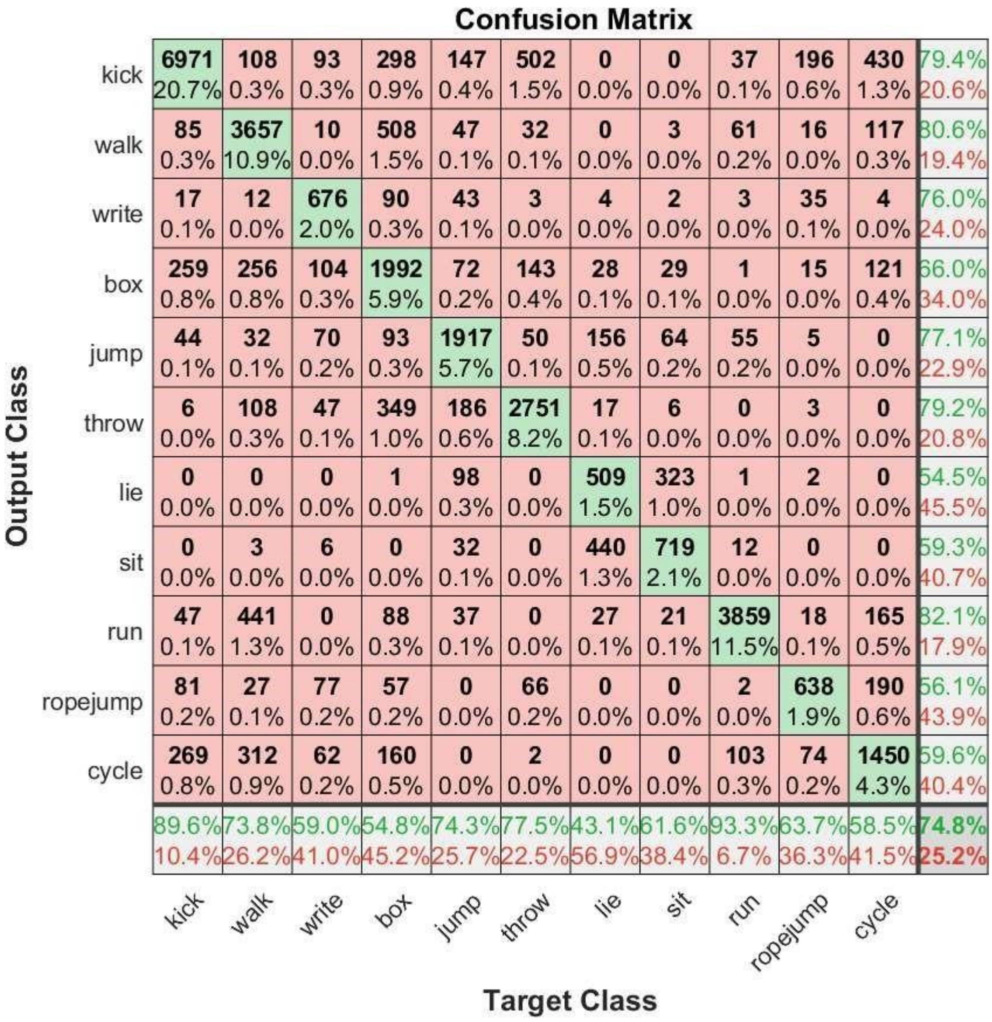


Figure C.49: 0.25 second epoch, LN accelerometer, Right wrist

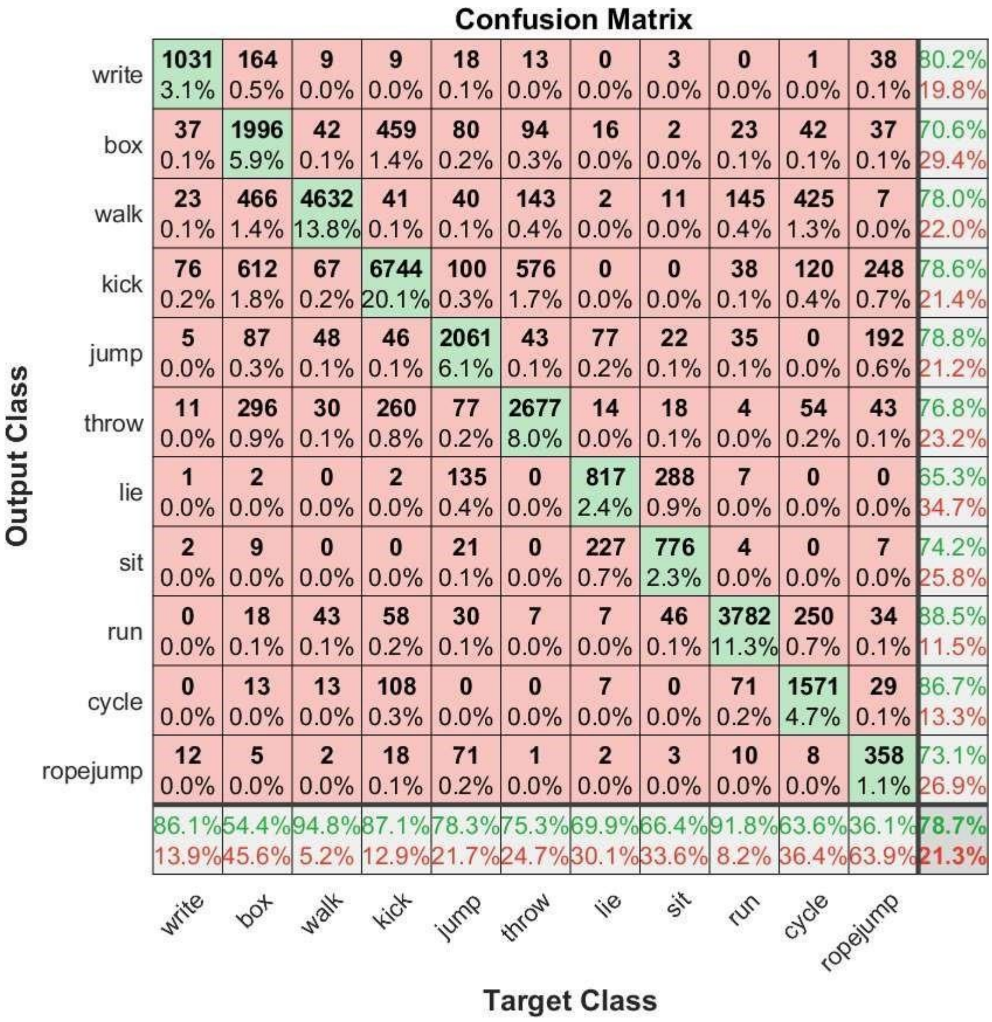


Figure C.50: 0.25 second epoch, WR accelerometer, Hip

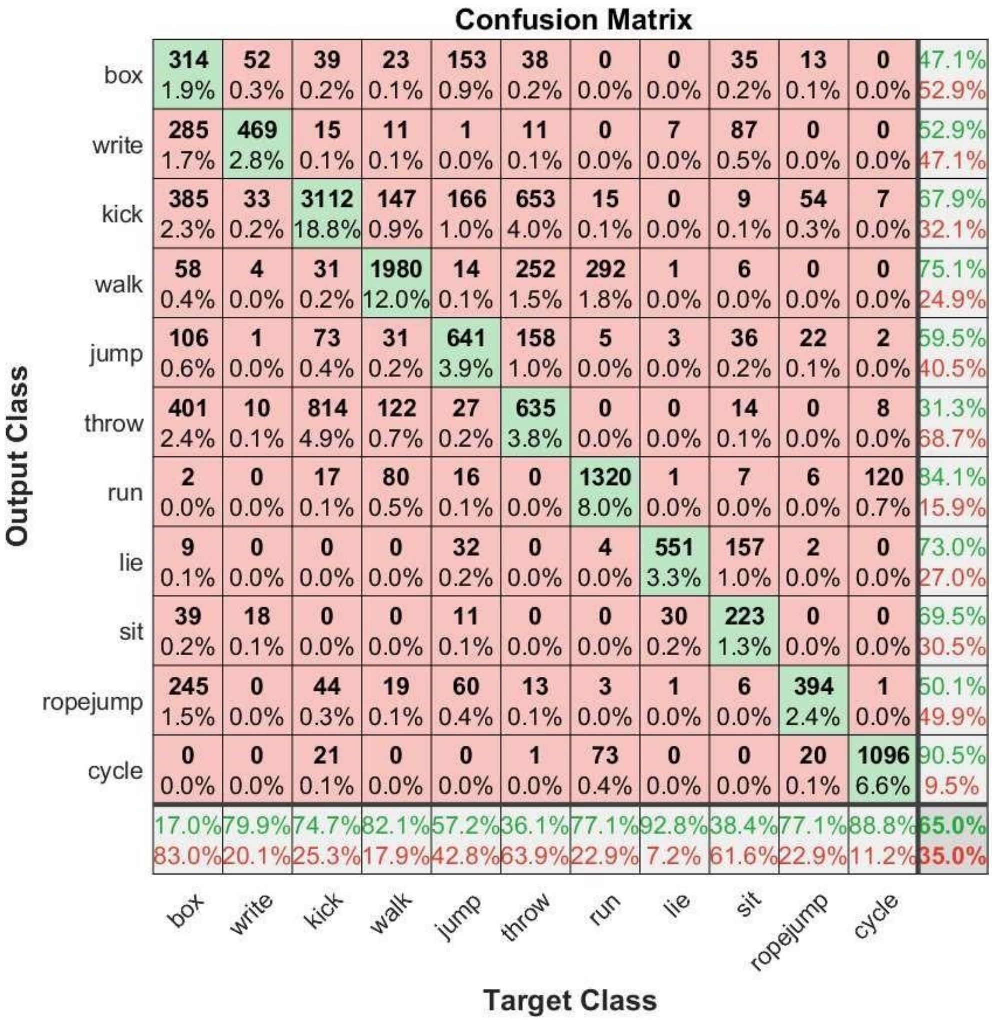


Figure C.51: 0.25 second epoch, WR accelerometer, Left ankle

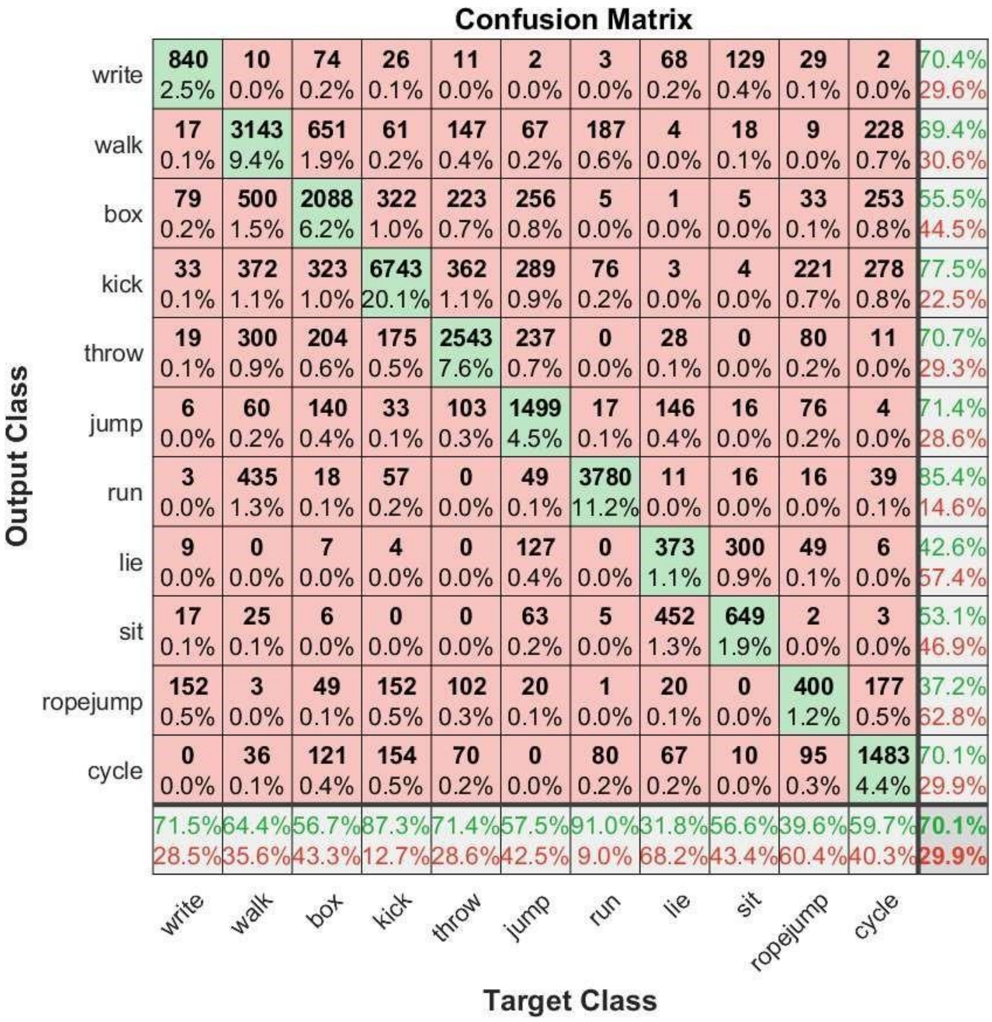


Figure C.52: 0.25 second epoch, WR accelerometer, Left wrist

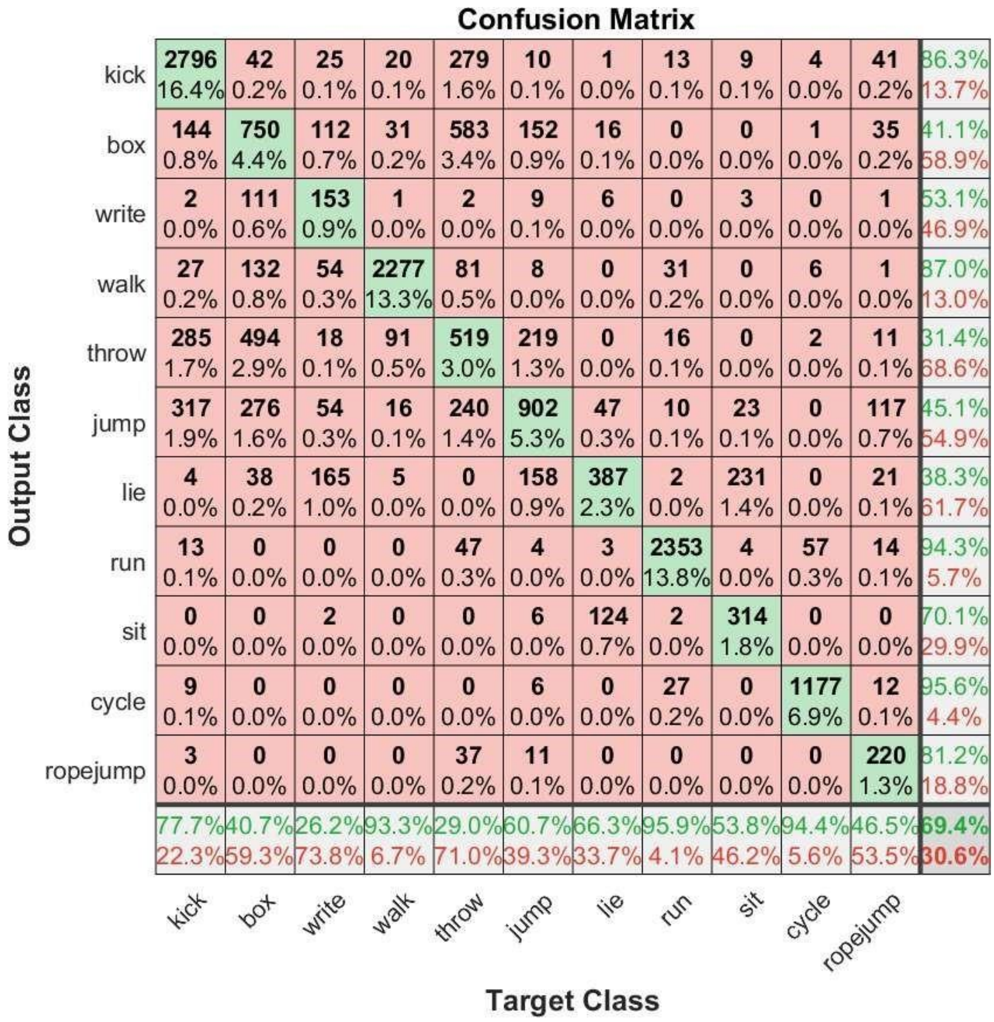
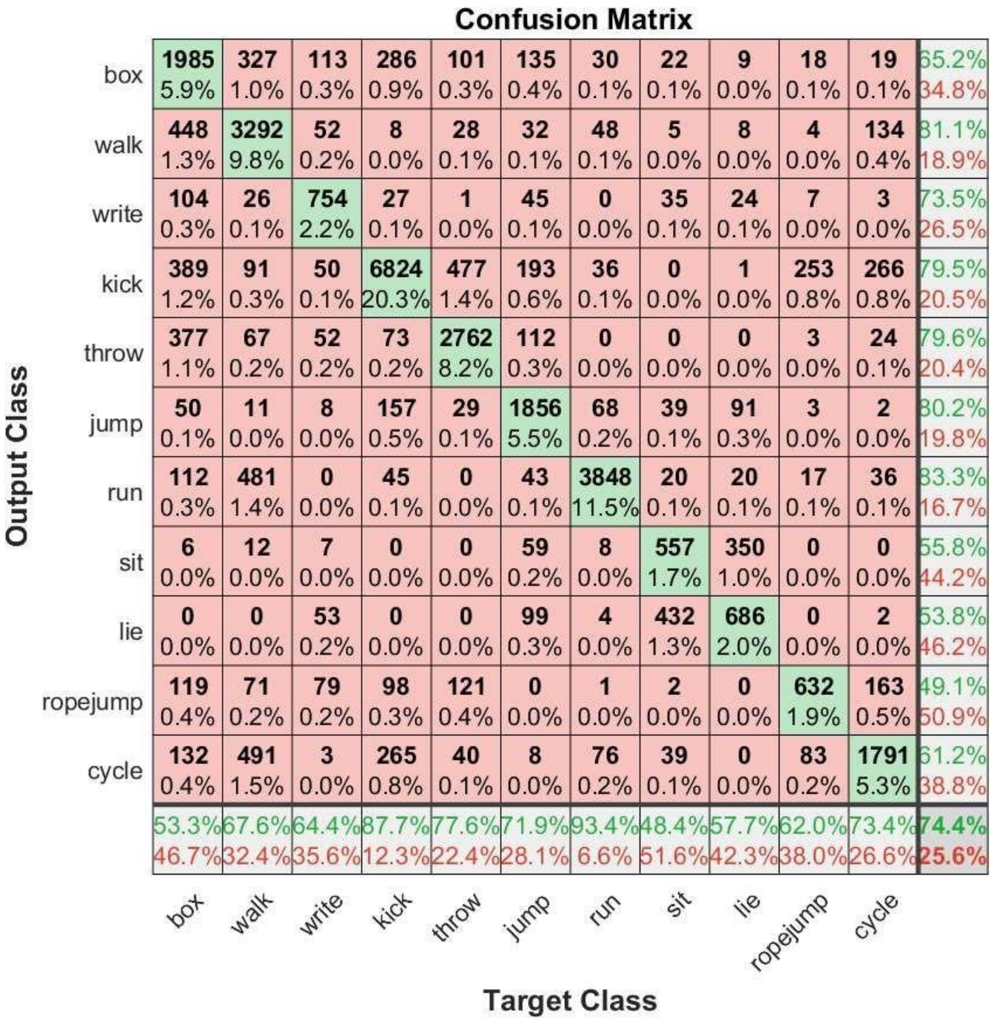


Figure C.53: 0.25 second epoch, WR accelerometer, Right ankle



	box	walk	write	kick	throw	jump	run	sit	lie	ropejump	cycle	
box	1985 5.9%	327 1.0%	113 0.3%	286 0.9%	101 0.3%	135 0.4%	30 0.1%	22 0.1%	9 0.0%	18 0.1%	19 0.1%	65.2% 34.8%
walk	448 1.3%	3292 9.8%	52 0.2%	8 0.0%	28 0.1%	32 0.1%	48 0.1%	5 0.0%	8 0.0%	4 0.0%	134 0.4%	81.1% 18.9%
write	104 0.3%	26 0.1%	754 2.2%	27 0.1%	1 0.0%	45 0.1%	0 0.0%	35 0.1%	24 0.1%	7 0.0%	3 0.0%	73.5% 26.5%
