Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

Roar Munkeby Fenne

Validation of the NTNU-HAR model for assessment of physical activity in typically developing children and adolescents, and in children and adolescents with cerebral palsy.

Master's thesis in Human Movement Science
Supervisor: Ellen Marie Bardal
June 2020

NTNU
Norwegian University of
Science and Technology

Roar Munkeby Fenne

Validation of the NTNU-HAR model for assessment of physical activity in typically developing children and adolescents, and in children and adolescents with cerebral palsy.

Master's thesis in Human Movement Science
Supervisor: Ellen Marie Bardal
June 2020

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

NTNU
Norwegian University of
Science and Technology

Abstract

Background: The NTNU-HAR is a machine learning based human activity recognition model that has been developed to recognize and predict different types of physical activity (PA) in healthy adults. Validation in other population groups is needed, including children and adolescents and people with physical disabilities.

Study Aim: To assess the validity of the NTNU-HAR in detecting and classifying types of PA in both typically developing (TD) children and adolescents and also children and adolescents with cerebral palsy (CP). Another aim was to assess if the performance of the model changes with age and length of activity bouts.

Methods: 67 TD children and adolescents and 16 children with CP were equipped with two triaxial accelerometers and a chest mounted camera. Two protocols were conducted: One semi-structured protocol with different activities lasting for short periods of time (<30 seconds), and one protocol consisting of activities of longer duration (>3minutes). Annotation of video recordings were used as gold standard to assess the validation of the NTNU-HAR in classifying types of PA. Four groups were tested, based on age and protocol: Adolescence (long bout), children (short bout), children (long bout) and CP (Short + long).

Overall accuracy were calculated for each group. Sensitivity, specificity, and positive predictive values (PPV) were calculated for each PA type.

Results: Overall accuracy was 94.9 % for adolescents (long bout), 90.9% for children (long bout), 67.5% for children (short bout), and 73.3% for CP. In the long bout protocol, sensitivity was high (>90%) for walking, running, standing, sitting, and cycling (sit). Most misclassifications were due to shuffling being misclassified as standing and walking. Sensitivity and PPV decreased for all categories in the short bout protocol. This was also the case for the CP group.

Conclusion: The NTNU-HAR is a valid tool for classifying PA types in TD children and adolescents if activities are performed over longer periods of time. The length of the activity periods affects the performance of the HAR-model in predicting PA-types, with accuracy decreasing if the length of the accuracy bouts are short. This might indicate a poorer validity of the model during free play. There was no effect of age on the performance of the model. For children and adolescents with CP, promising results were found, but differences in protocol makes it hard to draw conclusions whether a new ML-model needs to be developed exclusively for this group, or only trained on short/complex physical activities.

Sammendrag

Bakgrunn: NTNU-HAR er en maskinlæringsbasert aktivitetsgjenkjenningsmodell som har blitt utviklet til å gjenkjenne og predikere forskjellige typer av fysisk aktivitet (FA) hos friske voksne. Validering i andre grupper av befolkningen er nødvendig. Dette inkluderer barn og personer med fysiske funksjonshemninger.

Mål: Å evaluere validiteten til NTNU-HAR til å detektere og klassifisere typer av FA hos både normalt utviklede barn (TD) og hos barn med cerebral parese (CP). Et annet mål var å undersøke om validiteten til modellen endres med alder og lengde av aktivitetsperiodene.

Metode: 67 TD barn og 16 barn med CP ble utstyrt med to triaksiale akselerometre og et kamera montert på brystet. To protokoller ble utført: Én semistrukturert protokoll hvor forskjellige aktiviteter varte i korte tidsperioder (<30 sekunder), og én protokoll bestående av aktiviteter med lengre varighet (>3 minutter). Annotering av videoopptak ble brukt som gullstandard for å evaluere validiteten til NTNU-HAR til å klassifisere typer av FA. Fire grupper ble testet, basert på alder og protokoll: Ungdom (lang økt), barn (kort økt), barn (lang økt) og CP (kort + lang). Overordnet nøyaktighet ble regnet ut for hver gruppe. Sensitivitet, spesifisitet og positiv prediktiv verdi (PPV) ble regnet ut for hver FA-type.

Resultat: Overordnet nøyaktighet var 94.9% for ungdom (lang økt), 90.9% for barn (lang økt), 67.5% for barn (kort økt) og 73.3% for CP. Sensitiviteten var høy (≥90%) for kategoriene gå, løpe, stå, sitte og sykle(sittende). Flest feilklassifiseringer var på grunn av at shuffling-kategorien ble feilklassifisert som stå og gå. Sensitivitet og PPV sank for alle kategorier i den korte protokollen. Det samme var tilfelle i CP gruppen.

Konklusjon: NTNU-HAR er et gyldig verktøy for å klassifisere FA-typer hos TD barn og ungdom hvis aktivitetene er utøvd over lengre tidsperioder. Lengden på aktivitetsperiodene påvirker prestasjonen til NTNU-HAR i å predikere FA-typer, da nøyaktigheten synker hvis lengden på aktivitetsperiodene er av kort varighet. Dette indikerer en dårligere validitet for modellen under fri lek. Det var ingen effekt av alder på prestasjonen til modellen. For barn og ungdom med CP var resultatene lovende, men forskjell i protokoll både sammenlignet med TD gruppen og mellom deltakerne i CP gruppen gjør det vanskelig å trekke konklusjoner på hvorvidt en ny maskinlæringsmodell må utvikles kun for denne gruppen, eller om den nåværende kun trenger å bli trent på korte/komplekse fysiske aktiviteter.

Acknowledgements

A lot of people have contributed to this project.

I would first like to thank my supervisor Ellen Marie Bardal for her good guidance and supervision throughout the year. To Hilde Bremseth Bårdstu and Atle Kongsvold, thanks for all the help with the video annotation. I would like to thank my fellow student Sondre Hovda Dahlskaas for the good discussions during the year,

I would also like to thank everyone at the Department of Computer and Information Science (IDI) responsible for developing the NTNU-HAR model. A special thanks goes to Håkon Slåtten Kjærnli, for handling the testing process, as well as answering all my questions regarding machine learning.

Of course, thank you to everyone involved in the planning and execution of this exciting project before I joined. A special thanks goes to the participants for their contribution.

Finally, a thank you to my fellow students Stine Øverengen Trollebø, Vera Heggenhougen, Gunnveig Berge, and the rest of the AFO-team, for giving me valuable project planning- and lab experience by letting me be a part of their data collection team.

Table of contents

1. Introduction

It is well known that physical inactivity is associated with an increased risk of several non-communicable diseases such as obesity, diabetes, cardiovascular disease, and cancer in adults (1, 2). Regular bouts of physical activity (PA) on the other hand is associated with a reduced risk of the same diseases (3)**.** The relationship between PA and non-communicable diseases may not be as strong in children and adolescents as it is in adults (4). However, many of the same adaptions and benefits of PA in adults also applies to children and adolescents. These include improved cardiorespiratory- and muscular fitness, bone health, and cardiovascular- and metabolic health markers (5). The weak relation between PA and non-communicable diseases in children might be explained by a relative short available time frame for both exposure of physical inactivity and for development of disease. However, the benefits of physical activity in childhood and adolescence carry on into adulthood and might further promote a healthy and physically active lifestyle throughout life (6, 7). Indeed, those who maintains their PA levels from adolescence into adulthood have a lower risk of cardiovascular disease, and a better mental health, compared with those who doesn't maintain their PA level (8). There is also evidence showing that children whose parents are physically inactive most likely will end up being physically inactive as well throughout life (9). Thus, it is highly important to ensure a physically active lifestyle for children and adolescents.

The World Health Organization (WHO) recommends that children and adolescents should accumulate at least 60 minutes of moderate- to vigorous-intensity physical activity daily (5). However, 81 % of all adolescents worldwide fails to meet current physical activity guidelines (10). In addition to low physical activity level among children and adolescents, the amount of time spent in sedentary behavior poses additional negative effects on their health prospects. This includes a higher risk of cardiovascular diseases, metabolic syndrome, and depression (11, 12). Sedentary behavior can be defined as "any waking behavior that are done in sitting or reclining posture that expends ≤1.5 metabolic equivalents (METs)" (13). Keane et al. found that among 826 children aged 8-11 years, they spent on average 61% of their waking time sedentary (14).