kick	389 1.2%	91 0.3%	50 0.1%	6824 20.3%	477 1.4%	193 0.6%	36 0.1%	0 0.0%	1 0.0%	253 0.8%	266 0.8%	79.5% 20.5%
throw	377 1.1%	67 0.2%	52 0.2%	73 0.2%	2762 8.2%	112 0.3%	0 0.0%	0 0.0%	0 0.0%	3 0.0%	24 0.1%	79.6% 20.4%
jump	50 0.1%	11 0.0%	8 0.0%	157 0.5%	29 0.1%	1856 5.5%	68 0.2%	39 0.1%	91 0.3%	3 0.0%	2 0.0%	80.2% 19.8%
run	112 0.3%	481 1.4%	0 0.0%	45 0.1%	0 0.0%	43 0.1%	3848 11.5%	20 0.1%	20 0.1%	17 0.1%	36 0.1%	83.3% 16.7%
sit	6 0.0%	12 0.0%	7 0.0%	0 0.0%	0 0.0%	59 0.2%	8 0.0%	557 1.7%	350 1.0%	0 0.0%	0 0.0%	55.8% 44.2%
lie	0 0.0%	0 0.0%	53 0.2%	0 0.0%	0 0.0%	99 0.3%	4 0.0%	432 1.3%	686 2.0%	0 0.0%	2 0.0%	53.8% 46.2%
ropejump	119 0.4%	71 0.2%	79 0.2%	98 0.3%	121 0.4%	0 0.0%	1 0.0%	2 0.0%	0 0.0%	632 1.9%	163 0.5%	49.1% 50.9%
cycle	132 0.4%	491 1.5%	3 0.0%	265 0.8%	40 0.1%	8 0.0%	76 0.2%	39 0.1%	0 0.0%	83 0.2%	1791 5.3%	61.2% 38.8%
	53.3% 46.7%	67.6% 32.4%	64.4% 35.6%	87.7% 12.3%	77.6% 22.4%	71.9% 28.1%	93.4% 6.6%	48.4% 51.6%	57.7% 42.3%	62.0% 38.0%	73.4% 26.6%	74.4% 25.6%

Figure C.54: 0.25 second epoch, WR accelerometer, Right wrist

C.4.1. Hip and right wrist: 2 placements

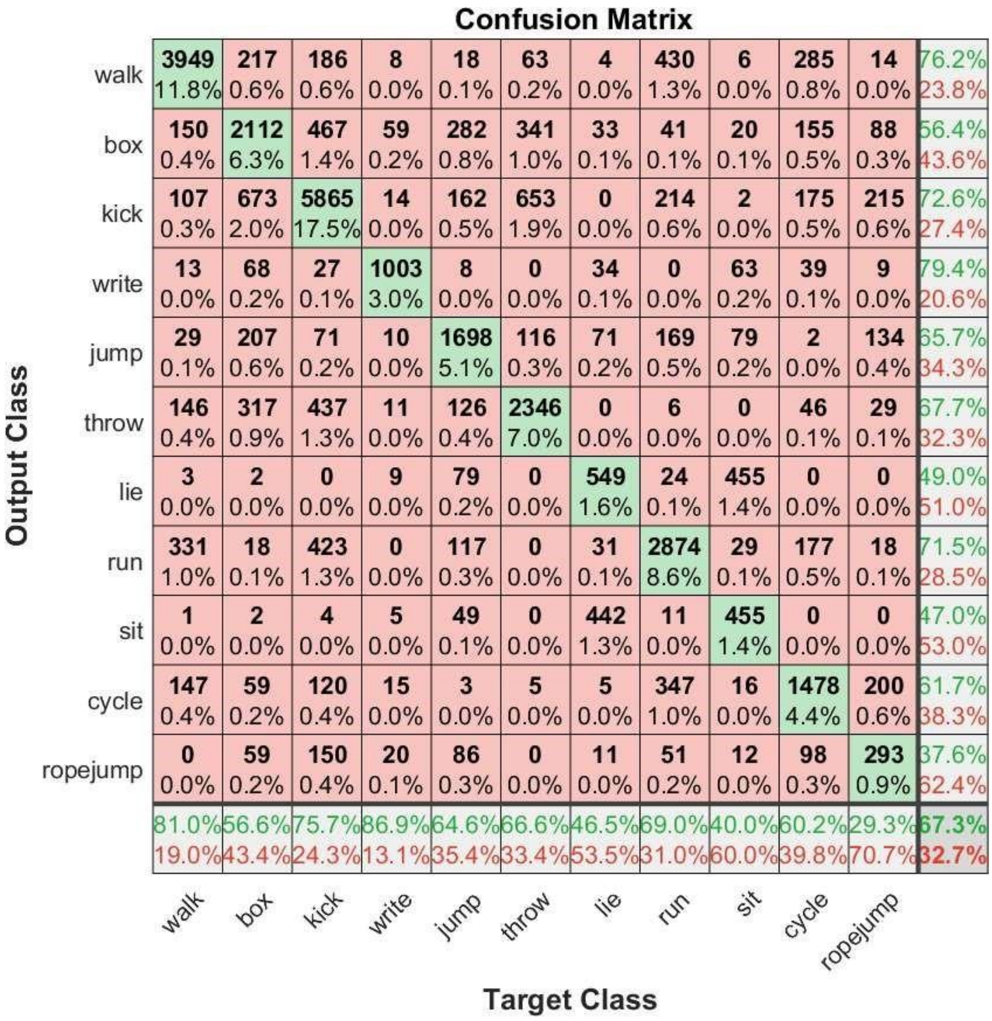


Figure C.55: 0.25 second epoch, LN accelerometer, Hip & Right wrist

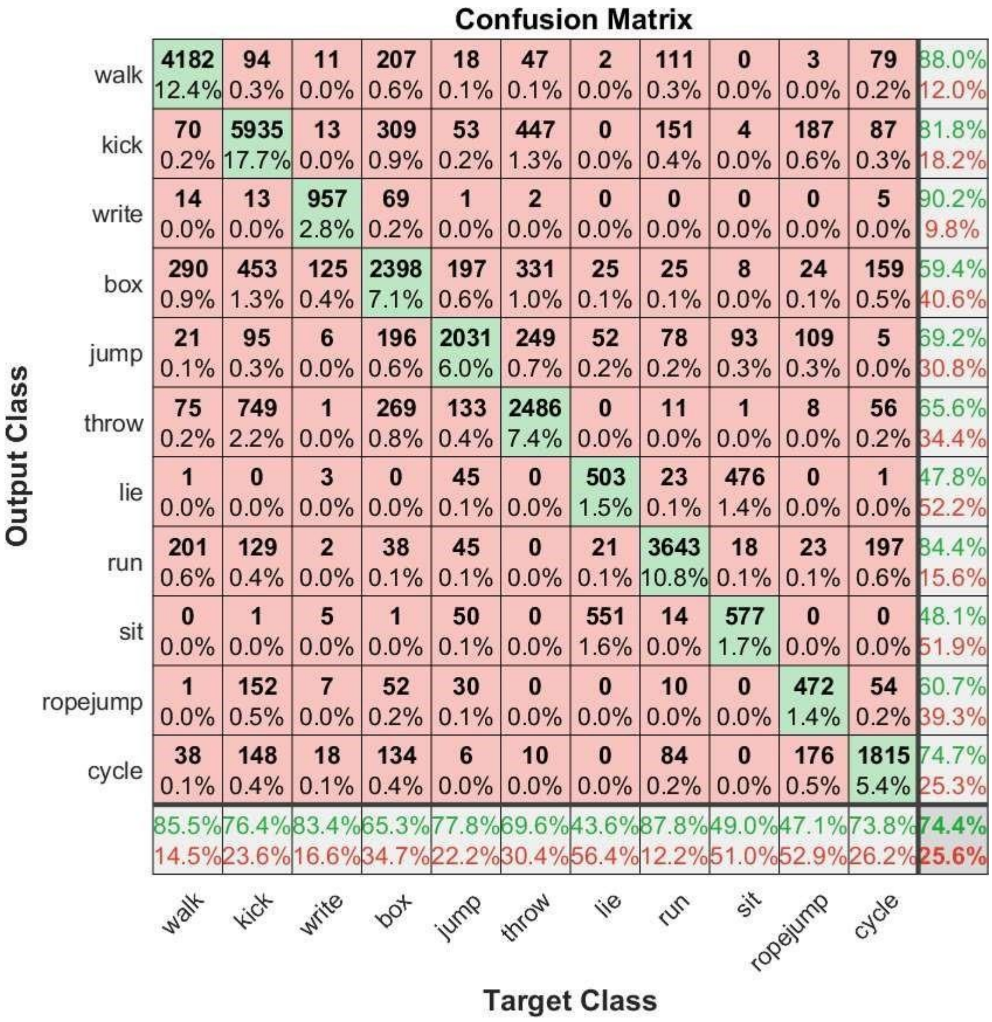


Figure C.56: 0.25 second epoch, WR accelerometer, Hip & Right wrist

C.4.2. Hip and right wrist: Activity - Intensity

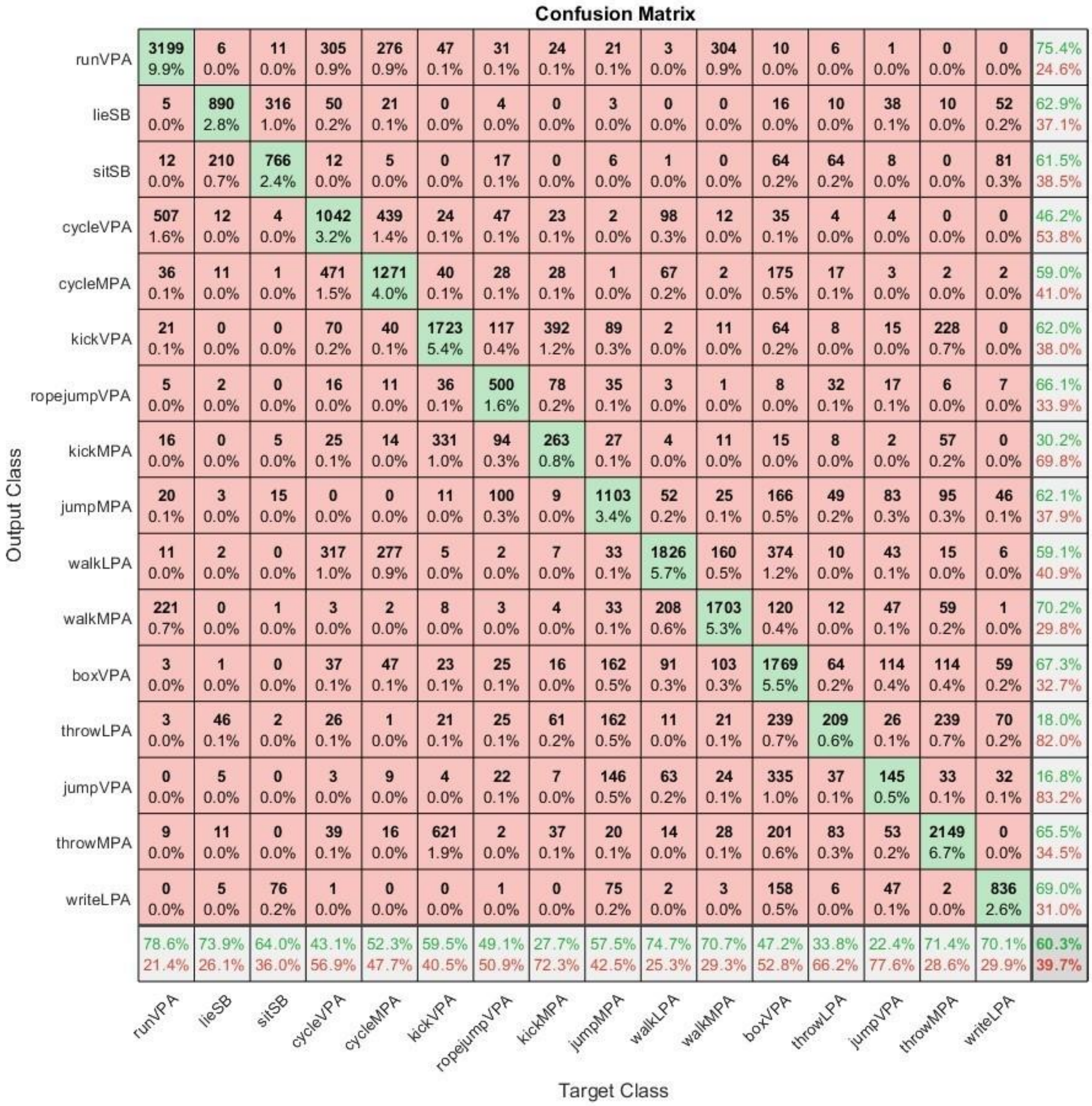


Figure C.57: 0.25 second epoch, LN accelerometer, Hip

Confusion Matrix

Output Class \ Target Class	cycleVPA	sitSB	lieSB	cycleMPA	runVPA	kickVPA	ropejumpVPA	kickMPA	jumpMPA	boxVPA	throwMPA	throwLPA	walkLPA	jumpVPA	walkMPA	writeLPA	
cycleVPA	819 2.5%	14 0.0%	40 0.1%	851 2.6%	86 0.3%	42 0.1%	23 0.1%	47 0.1%	1 0.0%	135 0.4%	43 0.1%	0 0.0%	181 0.6%	10 0.0%	18 0.1%	28 0.1%	35.0% 65.0%
sitSB	19 0.1%	606 1.9%	452 1.4%	36 0.1%	0 0.0%	0 0.0%	3 0.0%	0 0.0%	0 0.0%	3 0.0%	0 0.0%	6 0.0%	1 0.0%	2 0.0%	0 0.0%	2 0.0%	53.6% 46.4%
lieSB	28 0.1%	458 1.4%	533 1.7%	21 0.1%	4 0.0%	0 0.0%	0 0.0%	0 0.0%	33 0.1%	14 0.0%	13 0.0%	23 0.1%	4 0.0%	8 0.0%	0 0.0%	8 0.0%	46.5% 53.5%
cycleMPA	758 2.4%	35 0.1%	19 0.1%	790 2.5%	32 0.1%	162 0.5%	63 0.2%	41 0.1%	7 0.0%	96 0.3%	13 0.0%	1 0.0%	33 0.1%	8 0.0%	1 0.0%	1 0.0%	38.3% 61.7%
runVPA	313 1.0%	10 0.0%	22 0.1%	139 0.4%	3719 11.6%	19 0.1%	15 0.0%	24 0.1%	3 0.0%	10 0.0%	2 0.0%	6 0.0%	5 0.0%	4 0.0%	399 1.2%	2 0.0%	79.3% 20.7%
kickVPA	58 0.2%	0 0.0%	0 0.0%	163 0.5%	46 0.1%	1908 5.9%	161 0.5%	392 1.2%	16 0.0%	24 0.1%	86 0.3%	13 0.0%	0 0.0%	10 0.0%	3 0.0%	0 0.0%	66.3% 33.8%
ropejumpVPA	127 0.4%	8 0.0%	6 0.0%	60 0.2%	0 0.0%	92 0.3%	566 1.8%	122 0.4%	8 0.0%	31 0.1%	128 0.4%	9 0.0%	6 0.0%	32 0.1%	10 0.0%	15 0.0%	46.4% 53.6%