Therefore, there is a need to implement policies and measures to increase physical activity in children and adolescents. This requires accurate measurement methods to correctly quantify the activity levels in the population, and to further assess the dose-response relationship between PA and health benefits. Valid and precise measurement method will then lead to even more accurate health recommendations as the research behind them will be more accurate.

The main body of physical activity research today are using body worn sensors to assess physical activity levels. Body worn sensors gives an objective measure of a subject's movement during a given time period and may provide an accurate quantification of physical activity behaviors. The technology used in body worn sensors has improved significantly the last ten years. The latest sensors are small and lightweight, with large battery capacity and internal memory (15). This have made body worn sensors, and especially accelerometers, a preferred tool to assess physical activity, not only in experimental studies, but also in many population studies.

The first generation of accelerometers quantified movement as activity "counts", by counting the number of times the acceleration signal reached above a threshold value, over a given time period (16). Energy expenditure was then estimated using regression-based cut-off values. However, different manufacturers have used different algorithms to generate counts, and these have traditionally been kept as a manufacturers secret (15). This makes it hard to compare results between studies using different types of accelerometers.

In the recent years the use of accelerometers that gives the raw, unfiltered, acceleration signals have been more widely used (15, 16). This gives the opportunity to extract more information beyond exercise intensity. The gravity component of the acceleration signal when a person is inactive makes it possible to detect posture (17), while the pattern of the acceleration signal during dynamic activities makes it possible to recognize different types of physical activity (15). However, use of raw acceleration sensors require access to an analytic tool to extract the outputs of interest.

One approach to develop such analytic tools is machine learning (ML). Machine learning is a field of artificial intelligence that automates analytical model building, based on the idea that systems can learn from data, identify patterns and make decisions on their own (18). A ML-based model can learn to recognize different types of PA based on how the pattern of the acceleration signal looks like when these activities are performed and can provide tailored output variables (19).

At the Norwegian University of Science and Technology (NTNU) Department of Computer Science (IDI), in a collaboration with the Department of Public Health and Nursing (SM) and Department of Neuromedicine and Movement Science (INB), have developed a ML-based Human Activity Recognition model (NTNU-HAR). This model detects postures (lying down, sitting, standing) and activities (walking, running, stairclimbing, cycling, picking, bending), based on data from two accelerometers placed on the lower back and thigh. The model will be used to analyze physical activity data collected from the fourth wave of

The North-Trøndelag Health Survey (HUNT 4) (20). In this study, over 38 000 people wore two accelerometers for seven days as a part of the study.

The NTNU-HAR-model has been trained, developed, and validated for healthy adults and has reached an accuracy of 94% in predicting types of activity (21). Similar studies often operates with accuracies above 80% as acceptable, and above 90% as high (e.g by Trost, Zheng and Wong (22)). However, a goal is to end up with one analytical tool that can be used for the whole population. A question that is yet to be answered is whether the present model is valid for analyzing accelerometer data from children and adolescents, or if we need to train a sub-model to analyze activity data for the younger population?

With regard to measuring PA, children and adolescents are historically treated as small adults (23). However, they show a deviation of both movement patterns and activity patterns. The methods developed for adults may therefore be less suitable for this young population. During childhood and adolescence, the body undergoes a lot of changes as it continues to grow. Bone structure is altered by growth in both length and width, increased mass, and bone mineral density. This has implications for biomechanical movement, as it results in increased limb length and stature. The growth spurt during adolescence leads to further changes in body proportions, as growth in leg length precedes the growth in trunk length. This is followed by a growth in the muscular system (increased length, cross sectional area and mass), further resulting in changes in body composition (24). These are factors that might change the movement patterns of the growing child and might therefore have an impact on the performance of a HAR-model that is trained on adults. Especially when considering the differences in proportion that changes before and during growth in children and adolescents. As these changes occur at different times for every individual, it might be possible that the present HAR-model performs better when looking at certain age groups or in different growth phases.

Children also show a more spontaneous and transitory nature of physical activity characterized by frequent and short bouts of activity, with rapid changes in tempo and intensity. Especially during play these characteristics are evident. Free play is often distinguished by frequent bouts of short, low intensity PA, interspersed with less frequent PA of high intensity. During high intensities, the length of the activity bouts might range from anywhere between 3-20 seconds (25, 26).

Therefore, it is possible that the HAR-model will encounter problems in detecting these rapid changes, as the pattern from the acceleration signal is likely to vary a lot over a short time period. Thus, we might assume that the length of the activity bouts will affect the performance of the model, as short bursts of activity might be harder to detect than longer bursts of activity.

Body worn sensors serve as a tool to evaluate physical activity levels not only in typically developing (TD) children, but also in children with physical impairments.
Cerebral Palsy (CP) is the most common physical disability in children, with a prevalence of 1,89 per 1000 live births in Norway and 2,11 per 1000 in the developed world (27, 28). CP is a group of disorders caused by damage of the fetal or infant brain which affect the development of movement and posture (29). Depending on which areas of the brain that are affected, common movements disorders that can occur are stiff muscles (spasticity), uncontrollable movements (dyskinesia), and poor balance and coordination (ataxia) (30).

There are many barriers for children with CP to participate in physical activity. Whilst many of the factor are social/contextual, the impairments in body structures and motor function is of special relevance when assessing time spent in activity and what type of activities are being performed. Accurate measurements are important to evaluate the effect of treatments to increase PA, like surgery, injection of botulinum toxin (Botox) or physical therapy.

The factors that are thought to affect the performance of a HAR model trained on adults when tested on TD children might also apply for children with CP. However, the different motor impairments and the varied severity of them might produce different acceleration patterns when performing the same activity. This might lead to further difficulties for the HAR model that will affect the accuracy in detecting types of PA for children with CP.

The aim of the present study is therefore to assess the validity of the NTNU HAR in detecting and classifying types of physical activity in both typically developing children and adolescents and in children and adolescents with cerebral palsy.
We will also assess if the validity of the model changes with age, and length of activity bouts.

2. Method

The present study is a part of a larger validation study on physical activity and energy expenditure in children and adolescents. The study protocol was approved by the Norwegian Centre for Research Data (NSD). The data used in the present study are only a selection of the variables that were collected in the larger validation study. The data collection started during the fall of 2017 and lasted until 2019.

2.1 Participants

86 participants were included in the main validation study: 67 TD children and 16 children with CP. Prior to the study, the participants and their parents were informed of the aims of the study, and a written consent was signed by the parents prior to participation.

2.1.1 Typically developing children

The TD participants were recruited from a primary- and secondary school outside Trondheim. In the TD group, 67 children and youth, age ranging from 7-16 years, participated in the study. To ensure an equal distribution of age and gender, three boys and three girls from each school class was recruited. Of these, 48 subjects were included in the analyses of the present study (27 boys, 21 girls). 47 out of 48 participated in the long bout protocol, 13 of the participants completed both protocols, and one participant completed the short bout protocol only. The data collection was performed at the school area. The reason only 48 subjects were included was mainly due to time constraints, as analysis of video recordings are time consuming. Also, some participants had missing data that occurred during or after data collection.

2.1.2 Children with CP

The CP participants were recruited from habilitation clinics in both mid-Norway and south-east Norway and the data collection was performed at the different out-patient clinics.
In the CP group, 16 children and youth (8 boys, 8 girls), age ranging from 9 – 17 years participated. The inclusion criteria were that they were able to walk independently without using supporting assistance technology like crutches etc. and being able to understand and follow instructions. However, the use of an ankle-foot orthosis (AFO) was allowed as it is an aid to further improve or stabilize their ambulatory movement. The Gross Motor Functioning Classification System for Cerebral Palsy (GMFCS) classifies the gross motor function in five levels (I-V) based on functional limitations, and need of assistive technology etc. The severity of the physical impairment varies on an individual level, and can range from being able to walk without assistance to being fully dependent on assistive technology like electronic

wheelchairs etc. (31). To be included in the present study, the participants should have a GMFCS level of either I or II. 9 of the participants had a GMFCS I, and 7 participants had a GMFCS II. In terms of type of CP, the majority had bilateral spasticity, one hemiplegic on the left side, and one dyskinetic. All participants completed a short bout protocol. 7 participants also completed a long bout protocol. Table 1 shows descriptive characteristics of the two groups.

Table 1 Descriptive characteristics of the participants. The values are mean ± standard deviation. Divided into participation of the two different protocols. Many of the subjects participated in both protocols.