kickMPA	83 0.3%	0 0.0%	15 0.0%	145 0.5%	19 0.1%	435 1.4%	135 0.4%	179 0.6%	9 0.0%	20 0.1%	37 0.1%	12 0.0%	1 0.0%	3 0.0%	4 0.0%	0 0.0%	16.3% 83.7%
jumpMPA	18 0.1%	0 0.0%	36 0.1%	15 0.0%	5 0.0%	134 0.4%	2 0.0%	67 0.2%	1219 3.8%	97 0.3%	67 0.2%	69 0.2%	22 0.1%	42 0.1%	40 0.1%	69 0.2%	64.1% 35.9%
boxVPA	39 0.1%	11 0.0%	27 0.1%	99 0.3%	2 0.0%	74 0.2%	5 0.0%	30 0.1%	256 0.8%	2033 6.3%	121 0.4%	67 0.2%	96 0.3%	119 0.4%	179 0.6%	70 0.2%	63.0% 37.0%
throwMPA	60 0.2%	0 0.0%	0 0.0%	6 0.0%	0 0.0%	20 0.1%	30 0.1%	33 0.1%	85 0.3%	172 0.5%	2202 6.8%	64 0.2%	11 0.0%	64 0.2%	62 0.2%	4 0.0%	78.3% 21.7%
throwLPA	0 0.0%	10 0.0%	24 0.1%	26 0.1%	0 0.0%	0 0.0%	15 0.0%	10 0.0%	63 0.2%	164 0.5%	115 0.4%	159 0.5%	14 0.0%	63 0.2%	126 0.4%	125 0.4%	17.4% 82.6%
walkLPA	20 0.1%	3 0.0%	0 0.0%	54 0.2%	10 0.0%	6 0.0%	3 0.0%	1 0.0%	21 0.1%	254 0.8%	22 0.1%	11 0.0%	1805 5.6%	36 0.1%	467 1.5%	14 0.0%	66.2% 33.8%
jumpVPA	2 0.0%	0 0.0%	0 0.0%	3 0.0%	0 0.0%	2 0.0%	1 0.0%	4 0.0%	80 0.2%	235 0.7%	68 0.2%	35 0.1%	32 0.1%	120 0.4%	25 0.1%	83 0.3%	17.4% 82.6%
walkMPA	18 0.1%	0 0.0%	1 0.0%	13 0.0%	123 0.4%	0 0.0%	2 0.0%	1 0.0%	33 0.1%	418 1.3%	60 0.2%	29 0.1%	227 0.7%	48 0.1%	1069 3.3%	14 0.0%	52.0% 48.0%
writeLPA	50 0.2%	52 0.2%	13 0.0%	19 0.1%	1 0.0%	0 0.0%	0 0.0%	7 0.0%	65 0.2%	62 0.2%	5 0.0%	150 0.5%	12 0.0%	66 0.2%	2 0.0%	764 2.4%	60.3% 39.7%
	34.0% 66.0%	50.2% 49.8%	44.9% 55.1%	32.4% 67.6%	91.9% 8.1%	65.9% 34.1%	55.3% 44.7%	18.7% 81.3%	64.2% 35.8%	54.0% 46.0%	73.8% 26.2%	24.3% 75.7%	73.7% 26.3%	18.9% 81.1%	44.4% 55.6%	63.7% 36.3%	57.5% 42.5%

Figure C.58: 0.25 second epoch, LN accelerometer, Right wrist

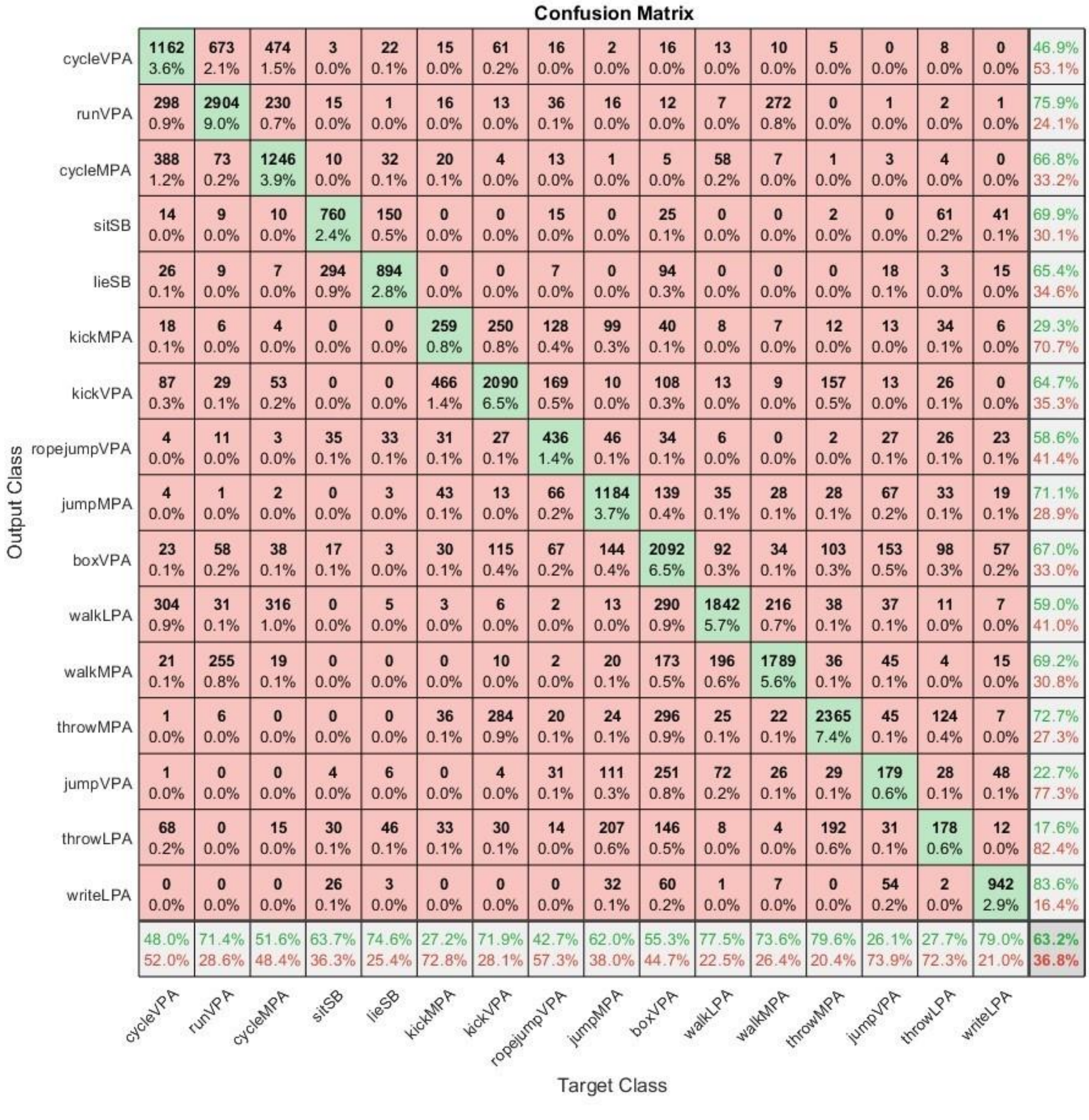


Figure C.58: 0.25 second epoch, WR accelerometer, Hip

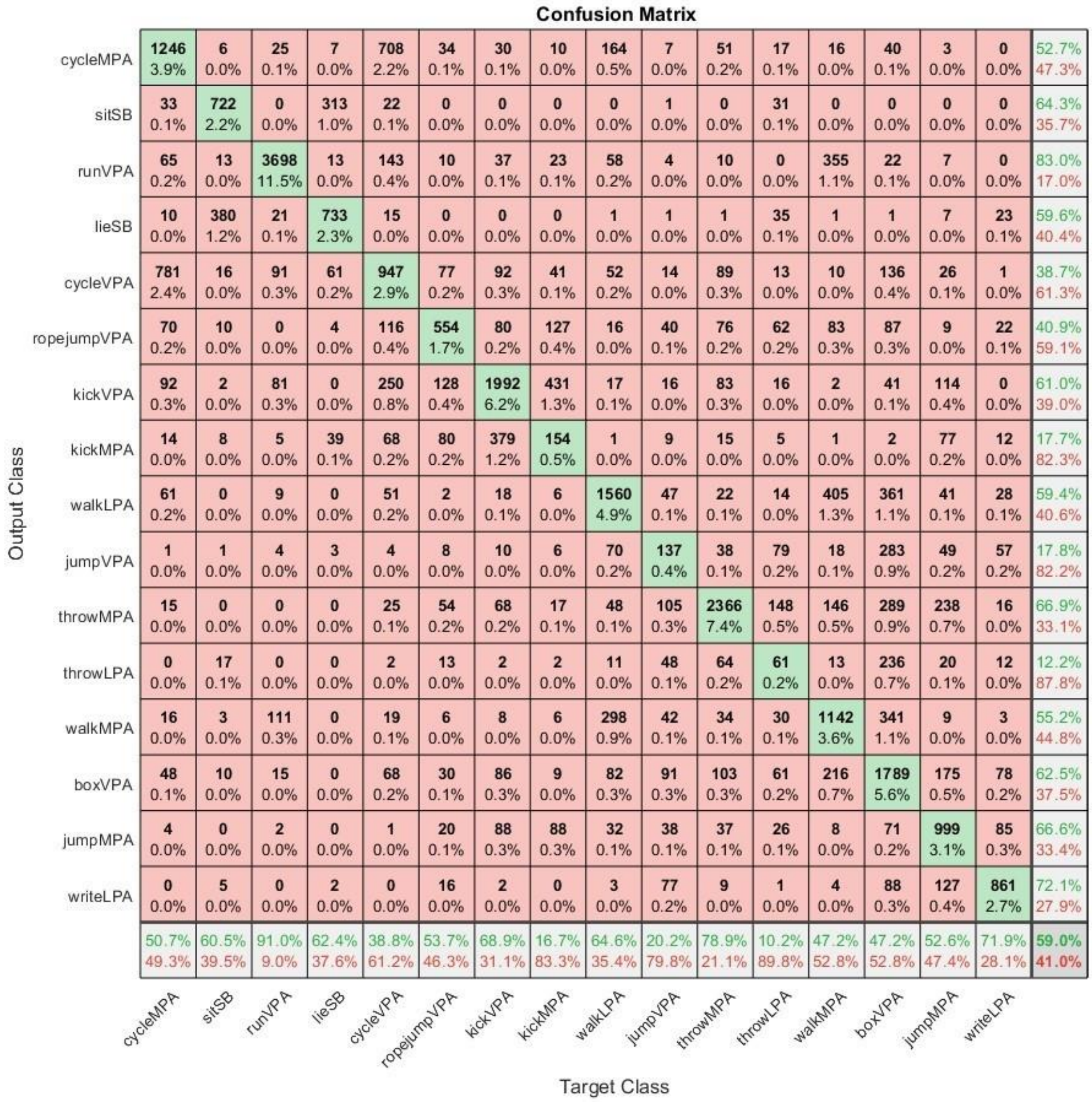


Figure C.58: 0.25 second epoch, WR accelerometer, Right wrist

C.4.3. Hip and right wrist: ”Unknown” class with KNN algorithm

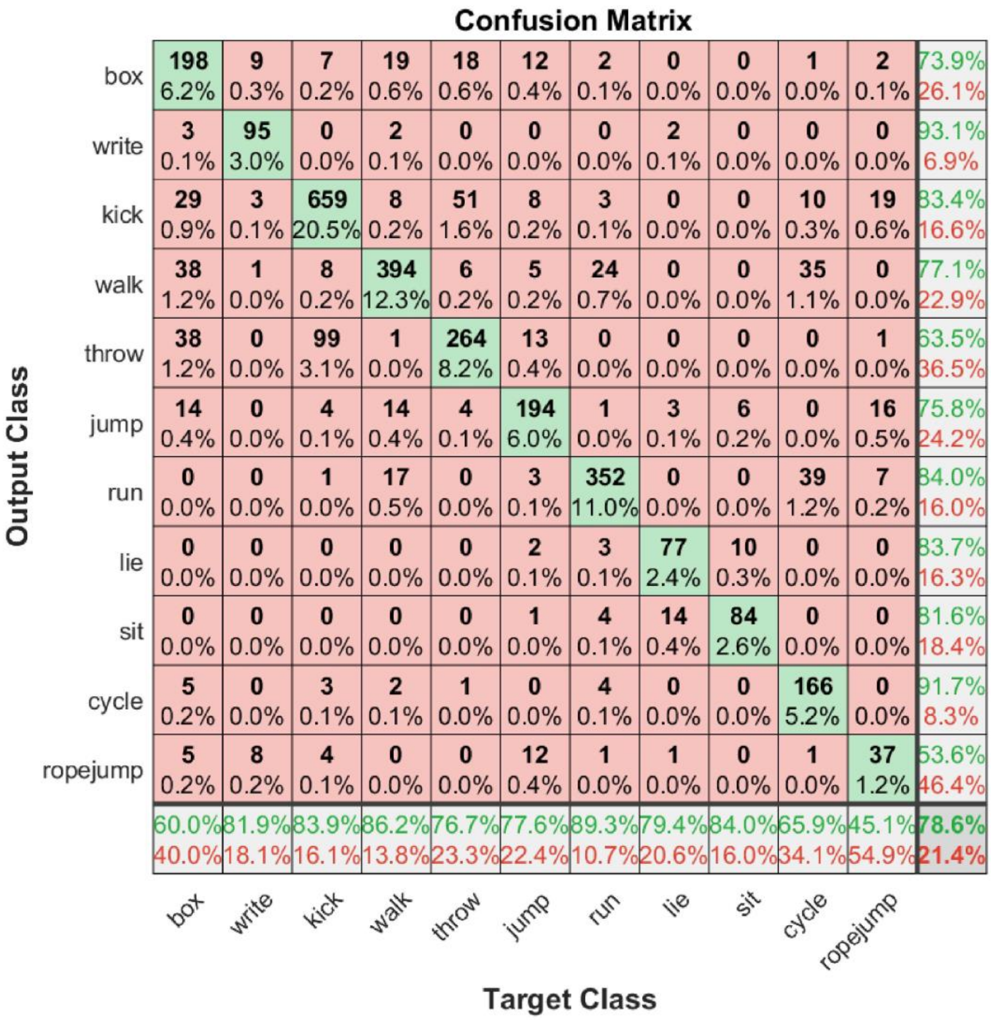


Figure C.61: 0.25 second epoch, LN accelerometer, Hip, Unknown with KNN

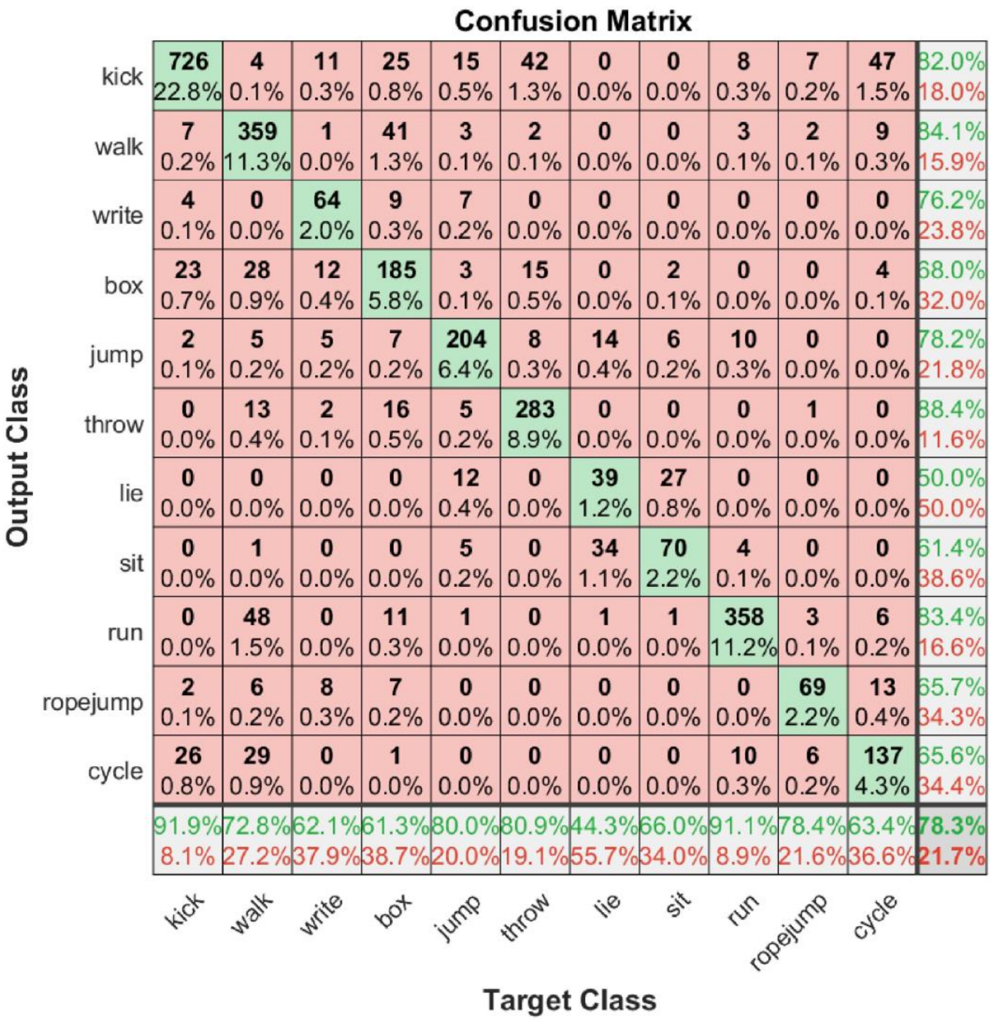