	TD (Total N= 48)		**CP (Total N=16)**	
	Short bout	Long bout	Short bout	Long bout
N	14	47	16	7
Age	11,14 ± 0,66	10,96 ± 2,70	11,44 ± 2,38	11,43 ± 3,31
Weight (kg)	46,91 ± 10,53	45,44 ± 13,20	43,23 ± 11,31	40,33 ± 13,94
Height (cm)	156,07 ± 7,7	152,34 ± 16,23	146,44 ± 11,17	145, 21 ± 15,13
Leg length (cm)	92,85 ± 5,21	90,67 ± 11,36	N/A	N/A

2.2 Data collection and equipment

In the present study, two protocols were performed: One semi-structured activity protocol containing activity bouts with short duration and frequent transitions and one structured activity protocol containing activity bouts with longer duration. To be able to test the validity of the NTNU-HAR model, the participants wore two tri-axial accelerometers and an action camera to document the performed activities.

2.2.1 Protocol

Short bouts

The participants in the TD group were asked to complete two activity protocols. The first protocol was a semi-structured protocol where the children performed different activities with short duration (<30 seconds), frequent transitions and several repetitions. It was developed to include activities often seen during free play in children. The activities ranged from sedentary activities like sitting, standing, and lying to more vigorous and complex activities like running, jumping, agility drills, scavenger hunt, and playing soccer or handball. This protocol was performed indoors and conducted in groups of 4 children at the time. The data collection was led by four research assistants.

Long bouts

The second protocol was a structured protocol with longer periods of activities (>3 minutes). The activities performed were activities common in daily life. This included walking, jogging, running, cycling, standing, and sitting. This protocol was performed outdoors on a running track if the weather permitted it, or indoors if not.

CP

The participants in the CP groups performed a similar protocol as described above, either a combination of the short bouts and long bouts, and/or a shorter and simplified version, adapted to their function and level of fatigue.

For a more detailed description of the protocols, see appendix 1

2.2.2 Axivity AX3

To assess physical activity, two Axivity AX3 (Axivity Ltd, Newcastle, UK) accelerometers were used. Acceleration was sampled at 100Hz, with a range of ± 8g. AX3 is a tri-axial accelerometer that measures raw acceleration in three axes (x, y, z). Its dimensions are 23 x 32,5 x 7,6 mm with a weight of 11grams (32).

The participants wore the two sensors fixed to the body. Elastic tape (Fixomull) was first attached to the skin, then surgical tape was used to attach the sensors on the elastic tape. The sensors were placed on the lower back (on the L3 spinal segment) and on the middle of the right thigh of the participants (Figure 1). For the CP subjects the accelerometer on the thigh was placed on the least affected side.

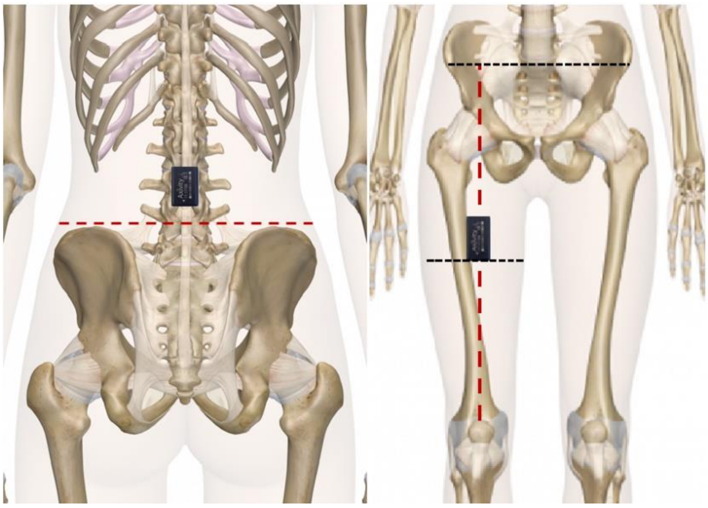


Figure 1 Anatomical placement of the Axivity AX3 sensors

2.2.3 Video recording

Video recording was used for observation of the physical activities being performed. During the data collection, the participants were recorded using a GoPro Hero 3+ camera. The camera was mounted on the body using a chest strap, recording from the chest and down on their feet, or placed on a tripod capturing the whole body during some parts of the protocol. The recordings were sampled at 60 frames per second (fps), with a resolution of 1080x720pixels.

2.3 Analysis

2.3.1 Video annotation

The video recordings were analyzed manually by labelling the type of PA being performed by the participants. This process is called annotation and was used as gold standard to assess the validation of the NTNU-HAR model in predicting and classifying types of PA.
Before the videos could be annotated, they were converted from MP4 to AVI format and down sampled from 60fps to 25fps. This was done using the MPEG Streamclip 1.2 for Windows (Squared 5 srl).

The annotation was done using the annotation tool ANVIL 6 (Kipp). This was done frame-by-frame for each participant. A total of 18 different types of PA were annotated according to a list of predefined activity definitions, based on work from previous or similar validation studies at NTNU. This included sitting, standing, walking, shuffling, stairs ascending, stairs descending, lying (prone, supine, right side, left side), cycling standing, cycling sitting, running, jumping, bending, picking, non-vigorous activity and undefined activity. All activities have a clearly defined description for when it starts and when it ends, as well as a general description of the activity. Also included are different postures, and transitions between posture and activities. E.g. walking was defined as:

*"Locomotion towards a destination with 1 stride or more, (1 step with both feet, where 1 foot is placed at the other side of the other), walking could occur in all directions e.g. forward, backwards, sideways. Walking along a curved line is allowed. From standing walking starts when walking direction is set and **heel off** occurs on the first foot. From transition or shuffling walking starts when walking direction is set and **heel off** occurs on the latter foot."*

The full list of activity definitions can be found in Appendix 2**.**
The completed annotations were exported to .txt format and synchronized with the acceleration signals.

2.3.2 Synchronization

To ensure that the accelerometers and the video recordings can be synchronized with each other, a reference point that is easily detected both in the accelerometer signal, and in the video recording, is needed. To make this reference point, the participants performed three heel drops (or jumps), before and after the data collection was completed. A heel drop is performed by going up on your toes, and then slamming the heels down to the ground forcefully. This was chosen as it is easy to spot the impact between the heel and ground, both in the acceleration signal from the sensors and in the video recordings. The heel drops were also used to synchronize the accelerometers worn on the lower back and thigh before the signals were run through the NTNU-HAR model.

2.3.4 The NTNU-HAR model

The acceleration data was run through the NTNU-HAR model to obtain a time series of predicted activities. The NTNU-HAR is a supervised machine learning model, which is based on a Random Forest (RF) classifier. In short, RF is a collection of several decision trees algorithms (33). A decision tree is a simple tree-based method that automatically creates a set of rules, based on simple threshold values of the signal input features that are extracted to classify the data (19). A simplified example of this could be if the angle thigh sensor has an angle greater than 45°, this equal sitting. In this context, the signal features from the sensors are the attributes (e.g. mean, amplitude etc.) that are used to decide the category, i.e. type of PA.

In total, 138 attributes from the signal features are calculated from the two sensors. They are calculated for windows of 5 seconds, with no overlap between the windows. Example: To calculate the mean of the x-acceleration in the back sensor, one takes 500 samples (5s window x 100Hz) and calculate the mean of these measurements. This makes one attribute. For the next window, the next 500 measurements are used, and so on. In a decision tree, the most important attributes are used first to classify the data. This means that during the training, the attribute that was capable of classifying most of the data is the most important, and so on.

A RF grows multiple decision trees, where each tree is independent of the others. This means that each tree is given a random subset of the data (bagging). In addition, in a RF the nodes/branches in a tree are given a random subset of the attributes/features. This creates diversity in the decision trees that are created, something that in general leads to better models (21, 33). In the end, each tree in the algorithm cast a vote for the most popular prediction. The final prediction class is the majority vote of all the trees. In other words, if 42 of 50 decision trees predicted the class as “sitting” and the remaining 8 trees predicted “standing”, the activity would be predicted as sitting.

2.3.4 Statistical analysis

The statistical analysis was done using MATLAB R2019b (The MathWorks Inc., US), Excel (Microsoft Office 2016) and IBM SPSS Statistics for Windows (Version 26, SPSS Inc., Chicago). A Pearson’s correlation test was conducted to check for the relation between age and the accuracy of the model.