Figure C.62: 0.25 second epoch, LN accelerometer, Right wrist, Unknown with KNN

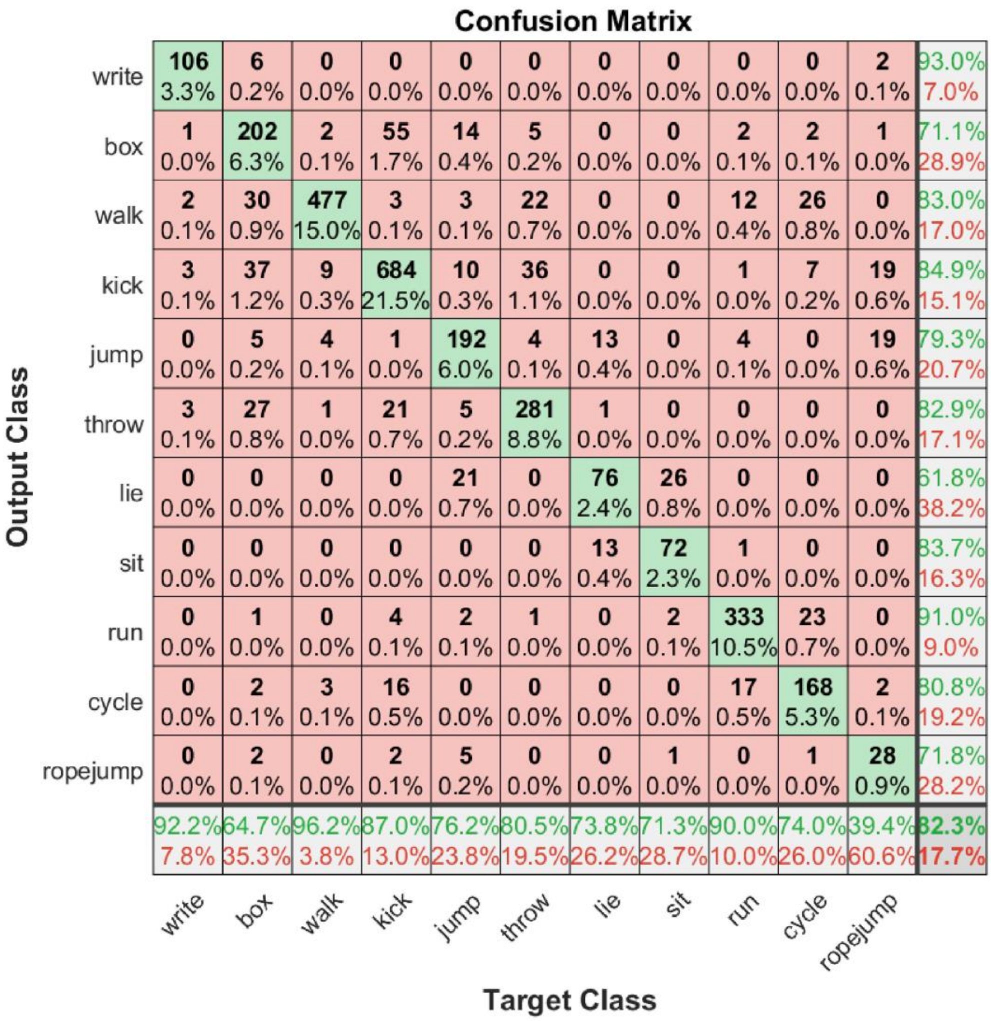


Figure C.63: 0.25 second epoch, WR accelerometer, Hip, Unknown with KNN

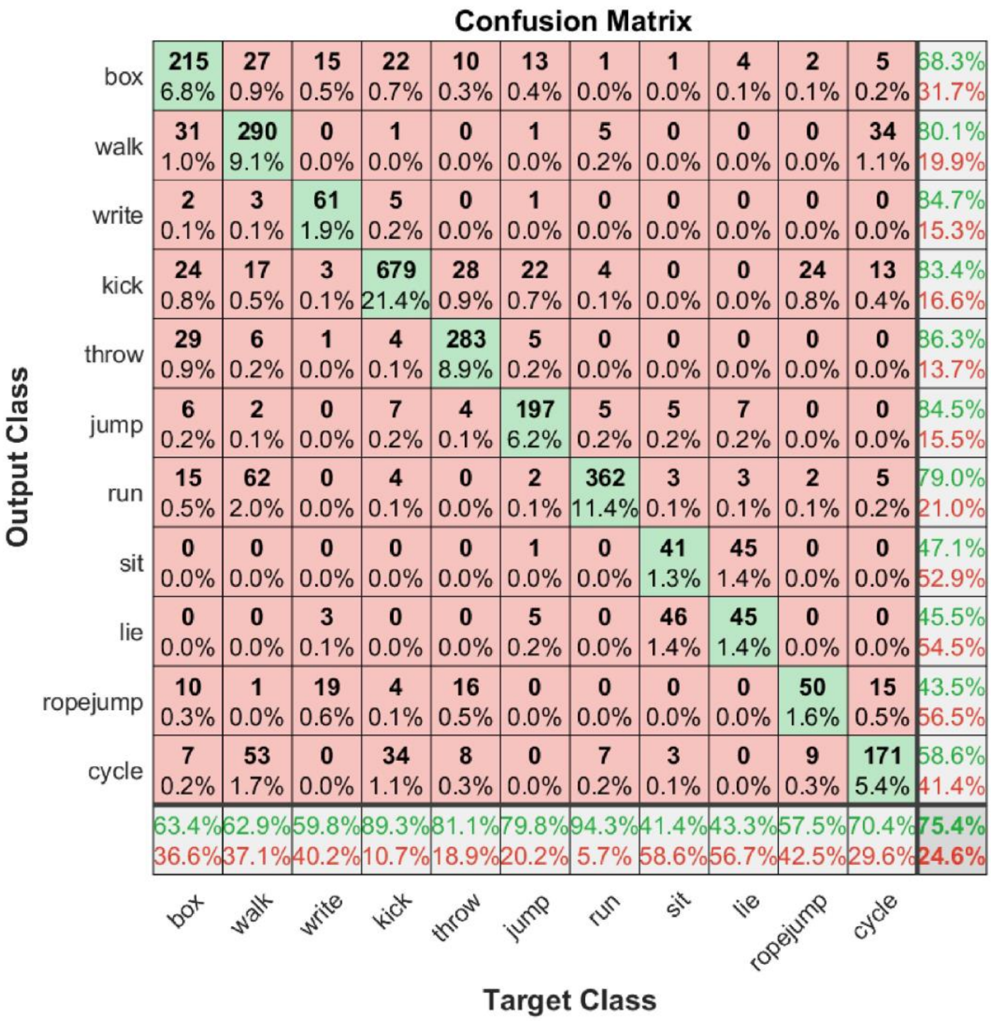


Figure C.64: 0.25 second epoch, WR accelerometer, Right wrist, Unknown with KNN

C.4.4. Hip and right wrist: "Unknown" class with threshold

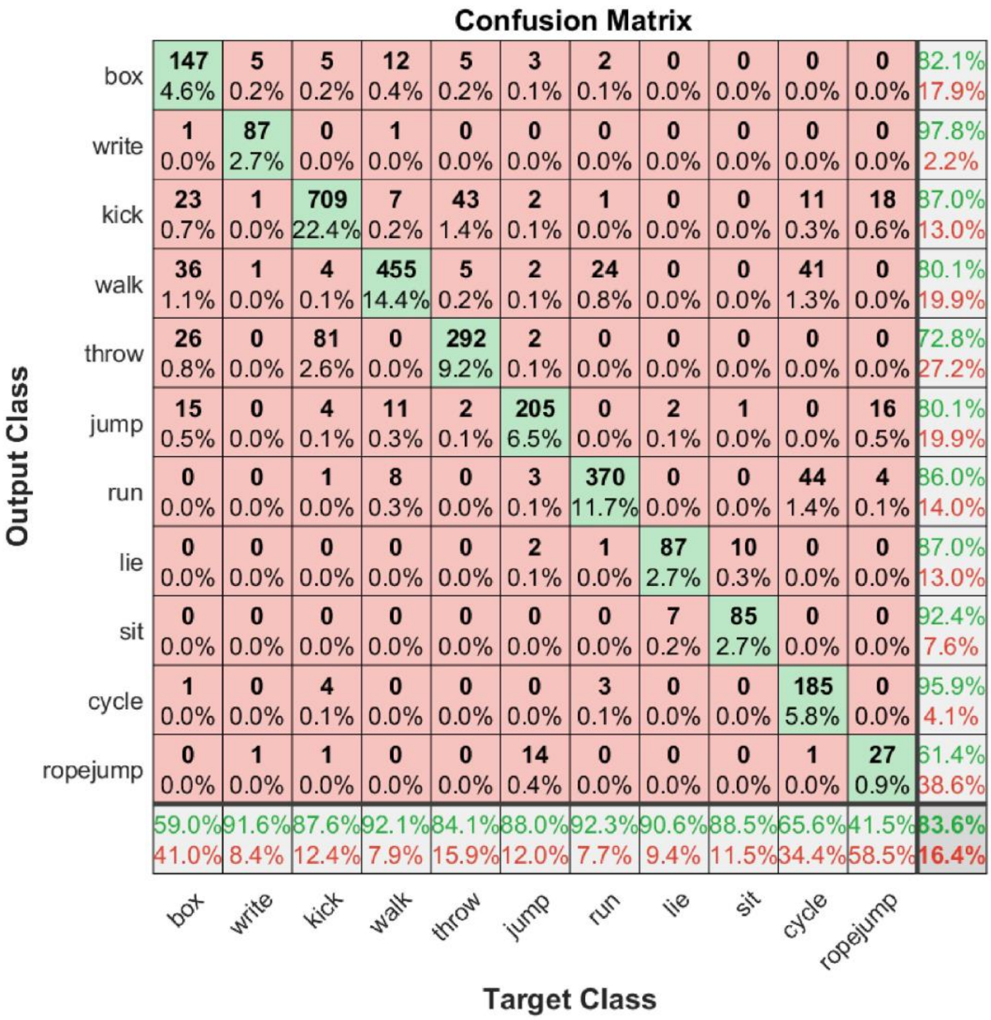


Figure C.65: 0.25 second epoch, LN accelerometer, Hip, Unknown with threshold at 0.775