The output from the testing of the NTNU-HAR model was .csv files containing the predicted activity class vs the annotated activity class. The results were given for each individual, as well as overall results for each group and protocol (TD short, long, (primary school, secondary school), CP).

All individual predictions from the HAR model were plotted against the annotations, for visual inspection, to ensure that the signals were synchronized.

To illustrate the performance of NTNU-HAR model, the predicted and annotated data will be presented in a confusion matrix. This is a table that shows the amount of correctly and incorrectly classified instances of each type of PA. From this we can assess the distribution of the classifications as true positives (TP), true negatives (TN), false positives (FP) and false negatives (NP). This distribution was used to calculate overall accuracy for each of the groups and protocols. Accuracy was calculated as the ratio of correctly classified instances to the total amount of instances:

$$Eq1: Accuracy = \frac{TP+TN}{TP+TN+FP+FN}.$$

To give a detailed assessment of the models’ performance on each class of PA, sensitivity, specificity, and positive predictive value (PPV) was also calculated. Sensitivity represents the ability of the classifier to select instances of a certain activity class. If we take the “walking” category as an example: when it is actually “walking”, how often does it predict as “walking”? Sensitivity, or the true positive rate, is the proportion of the correctly classified instances of a PA type over the actual number of instances of that PA type:

$$Eq2: Sensitivity = \frac{TP}{FN+TP}.$$

Specificity represents the ability of the classifier to not select instances of a certain activity class when it was not that activity class. Back to the walking-example: when it is not “walking”, how often does it classify as not “walking”? Specificity, or the true negative rate, is the ratio of how much of a PA-type was correctly classified as not belonging to that PA-type when it was not.

$$Eq3: Specificity = \frac{TN}{TN+FP}$$

PPV represents the probability that a detection of a particular activity is correct. Example: If an instance is predicted to be “walking”, how likely is it that that instance truly belongs to “walking”? PPV, or precision, is the proportion of correctly classified instances of a PA-type over all instances that was classified as that type of PA.

$$Eq4: PPV = \frac{TP}{TP+FP}.$$

There are no clear guidelines for what is considered as acceptable measures in human activity recognition research. This varies based on the goal of the research and the complexity of the study design etc. Based on a similar study, the results in the present study were considered excellent if they reached an accuracy ≥90 %, acceptable ≥80%, modest <80 %, and low <50% (22). Similarly, values for sensitivity, specificity and PPV were regarded as high above 90%, acceptable above 80%, modest between 50-80% and low below 50%

3. Results

The main results are split into four groups according to age and protocol performed. These are adolescents (long bout protocol), children (short bout) children (long bout) and CP (long + short). For some participants, an error in the synchronization of signals between the two accelerometers was encountered during the testing. This unfortunate event lead to exclusion of data from 6 participants in the adolescent group, 1 in the long bout protocol, and 2 in the short bout protocol in the TD children group, and 1 in the CP group.

The NTNU-HAR model achieved excellent accuracy for the long bouts protocol both in adolescence and in children, with 94.9% and 90.9 % respectively. For the short bouts, the model achieved an accuracy of 67.5%. In the CP group, the overall accuracy was 73.3 %. Table 2 shows the overall accuracy, and sensitivity, specificity and PPV for all PA types.

Table 2 Sensitivity, specificity, PPV and overall accuracy for TD and CP. Separated into age (adolescents and children), and type of protocol (long bout or short bout).

Activity	**Adolescents (Long)**			**Children (Long)**			**Children (Short)**			**CP (Long + short)**		
	Sens.	Spec.	PPV	Sens.	Spec.	PPV	Sens.	Spec.	PPV	Sens.	Spec.	PPV
Walking	0.96	0.97	0.96	0.91	0.95	0.92	0.88	0.79	0.50	0.80	0.87	0.62
Running	0.99	0.98	0.95	0.91	0.99	0.92	0.76	0.95	0.61	0.63	0.99	0.58
Shuffling	0.22	0.99	0.27	0.10	1.00	0.23	0.09	0.99	0.61	0.04	1.00	0.46
Stairs A	N/A	N/A	N/A	N/A	N/A	N/A	0.00	1.00	0.00	0.04	1.00	0.18
Stairs D	N/A	N/A	N/A	N/A	N/A	N/A	0.09	1.00	0.14	0.00	1.00	0.00
Standing	0.93	0.99	0.96	0.91	0.98	0.93	0.81	0.97	0.84	0.85	0.96	0.91
Sitting	0.99	1.00	0.98	0.96	0.99	0.94	0.93	0.97	0.89	0.93	0.98	0.92
Lying	N/A	N/A	N/A	N/A	N/A	N/A	0.84	0.99	0.94	0.85	0.99	0.81
Transition	N/A	N/A	N/A	N/A	N/A	N/A	0.38	1.00	0.59	0.26	0.99	0.23
Bending	N/A	N/A	N/A	N/A	N/A	N/A	0.02	1.00	0.13	0.05	1.00	0.36
Picking	N/A	N/A	N/A	N/A	N/A	N/A	0.03	1.00	1.00	0.03	1.00	0.50
Cycling (sit)	N/A	N/A	N/A	0.91	1.00	0.73	N/A	N/A	N/A	0.62	0.98	0.28
Cycling (stand)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.82	0.98	0.04
Non-vig. Act	N/A	N/A	N/A	0.00	1.00	0.00	0.09	1.00	0.03	0.21	0.98	0.09
Jumping	N/A	N/A	N/A	N/A	N/A	N/A	0.00	1.00	N/A	0.00	1.00	N/A
Accuracy	**0.949**			**0.909**			**0.675**			**0.733**		

Note: PPV=Positive predictive value; N/A=Not applicable. Activities not part of the protocol; Stairs A = Stairs ascending; Stairs D = Stairs descending; Non-vig. Act. = Non-vigorous activity.

During the long activity bouts protocol, sensitivity, specificity and PPV was high (>0.90) for the activities walking, running, standing, and sitting, both for the children and the adolescents. Specificity was high (>0.90) for all PA-types in every protocol, with the exception of walking (0.79) for the short activity bouts protocol and for CP (0.87). In the short bouts, sensitivity and PPV were lower for all activities compared to the long bout protocol.

Confusion matrixes showing how the activities are classified by the HAR-model for the different groups are presented in table 3-6.

3.1 Adolescents

The best performance of the NTNU-HAR model was found in the adolescents' group (age 13-15 years). The model correctly classified 4591 of 4835 instances (table 3), yielding an accuracy of 94.9 %. The categories walking, running, standing, and sitting had high sensitivity and PPV (>93%), with running and sitting having almost perfect sensitivity of 99%. Walking was mostly misclassified as running in 3% of total instances.
Sensitivity was low for shuffling (22%). Most of shuffling was misclassified as standing in 49% of the cases, and as walking 22% of the cases.
Standing was misclassified as walking 4% of the time, and 2 % as shuffling.

3.2 Children long bout protocol

For the long bout protocol in the children's group, 14253 of a total number of 15674 instances was classified correctly (table 4), giving an accuracy of 90.1%. Sensitivity and PPV was high (>91%) for walking, running, standing, sitting, and for cycling (sit), except for a lower PPV of 73% for cycling (sit).
Walking was mostly misclassified as running (2.8%), standing (1.8%) and cycling sit (1.7%). Running was misclassified as walking 7% of the time. Sensitivity was low for shuffling (10%). Shuffling was misclassified as standing 57% of the time and 27.6% of the time as walking. Some standing was misclassified as walking (6%).
Cycling sit was misclassified as sitting 7% of the time, and sometimes as standing (2%).

3.3 Children short bout protocol

In the short bout protocol, 4638 of 6867 instances were classified correctly (table 5), yielding an accuracy of 67.5%. Sensitivity was high for sitting (93%), and good (>80%) for walking, standing, and lying. Running had a lower sensitivity of 73%. Sensitivity was low for the remaining categories. PPV was high for lying (94%), good (<80) for standing and sitting, and modest (>50%) for running, transition, and walking.
Most notably, running was misclassified as walking 23.5% of the time. Shuffling was in most cases misclassified as walking (62%) and as standing (14%) and running (13%).
Ascending stairs was not correctly classified once but was misclassified as walking (61.4%) and running (38.6%). Similarly, descending stairs was misclassified as walking (49.1%) and running (42.1%).
Standing was misclassified as walking 15% of the time total instances. Lying was sometimes

misclassified as sitting (13%) and transition (3%).
Transition was most often misclassified as lying (20%), walking (16%) and non-vigorous activity (13%).
Bending and picking was poorly recognized by the model (38% and 2 %, respectively). Bending was misclassified mainly as walking (56.4%) and running (35,5%), while picking was misclassified mostly as sitting (50%) and walking (19.4%).
Jumping was misclassified in all instances, as the model was not yet trained to recognize this category. 79.3% was misclassified as walking and 20.7% as running.

3.4 CP

In the CP group, the model correctly classified 7442 of 10150 instances (table 6), giving an accuracy of 73.3%. Sensitivity was high for sitting (93%) and good (>80%) for walking, standing, lying, and cycling stand. It was modest (<63%) for running and cycling sit. It was low (<26%) for the rest of the categories.
Except for walking (87%), specificity was high (>96%) for all categories. PPV was high (>90%) for standing and sitting, good for lying (81%), and modest (≥59%) for walking running. It was low (≤46%) for the remaining categories.

Walking was mostly misclassified as running (5%) and standing (4%). Running was misclassified as walking in 34% of the total instances.
Shuffling was mostly misclassified as walking (56%), and some standing (19%).
Ascending stairs mainly misclassified as walking (76%) and cycling sit (13%). The model failed to classify descending stairs correctly. As with ascending stairs it was mostly misclassified as walking (90%).
Standing was misclassified as walking in 11% of the total instances. Transition was misclassified mostly as non-vigorous activity (23%) and lying (22%).
As with the TD short group, bending was mostly misclassified, mainly as walking (34 %), non-vigorous (26%) and cycling sit (15%). Picking was misclassified as walking (24%), non-vigorous activity (24%), sitting (15%) and cycling sit (14 %).
Cycling sit was misclassified as sitting in 14% of the total instances, 12 % as cycling stand, and 9% as walking. Of the few instances of cycling stand, most was classified correctly (81%). The remaining 2 instances was misclassified as cycling sit (18%.). Non-vigorous activity was misclassified mostly as transition (24%), walking (19 %) and standing (8.6%) and cycling sit (8.6%). Jumping was misclassified in all instances. 85.9% was misclassified as walking and 14.1% as running.

Table 3 Confusion matrix for the adolescents (N=10). The numbers in the cell are instances (top, black) and (nearest) percentage (%) of total number of instances (bottom, in grey). The colored boxes represent the amount of correctly identified instances of each PA-type. Rows represent the labeled PA-types, and the columns the predictions from the NTNU-HAR. The column to the right is the total number of instances annotated in each category. The bottom row is the total number of instances detected by the model. Activities that were neither annotated nor predicted, was not included in the table.

Annotated activity	Predicted activity									
	Walking	Running	Shuffling	Stairs ascend	Stairs descend	Standing	Sitting	Bending	Cycling (sit)	Total Annotated
Walking	1871 96	56 3	6 <1	3 <1	0 -	12 <1	3 <1	0 -	0 -	**1951**
Running	12 1	1091 99	0 -	0 -	0 -	1 <1	0 -	0 -	0 -	**1104**
Shuffling	13 22	1 2	13 22	0 -	0 -	29 49	2 3	0 -	1 2	**59**
Stairs ascend	3 75	0 -	0 -	1 25	0 -	0 -	0 -	0 -	0 -	**4**
Stairs descend	4 100	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**4**
Standing	50 4	4 <1	29 2	0 -	0 -	1132 93	3 <1	0 -	5 <1	**1223**
Sitting	3 <1	0 -	0 -	0 -	0 -	1 -	483 99	0 -	0 -	**487**
Bending	0 -	0 -	0 -	0 -	0 -	3 100	0 -	0 -	0 -	**3**
Cycling (sit)	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**0**
Total Predicted	**1956**	**1152**	**48**	**4**	**0**	**1178**	**491**	**0**	**6**	**4835**

Table 4 Confusion matrix for TD children (long bout protocol) (N=30). The numbers in the cell are instances (top, black) and (nearest) percentage (%) of total number of instances (bottom, in grey). The colored boxes represent the amount of correctly identified instances of each PA-type. Rows represent the labeled PA-types, and the columns the predictions from the NTNU-HAR. The column to the right is the total number of instances annotated in each category. The bottom row is the total number of instances detected by the model. Activities that were neither annotated nor predicted, was not included in the table.

Annotated activity	Predicted activity													
	Walking	Running	Shuffling	Stairs ascend	Stairs descend	Standing	Sitting	Lying	Transition	Bending	Picking	Cycling (sit)	Cycling (stand)	Non-vig. Act.	Total Annotated
Walking	5437 91	169 3	22 <1	34 <1	5 <1	111 2	49 <1	0 -	1 <1	1 <1	0 -	100 2	15 <1	1 <1	**5945**
Running	177 7	2226 91	0 -	3 <1	0 -	10 -	28 1	0 -	0 -	0 -	0 -	1 <1	0 -	0 -	**2445**
Shuffling	54 28	0 -	19 10	0 -	0 -	112 57	6 3	0 -	0 -	0 -	0 -	3 2	1 <1	0 -	**195**
Stairs ascend	1 100	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**1**
Stairs descend	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**0**
Standing	230 6	30 <1	37 <1	0 -	0 -	3731 91	38 <1	0 -	0 -	1 <1	5 <1	20 <1	0 -	3 -	**4095**
Sitting	37 1	0 -	5 <1	0 -	0 -	37 1	2480 96	4 <1	0 -	0 -	0 -	8 <1	0 -	1 <1	**2572**
Lying	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**0**
Transition	0 -	0 -	1 14	0 -	0 -	2 29	4 57	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**7**
Bending	2 100	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**2**
Picking	0 -	1 20	0 -	0 -	0 -	1 20	2 40	0 -	0 -	0 -	0 -	1 20	0 -	0 -	**5**
Cycling (sit)	0 -	0 -	0 -	0 -	0 -	9 2	26 7	0 -	0 -	0 -	0 -	360 91	0 -	0 -	**395**
Cycling (stand)	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**0**
Non-vig. Act.	2 17	0 -	0 -	0 -	0 -	2 17	6 50	0 -	0 -	0 -	0 -	2 17	0 -	0 -	**12**
Total Predicted	**5940**	**2426**	**84**	**37**	**5**	**4015**	**2639**	**4**	**1**	**2**	**5**	**495**	**16**	**5**	**15674**

Table 5 Confusion matrix for TD children (short bout protocol) (N=12). The numbers in the cell are instances (top, black) and (nearest) percentage (%) of total number of instances (bottom, in grey). The colored boxes represent the amount of correctly identified instances of each PA-type. Rows represent the labeled PA-types, and the columns the predictions from the NTNU-HAR. The column to the right is the total number of instances annotated in each category. The bottom row is the total number of instances detected by the model.

Annotated activity	Predicted activity														
	Walking	Running	Shuffling	Stairs ascend	Stairs descend	Standing	Sitting	Lying	Transition	Bending	Picking	Cycling (sit)	Cycling (stand)	Non-vig. Act.	Jumping	Total Annotated
Walking	1345 88	58 4	26 2	12 <1	21 1	30 2	20 1	0 -	2 <1	0 -	0 -	6 <1	2 <1	4 <1	0 -	**1526**
Running	164 23	530 76	0 -	1 <1	3 <1	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**698**
Shuffling	447 62	93 13	67 9	1 <1	5 <1	100 14	6 <1	0 -	0 -	4 <1	0 -	1 <1	0 -	2 <1	0 -	**726**
Stairs ascend	35 61	22 39	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**57**
Stairs descend	28 49	24 42	0 -	0 -	5 9	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**57**
Standing	153 15	9 <1	13 1	1 <1	1 <1	834 81	13 1	0 -	0 -	1 <1	0 -	1 <1	0 -	0 -	0 -	**1026**
Sitting	26 2	0 -	2 <1	0 -	0 -	22 2	1256 93	19 1	6 <1	0 -	0 -	5 <1	0 -	9 <1	0 -	**1345**
Lying	0 -	0 -	0 -	0 -	0 -	0 -	88 13	561 84	17 3	0 -	0 -	0 -	0 -	0 -	0 -	**666**
Transition	16 16	0 -	2 2	0 -	0 -	1 1	6 6	19 20	37 38	0 -	0 -	3 3	0 -	13 13	0 -	**97**
Bending	35 56	22 35	0 -	0 -	0 -	0 -	1 2	0 -	0 -	1 2	0 -	2 3	0 -	1 2	0 -	**62**
Picking	7 19	0 -	0 -	0 -	0 -	1 3	18 50	0 -	1 3	2 6	1 3	3 8	0 -	3 8	0 -	**36**
Cycling (sit)	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**0**
Cycling (stand)	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**0**
Non-vig. Act.	3 27	1 9	0 -	0 -	0 -	1 9	4 36	0 -	0 -	0 -	0 -	1 9	0 -	1 9	0 -	**11**
Jumping	444 79	116 21	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**560**
Total Predicted	**2703**	**875**	**110**	**15**	**35**	**989**	**1412**	**599**	**63**	**8**	**1**	**22**	**2**	**33**	**0**	**6867**

Table 6 Confusion matrix for CP group ((N=15). The numbers in the cell are instances (top, black) and (nearest) percentage (%) of total number of instances (bottom, in grey). The colored boxes are the amount of correctly identified instances of each PA-type. Rows represent the labeled PA-types, and the columns the predictions from the NTNU-HAR. The column to the right is the total number of instances annotated in each category. The bottom row is the total number of instances detected by the model.