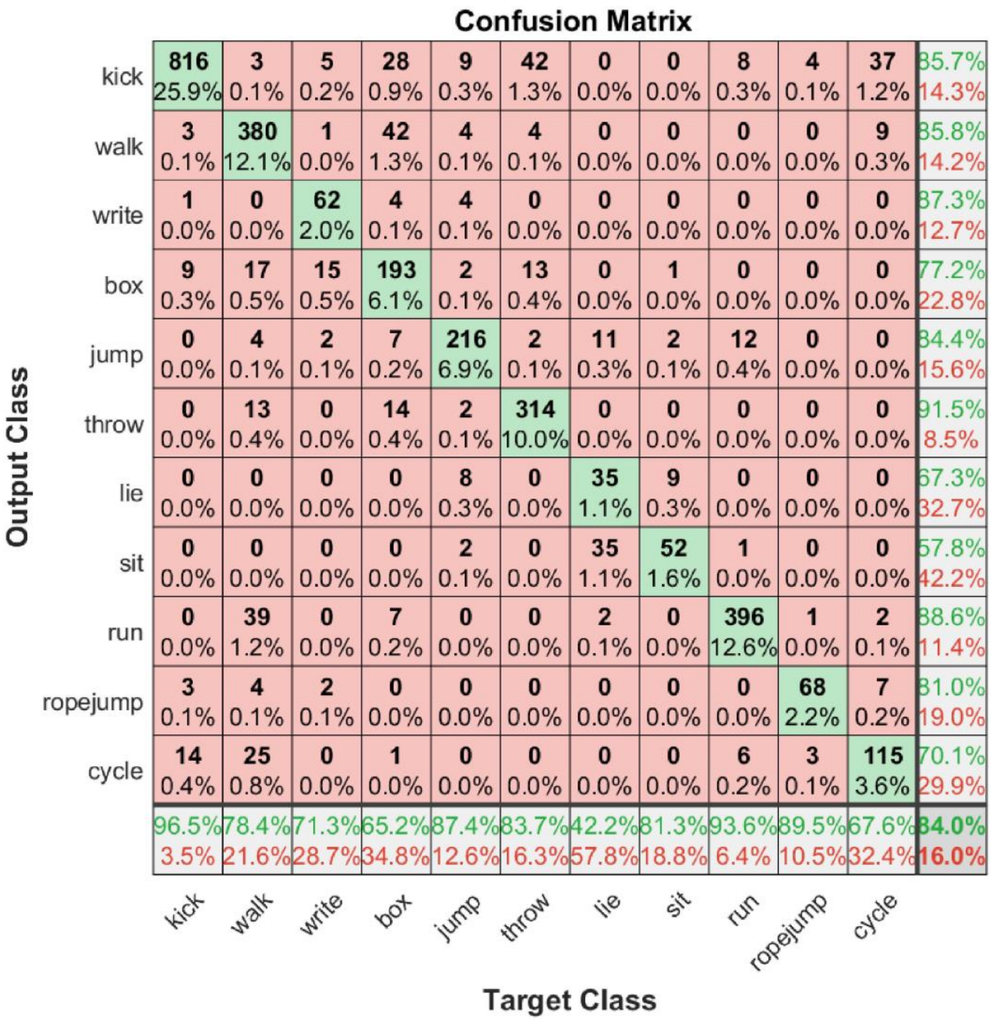


Figure C.66: 0.25 second epoch, LN accelerometer, Right wrist, Unknown with threshold at 0.775

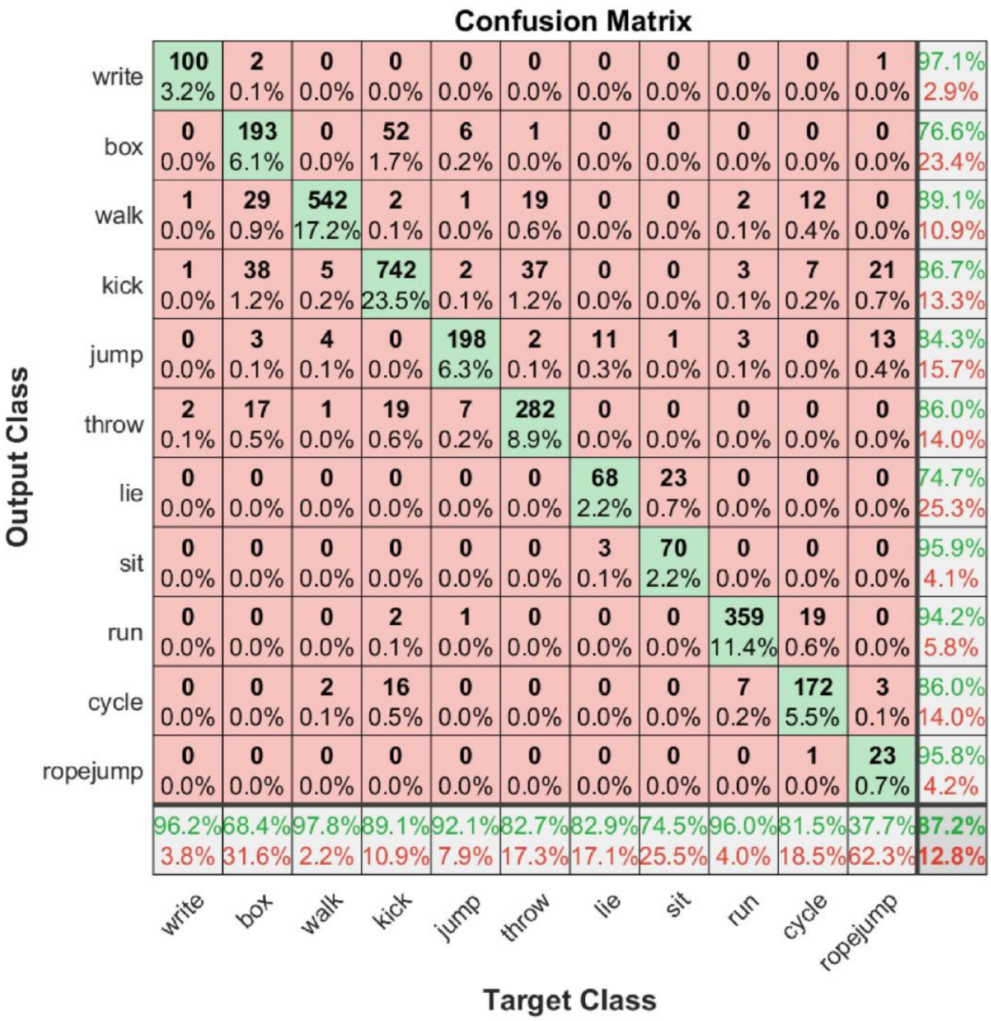


Figure C.67: 0.25 second epoch, WR accelerometer, Hip, Unknown with threshold at 0.775

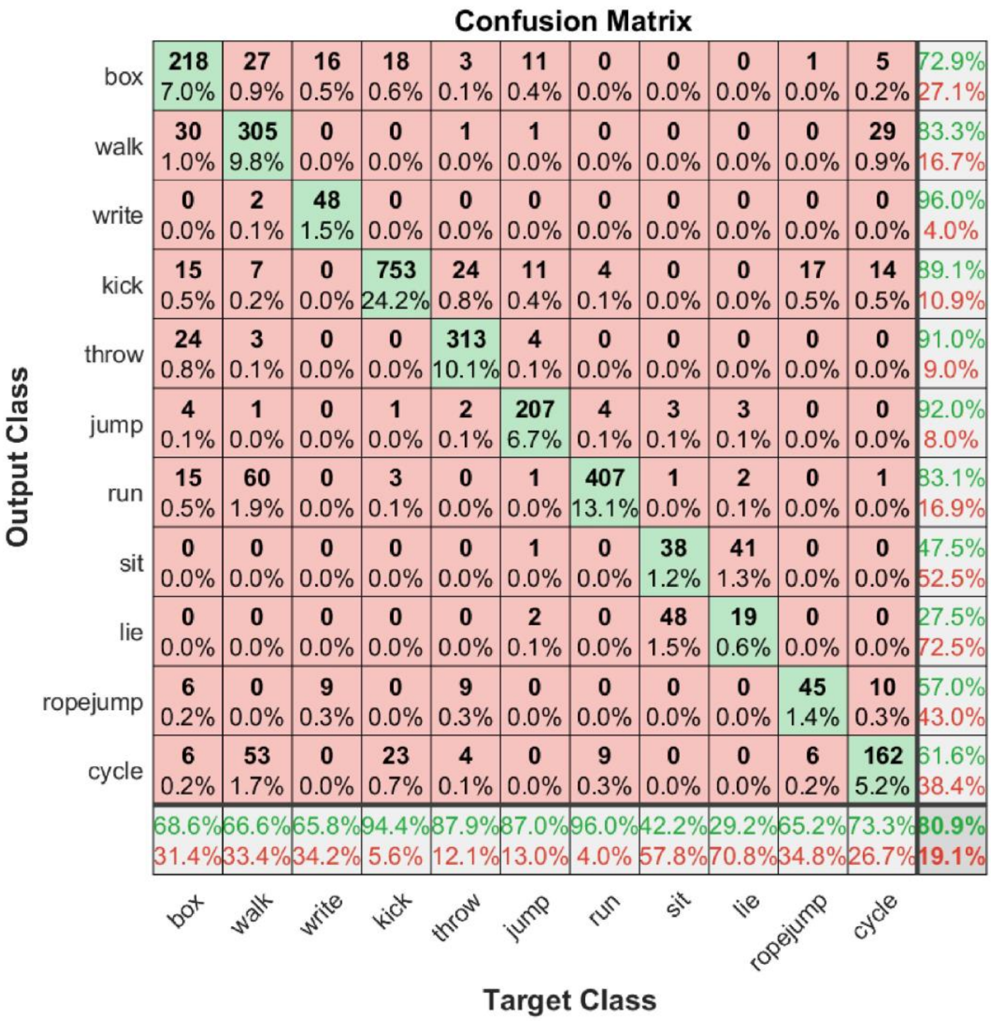


Figure C.68: 0.25 second epoch, WR accelerometer, Right wrist, Unknown with threshold at 0.775

D

Detailed results

Accuracies for all five wear-sites and four epoch lengths

Placement	Low noise accelerometer			
	1.0 s. epochs	0.5 s. epochs	0.33 s. epochs	0.25 s. epochs
Hip	31.2 ±22.7 (29.0 - 33.4)	52.9 ±22.9 (50.7 - 55.1)	65.3 ±18.9 (63.5 - 67.1)	76.1 ±19.8 (74.2 - 78.0)
Left ankle	21.7 ±16.4 (19.5 - 23.9)	59.2 ±21.1 (56.3 - 62.1)	60.1 ±22.5 (57.0 - 63.1)	65.1 ±21.4 (62.2 - 68.0)
Left wrist	24.9 ±20.3 (22.9 - 26.9)	49.1 ±22.5 (47.0 - 51.3)	67.2 ±20.5 (65.2 - 69.2)	71.7 ±17.7 (70.0 - 73.4)
Right ankle	21.0 ±16.2 (18.8 - 23.2)	52.2 ±28.3 48.3 - 56.1)	71.1 ±20.2 (68.4 - 73.9)	66.2 ±22.2 (63.2 - 69.2)
Right wrist	28.7 ±21.8 (26.6 - 30.8)	52.6 ±23.4 (50.4 - 54.9)	69.5 ±21.6 (66.3 - 70.5)	74.7 ±14.3 (73.3 - 76.0)
	Wide range accelerometer			
	1.0 s. epochs	0.5 s. epochs	0.33 s. epochs	0.25 s. epochs
Hip	32.4 ±23.5 (30.2 - 34.7)	56.2 ±22.8 (54.0 - 58.4)	69.5 ±17.1 (67.8 - 71.1)	78.8 ±13.3 (77.5 - 80.1)
Left ankle	25.3 ±22.2 (22.3 - 28.3)	63.9 ±22.3 (60.8 - 66.9)	68.4 ±19.3 (65.7 - 71.0)	66.2 ±21.8 (63.2 - 69.1)
Left wrist	24.2 ±19.9 (22.3 - 26.1)	43.1 ±22.0 (40.9 - 45.2)	62.9 ±24.7 (60.5 - 65.2)	69.4 ±16.6 (67.8 - 71.0)
Right ankle	24.0 ±18.9 (21.4 - 26.6)	62.0 ±25.8 (58.5 - 65.5)	70.8 ±21.3 (67.9 - 73.7)	70.0 ±17.4 (67.5 - 72.2)
Right wrist	21.9 ±18.1 (20.2 - 23.7)	55.5 ±20.9 (53.5 - 57.6)	69.0 ±21.2 (67.0 - 71.1)	74.0 ±16.6 (72.4 - 75.6)

Table D.1: The mean accuracy (%) with standard deviation and 95% confidence interval per wear-site, sensor type and epoch length.