Annotated activity	Predicted activity														
	Walking	Running	Shuffling	Stairs ascend	Stairs descend	Standing	Sitting	Lying	Transition	Bending	Picking	Cycling (sit)	Cycling (stand)	Non-vig. Act.	Jumping	Total Annotated
Walking	1926 80	114 5	19 <1	9 <1	8 <1	84 4	48 2	0 2	4 <1	3 <1	0 -	55 2	78 3	47 2	0 -	**2395**
Running	112 34	206 63	0 -	0 -	0 -	1 <1	0 -	0 -	0 -	0 -	0 -	3 <1	4 <1	2 <1	0 -	**328**
Shuffling	459 56	4 <1	33 2	0 -	1 <1	156 19	3 <1	0 -	1 <1	4 <1	0 -	58 7	63 8	30 4	0 -	**812**
Stairs ascend	34 76	0 -	0 -	2 4	0 -	1 2	0 -	0 -	0 -	0 -	0 -	6 13	1 2	1 2	0 -	**45**
Stairs descend	26 90	2 7	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	1 3	0 -	0 -	0 -	**29**
Standing	349 11	6 <1	15 <1	0 -	1 <1	2747 85	35 1	0 -	12 <1	5 <1	0 -	39 1	12 <1	11 <1	0 -	**3232**
Sitting	13 <1	0 -	2 <1	0 -	0 -	18 <1	2017 93	42 2	35 2	0 -	0 -	11 <1	1 <1	29 1	0 -	**2168**
Lying	6 1	0 -	0 -	0 -	0 -	3 <1	35 8	355 85	18 4	0 -	0 -	0 -	0 -	1 <1	0 -	**418**
Transition	11 9	0 -	0 -	0 -	0 -	2 2	12 10	26 22	31 26	3 3	0 -	4 3	3 3	28 23	0 -	**120**
Bending	66 34	7 4	2 1	0 -	0 -	4 2	2 1	2 1	4 2	10 5	0 -	16 8	29 15	50 26	0 -	**192**
Picking	19 24	0 -	0 -	0 -	0 -	1 1	12 15	5 6	5 6	3 4	2 3	11 14	2 3	19 24	0 -	**79**
Cycling (sit)	13 10	0 -	0 -	0 -	0 -	2 2	18 14	0 -	0 -	0 -	0 -	82 62	16 12	1 <1	0 -	**132**
Cycling (stand)	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	2 18	9 82	0 -	0 -	**11**
Non-vig. Act.	20 19	2 2	0 -	0 -	0 -	9 9	5 5	6 6	25 24	0 -	2 2	9 8	4 4	22 21	0 -	**104**
Jumping	73 86	12 14	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	0 -	**85**
Total Predicted	**3127**	**353**	**71**	**11**	**10**	**3028**	**2187**	**436**	**135**	**28**	**4**	**297**	**222**	**241**	**0**	**10150**

3.5 Effect of age and length of activity period

The results showed no clear effect of age. A Pearson correlation test revealed no significant correlations (R-value <±0.2, p-value >0.2) between age and accuracy in any of the groups. No significant correlation was found between GMFCS-level and accuracy in the CP group (R=-0.4, p=0.08). Figure 2 shows the relationship between the overall accuracy for each individual and age in the different protocols.

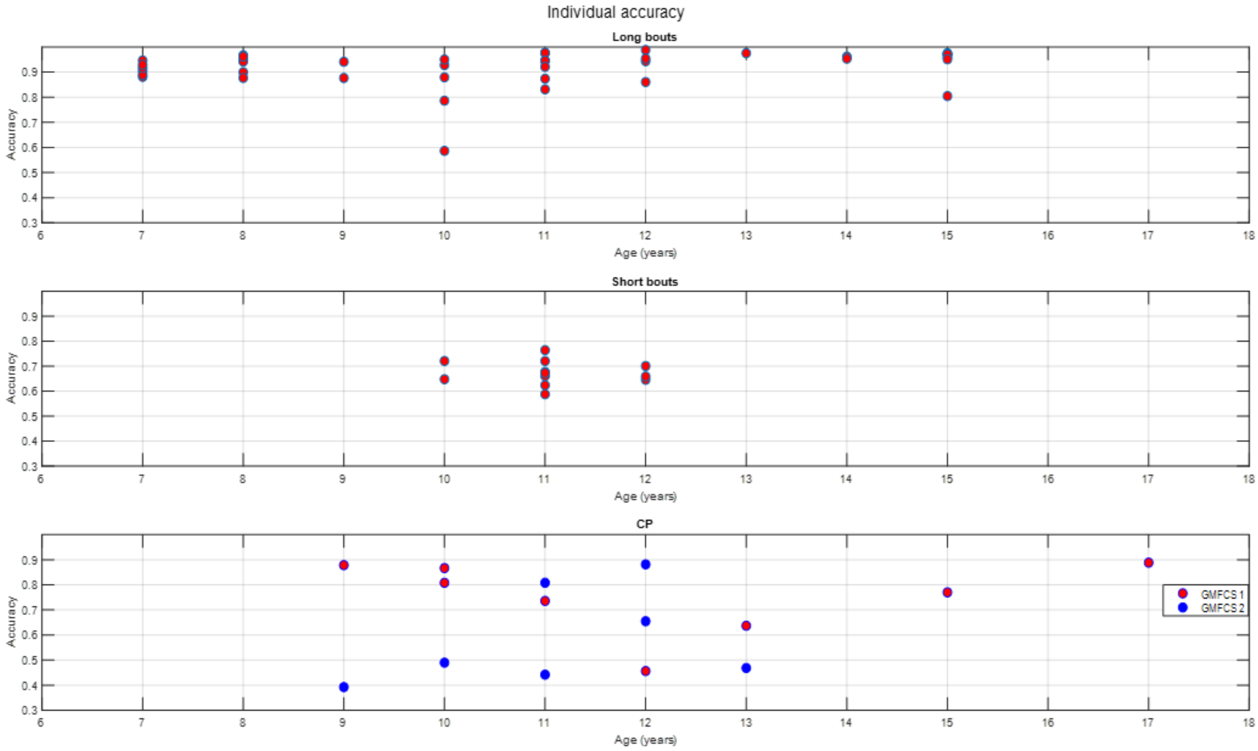


Figure 2 individual accuracy sorted after age. The figure on the top shows the accuracy for all participants in the long bout protocol. The middle figure shows the accuracy in the short bout protocol. The bottom figure shows the individual accuracy in the CP group, sorted after age, and their GMFCS level. Red = GMFCS 1, blue = GMFCS 2

The overall accuracy was higher for the long bout protocol compared to the short bouts protocol. Figure 3 shows the distribution of PA type from the long bout protocol for one participant, and the NTNU-HAR's ability to match the predictions against the annotations. Figure 4 shows the distribution from the short bout protocol for the same participant. The short bout protocol contains a lot more changes in PA-type, and each activity is performed for a short duration.

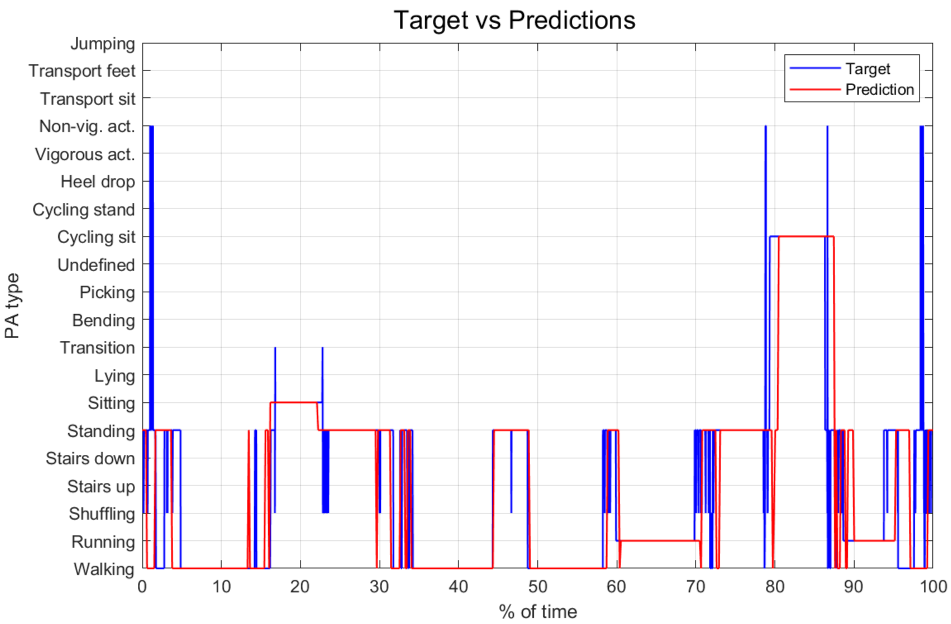


Figure 3 Plot of labeled activities (blue) and predicted activities (red) from one participant during the long bout protocol

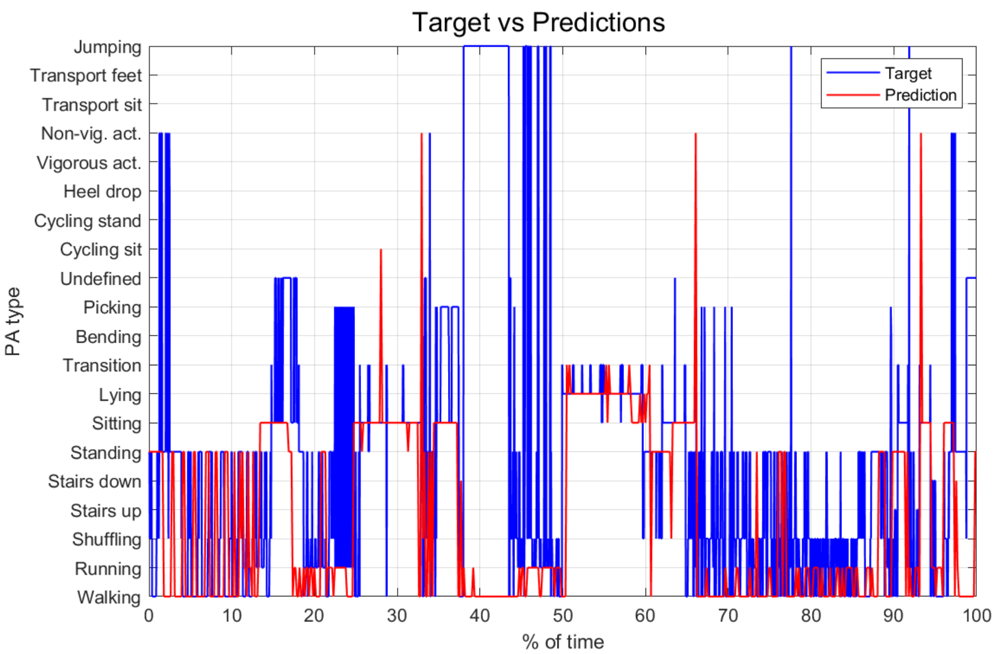


Figure 4 Plot of labeled activities (blue) and predicted activities (red) from the same participant in during the short bout protocol

4. Discussion

The purpose of this thesis was to assess the performance of the NTNU-HAR model in detecting and classifying types of PA in typically developing children and adolescents, and for children and adolescents with cerebral palsy. The main results show that the NTNU-HAR model performs well in TD children and adolescents when the activities are performed over a longer time span. The performance is poorer in the CP group. However, the mixed protocols in this group (short + long bouts duration) might affect their overall accuracy since the performance of the model seems to drop when the protocol include frequent changes in activity. There was no effect of age on the performance of the model.

4.1 Long bouts vs short bouts

The model achieved acceptable results when classifying activity from the long bout protocol, with an overall accuracy of 94.9% in adolescents and 90.9% in children. The activities walking, running, sitting, standing, and lying were classified with good to high (80-99 %) sensitivity, specificity, and PPV. This is similar, or close to, the overall accuracy of 94% that the NTNU-HAR has achieved in adults. Our results are also comparable to other studies done on activity recognition in children using raw acceleration signals, with overall accuracies ranging from 87.5-92.4% (22, 34-36). These studies vary in protocol, accelerometer placement, and type of PA predicted. Most notably none of the other studies placed accelerometers on the lower back or thigh, but rather on hip, wrist and/or ankle. This is a common challenge when comparing studies in the field of human activity recognition and machine learning as there is no general consensus regarding accelerometer placement, signal feature extraction and choice of machine learning technique (37)

Interestingly, the study by Mannini, Rosenberger, Haskell, Sabatini and Intille (35) first tested a support vector magnitude ML-model that was developed for adults on activity data from children reaching an accuracy of 85,9% and 89.7% on sensors placed on wrists and ankles, respectively. Then the model was further trained on features from the data collected on children and tested again. This led to the overall accuracy improving to 91.0% and 92.4% on wrist and ankle. This indicates that inclusion of data from children and adolescents in the training of a model might improve the performance of the model further. Future research should therefore try to train the NTNU-HAR model on the dataset collected in this study.

For the short bout protocol however, the model did not reach acceptable results, as the accuracy was 61.9 %. However, it was able to separate between the different postures, as sitting, standing, and lying were classified with good to high (80-99 %) sensitivity, specificity, and PPV.

Several factors might explain this difference in performance between the long- and short bout protocols. During the long bout protocol, the participants performed the same activity for several minutes before doing another type of PA. This could be for example walking 5 minutes, followed by standing and so on. The short bout protocol contained activities of shorter duration, and more complex tasks like agility drills, scavenger hunt, football etc. These tasks are composed of several different types of PA that changes during the exercise. E.g. during the football activity, there are a lot of changes between walking and running and quick change of direction. This might explain the some of the misclassification of running as walking in 22% of the instances. Thus, the rapid changes in types of PA and direction of movement cause misclassification for the NTNU-HAR model.

The design of the NTNU-HAR can also explain the difference in performance between the two protocols. The current iteration of the model operates with a window size of 5 seconds to calculate the signal features from the acceleration signal that are used to classify what activity is being performed. This might have led to activity bursts lasting shorter than 5 seconds to be "overlooked" during the classification. It is not unlikely that a smaller window size would be able to better detect and classify more of the frequent and quick changes in PA-type, thus improving the performance of the model. This is something that should be tested in future studies. However, this demands that the whole model needs to be retrained.

It is important to end up with a model that is able to classify PA-type during shorter activity periods to obtain as precise estimates of PA as possible. As a lot of children's PA occurs as free play and (ball)games, the model might underestimate their levels of PA if it is unable to correctly classify this type of activity. An alternative to retraining the model could be to run post analyzes that extracts periods with frequent changes in activity and classifies this as e.g. "other physical activity". Then these periods could be further analyzed to determine if this is activity of high or low intensity, based on e.g. the amplitude of the accelerometer signal, similar to what is done using activity counts.

The inclusion of two different protocols is a strength of this thesis. This gives a varied and diverse set of activities that covers a broad spectrum of everyday PA typically seen in children. The short bout protocol was semi-structured, both in terms of tasks being performed, and the way instructions were given. In the more complex part of the protocol, e.g. during the scavenger hunt, no specific instructions were given on how they were to perform the task (e.g. walk, or run etc.), only that they were to find and collect pieces of a puzzle. Thus, the participants did what was natural for them to do. I.e. some chose to alternate between running and walking, sit down on the floor to complete the puzzle and so on. This set-up helps to eliminate what is often termed as “laboratory walk” where participants in a standardized laboratory setting often performs an activity differently than what they would have done outside of the lab on their own.