Accuracies for gyroscope and magnitude data

Operation	Low noise accelerometer				Wide range accelerometer			
	0.33 s. epochs		0.25 s. epochs		0.33 s. epochs		0.25 s. epochs	
	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist	Hip	Right wrist
Gyroscope	61.5 ±20.9 (59.5 - 63.5) p = 0.049	59.9 ±20.6 (57.9 - 61.8) p<0.001	69.4 ±18.3 (67.6 - 71.2) p<0.001	59.1 ±21.4 (57.0 - 61.2) p<0.001	67.3 ±17.1 (65.4 - 69.2) p = 0.381	67.5 ±20.5 (65.5 - 69.4) p = 0.680	74.3 ±17.8 (72.6 - 76.0) p = 0.002	68.5 ±20.4 (66.5 - 70.4) p<0.001
Magnitude	59.1 ±20.8 (57.1 - 61.1) p<0.001	57.5 ±24.0 (55.2 - 59.8) p<0.001	64.4 ±19.5 (62.5 - 66.3) p<0.001	70.7 ±19.8 (68.8 - 72.7) p = 0.021	50.3 ±21.9 (48.2 - 52.4) p<0.001	50.1 ±20.1 (48.2 - 52.1) p<0.001	64.9 ±21.0 (62.9 - 66.9) p<0.001	50.3 ±18.6 (48.5 - 52.1) p<0.001
Gyroscope enriched magnitudes	53.8 ±23.2 (51.5 - 56.0) p<0.001	52.5 ±22.7 (50.3 - 54.7) p<0.001	68.7 ±19.3 (66.8 - 70.5) p<0.001	55.9 ±22.0 (53.7 - 58.0) p<0.001	62.5 ±20.5 (60.5 - 64.5) p<0.001	58.4 ±18.9 (56.5 - 60.2) p<0.001	71.3 ±18.8 (69.5 - 73.1) p<0.001	60.4 ±20.0 (58.5 - 62.3) p<0.001

Table D.2: The mean accuracy (%) with standard deviation and 95% confidence interval and significance with respect to accelerometer-only data of the gyroscope enriched, accelerometer magnitude and gyroscope and accelerometer magnitude data for measurements from the IMUs on the right wrist and hip summarized per 0.33 and 0.25 seconds.

Accuracies for the shorter sequences

Grouped activities	Low noise accelerometer 0.33 s. epochs Hip	Low noise accelerometer 0.33 s. epochs Right wrist	Low noise accelerometer 0.25 s. epochs Hip	Low noise accelerometer 0.25 s. epochs Right wrist
Cycle - Kick - Rope jump - Run	67.3 ±19.9 (63.4 - 71.2)	77.7 ±12.1 (75.3 - 80.0)	72.8 ±28.0 (67.4 - 78.3)	78.8. ±13.6 (76.2 - 81.5)
Box - Jump - Throw - Walk	58.9 ±24.8 (54.1 - 63.7)	60.3 ±30.1 (54.4 - 66.1)	69.9 ±18.4 (66.3 - 73.5)	71.2 ±17.6 (67.8 - 74.6)
Jump - Lie - Run - Sit	67.2 ±14.0 (64.5 - 69.9)	68.1 ±15.9 (65.0 - 71.2)	85.4 ±11.2 (83.2 - 87.5)	75.2 ±12.3 (72.9 - 77.6)
Box - Kick - Walk - Write	67.8 ±13.2 (65.2 - 70.3)	67.5 ±20.7 (63.4 - 71.5)	76.5 ±13.7 (73.8 - 79.1)	73.4 ±12.2 (71.0 - 75.8)
	Wide range accelerometer 0.33 s. epochs Hip	Wide range accelerometer 0.33 s. epochs Right wrist	Wide range accelerometer 0.25 s. epochs Hip	Wide range accelerometer 0.25 s. epochs Right wrist
Cycle - Kick - Rope jump - Run	71.4 ±22.2 (67.1 - 75.7)	71.0 ±20.3 (67.1 - 75.0)	76.4 ±14.9 (73.5 - 79.3)	81.1 ±13.2 (78.5 - 83.7)
Box - Jump - Throw - Walk	69.6 ±13.5 (67.0 - 72.2)	67.9 ±28.8 (62.3 - 73.5)	76.1 ±14.3 (73.3 - 78.9)	66.7 ±21.5 (62.5 - 70.9)
Jump - Lie - Run - Sit	66.7 ±17.9 (63.3 - 70.2)	71.5 ±13.6 (68.8 - 74.1)	82.0 ±10.6 (80.0 - 84.1)	75.2 ±13.1 (72.7 - 77.8)
Box - Kick - Walk - Write	70.1 ±13.0 (67.6 - 72.7)	65.7 ±18.9 (62.0 - 69.3)	80.8 ±12.2 (78.4 - 83.2)	72.9 ±13.9 (70.2 - 75.6)

Table D.3: The mean accuracy (%) with standard deviation and 95% confidence interval of the predictions per short sequence of activities for measurements from both accelerometers on the right wrist and hip summarized per 0.33 and 0.25 seconds.

Precision and recall for all activities

Activity		Low noise accelerometer 0.33 s. epochs Hip	Low noise accelerometer 0.33 s. epochs Right wrist	Low noise accelerometer 0.25 s. epochs Hip	Low noise accelerometer 0.25 s. epochs Right wrist	Wide range accelerometer 0.33 s. epochs Hip	Wide range accelerometer 0.33 s. epochs Right wrist	Wide range accelerometer 0.25 s. epochs Hip	Wide range accelerometer 0.25 s. epochs Right wrist
Box	Prc.	45.9%	69.5%	64.9%	66.0%	65.6%	72.0%	70.6%	65.2%
	Rcl.	36.7%	51.2%	49.7%	54.8%	49.3%	51.2%	54.4%	53.3%
Cycle	Prc.	66.9%	47.1%	79.8%	59.6%	79.1%	52.7%	86.7%	61.2%
	Rcl.	60.7%	57.0%	64.5%	58.5%	70.4%	58.5%	63.6%	73.4%
Jump	Prc.	65.7%	75.1%	72.3%	77.1%	83.1%	74.7%	78.8%	80.2%
	Rcl.	69.4%	67.0%	75.6%	74.3%	66.2%	73.1%	78.3%	71.9%
Kick	Prc.	71.6%	70.5%	80.7%	79.4%	75.0%	70.7%	78.6%	79.5%
	Rcl.	85.1%	86.1%	83.3%	89.6%	79.5%	80.1%	87.1%	87.7%
Lie	Prc.	52.8%	40.8%	77.1%	54.5%	53.1%	48.2%	65.3%	53.8%
	Rcl.	46.4%	27.3%	76.7%	43.1%	51.1%	47.8%	69.9%	57.7%
Rope jump	Prc.	52.0%	53.4%	60.0%	56.1%	61.0%	39.9%	73.1%	49.1%
	Rcl.	37.5%	66.1%	48.7%	63.7%	52.1%	42.7%	36.1%	62.0%
Run	Prc.	74.8%	72.1%	83.6%	82.1%	64.8%	79.2%	88.5%	83.3%
	Rcl.	67.9%	90.1%	90.1%	93.3%	75.1%	87.9%	91.8%	93.4%
Sit	Prc.	59.9%	47.2%	74.9%	59.3%	52.5%	41.1%	74.2%	55.8%
	Rcl.	52.9%	55.9%	74.4%	61.6%	51.1%	35.7%	66.4%	48.4%
Throw	Prc.	64.2%	78.9%	61.1%	79.2%	62.0%	69.2%	76.8%	79.6%
	Rcl.	60.7%	62.8%	70.4%	77.5%	77.0%	79.2%	75.3%	77.6%
Walk	Prc.	66.1%	81.1%	79.5%	80.6%	77.7%	74.9%	78.0%	81.1%
	Rcl.	69.9%	63.0%	85.7%	73.8%	77.1%	65.1%	74.8%	67.6%
Write	Prc.	59.1%	77.2%	78.1%	76.0%	75.7%	86.4%	80.2%	73.5%
	Rcl.	71.0%	57.5%	73.1%	59.0%	75.8%	54.3%	86.1%	64.4%
Average	Prc.	61.7%	64.8%	73.8%	70.0%	68.2%	64.5%	77.4%	69.3%
	Rcl.	59.8%	62.2%	72.0%	68.1%	65.9%	61.4%	71.3%	68.9%
Sit or Lie	Prc.	91.5%	84.3%	90.9%	92.8%	88.4%	91.4%	91.8%	89.1%
	Rcl.	80.6%	90.6%	90.4%	84.8%	85.5%	85.0%	90.2%	86.5%
	Acc.	+2.2%	+3.0%	+1.1%	+2.3%	+2.4%	+2.9%	+1.6%	+2.3%
Jump or Rope jump	Prc.	69.2%	72.2%	77.1%	70.7%	80.0%	67.2%	86.4%	69.2%
	Rcl.	69.4%	68.2%	75.6%	71.5%	65.0%	67.5%	74.0%	69.2%
	Acc.	+1.0%	+0.2%	+0.9%	+0.0%	+0.3%	+0.3%	+0.8%	+0.0%
Box or Throw	Prc.	68.7%	81.1%	73.5%	80.7%	73.7%	75.9%	80.2%	79.6%
	Rcl.	59.1%	62.1%	70.3%	72.9%	73.2%	70.2%	70.1%	71.2%
	Acc.	+2.3%	+1.1%	+2.2%	+1.4%	+2.2%	+1.1%	+1.2%	+1.4%
Correction accuracy		+5.5%	+4.3%	+4.2%	+3.7%	+4.9%	+4.3%	+3.6%	+3.7%

Table D.4: Precision and recall of classification method for individual activities and corrected for activities that are less relevant to distinguish, see Appendix C.