4.2 Cerebral palsy

The overall accuracy in the CP group was 73,3 %. Not many studies have been conducted on human activity recognition using raw acceleration on children and adolescents with cerebral palsy. Ahmadi et al. (38) developed and tested different machine learning models (Random Forest and SVM) for activity recognition in children and adolescents with CP using accelerometers worn on the hip and wrist, reaching overall accuracies from 86.2-89.0 % in predicting the categories sedentary, standing movements (e.g. folding laundry), comfortable walking and brisk walking. This corresponds to the categories sitting, lying, standing, and walking in our data. Our results are comparable as the NTNU-HAR reached a sensitivity, specificity and PPV above 80% for those categories, except for a PPV of 62 % in walking. In addition, Hegde et al. (39) assessed the performance of a machine learning procedure (Multinomial Logistic Discrimination (MLD)) in classifying sitting, standing, and walking on data collected from a shoe based wearable system using data from pressure sensors and an accelerometer. An average accuracy of 95.3% was reached in classifying these categories.

The overall accuracy of the NTNU-HAR model did not reach that high accuracy. There are several factors that can explain the modest accuracy of the NTNU-HAR model in the CP group compared to the studies from Ahmadi et al. and Hegde et al. First, our model was not trained and developed on data collected from children with CP, but on data from healthy adults. Further, the protocol in the studies mentioned above were standardized lab protocols. The protocol in our study was a combination of activities of long duration, and of short duration more in resemblance of free living/free play. Also, more PA-types were predicted in our study.

Compared with the results from the short bout protocol in TD children, the model achieved a higher accuracy (73,3% vs 60,9%) in the CP group. This can partially be explained in the differences in the protocol between them as many of the CP participants performed a combined protocol of both long and short activity bouts. This might indicate that, like their TD counterpart, it is the PA characteristics (e.g. length of activity bouts) and not the movement characteristics that affect the performance of the model the most. However, future research should investigate this more thoroughly by separating the two different protocols for this group as well. The main reason the protocols were mixed in this study was to adapt to their function and level of fatigue, while securing that elements from both protocols from the

TD group would be performed. It was difficult to separate between participation in short bout and long bout in the results as the participants performed the mixed protocol as one continuous trial.

CP is a diverse group with a lot more individual variation between the participants in terms of movement characteristics. There was some indication of a negative correlation (R=-0.4) that could be found between GMFCS-level and individual accuracy. However, this was not statistically significant at a 5 % significance level. It is more likely that the difference in protocol is the most important factor. The six participants with the lowest individual accuracies completed only the short bout protocol. This was done at the outpatient's centers. The data collection was done collectively in groups by three participants plus one instructor and consisted of a wide array of different games and activities similar to the short bout protocol in the TD group.

The differences in the protocol between the TD groups and the CP group is a limitation of this study, as a direct comparison cannot be properly made between them. Further, the variation between the CP participants in terms of the execution of the protocol complicates the results. Future research should seek to increase the sample size, as well as implement a protocol that is performed similarly by all participants.

4.3 Misclassified PA types

The shuffling category was a cause for a lot of misclassifications with sensitivity below 10% for all four groups. Most instances were misclassified as walking, followed by standing and running. Shuffling is defined as: "Stepping in place, by non-cyclical movement of the feet, turning on the spot with feet movement not as part of walking bout". Thus, shuffling is similar to walking and standing in execution. It is therefore hard to separate between those categories. Further, shuffling could occur in-between periods of walking or standing if the activity were interrupted by irregular movements of the feet. The 5 second window size of the NTNU-HAR means that a lot of PA types could have been overlooked during the classifications, especially during more complex periods with a lot of changes in PA, leading to more misclassifications. This is evident for shuffling, as the model predicted fewer instances compared to the annotations with a larger discrepancy in the short bout protocol. For the short bout protocol, shuffling was predicted in 110 instances vs 726 actual instances (67 correct). In the CP protocol, where the long and short bout protocol was combined, 71 instances were predicted out of 812 (33 correct).

The shuffling category is a cause of misclassification in adults as well. This led to it being removed in the most recent study/development of the NTNU-HAR model in adults as it is difficult to recognize (21). In a health perspective it is also questionable if shuffling has an independent effect of health. Compared with standing, it is likely that the energy expenditure is only slightly higher during shuffling due to movement of the feet. Thus, both activities would most like be classified as low intensity. This would also apply for shuffling during walking, though the energy expenditure would most likely be higher for walking than for shuffling.

Stair walking was a part of the short bout protocol. Both ascending and descending stairs were rarely correctly classified in both the TD group and the CP group with sensitivity ranging from 0-9%. Most of the misclassification was as walking and some running in the TD group. In the CP group, almost all instances were misclassified as walking, except that some ascending stairs was also misclassified as cycling sit. In the TD group descending stairs was quite evenly misclassified between walking and running. Compared to the CP group, many of the participants in the TD groups ran up and down the stairs during the data collection. This might explain the larger share of misclassifications as running compared to the CP participants. It is concerning that the current iteration of NTNU-HAR seemingly fails to classify stair walking. As with shuffling, stair walking was also a problem in adults (21). However, in contrast with shuffling, it can be argued that climbing stairs has another effect on health compared to walking flat. Climbing stairs require one to pull its bodyweight against gravity, resulting in a greater energy expenditure compared to regular walking. Thus, it is important to be able to separate between these two activities.

The transition category was often misclassified as lying, walking and non-vigorous activity. As this category describes the transition from one activity to another, transition often occurred between two categories. Consequently, similarities in the acceleration pattern between transition and the categories it precedes or follows, might explain some of this misclassification, as well as the length of the transition period.

Bending and picking were two other categories with low sensitivity scores. In the TD group it was mostly misclassified as walking, and some running. Some explanation might come from how the activities performed in the short bout protocol was designed. Most of the bending and picking was done during an agility drill where the goal was to fetch and collect objects lying on the ground as fast as possible. Thus, most of the bending (and picking) was done whilst running or walking fast, in a tempo that might have been too fast for the NTNU-HAR to be able to recognize it has bending. This certainly applies to the picking category, as this is the period in between bending down and rising again, where an object is picked up from the ground. In many instances this would correspond to only 1 frame of the video.

Bending is a category that is also often misclassified in adults by the NTNU-HAR model, with a sensitivity of 20.9% (21). A notable difference is that it is mostly misclassified as sitting in adults. A possible explanation of this could be that bending in general is quite similar to sitting, especially if it is executed over a long time period that might be more commonly seen in adults than children. Same as with shuffling, it might not be the most important to correctly classify instances of bending and picking in children. If these instances happens frequently over a short period, as was the case in the protocol, what would be important in a health perspective is to be able to recognize whether these events might be as a part of a complex activity that is likely of moderate or vigorous intensity, or not.

The NTNU-HAR is not yet trained to recognize and classify jumping. As seen in the results, this is misclassified as mostly walking and running, and even some sitting. Jumping accounted for 7% of all instances in the short bout protocol, and this misclassification have a negative impact on the results.

The activities that are most often misclassified in this study is a general weakness of the NTNU-HAR as it misclassifies these activities regardless of population (adults or children). I.e. the reasons for misclassification of the abovementioned PA-types in this study is not likely due to any movement- characteristics or behavior that is typical for children, but a general problem of the model.

4.4 Effect of age.

The plotting of individual accuracy and age indicated no effect of age in the performance of the model. However, when comparing the results from the long bout protocol between the adolescents and children groups, the adolescents still had a higher overall accuracy (94.7-8% vs 90.8%) as well as sensitivity, specificity and PPV for the different PA types. This might indicate that adolescents have become more similar to adults in terms of movement characteristics. At the same time, the sample size for the adolescents was relatively small, especially due to the omission of five subjects due to synchronization errors between the two accelerometers (N=10). This gives little power to suggest that there is a difference between the two groups. Also, only the long bout protocol data was analyzed in this group. To further assess this, data from the short bout protocol should be included.

4.5 Implications/Future steps

A main finding from the result indicate that it is not necessary the movement characteristics of children compared to adults that affects the performance, but the activity pattern. Children have a different activity pattern compared to adults, with a rapid and transitory nature of PA, especially during free play. This can also be extrapolated to highlight the challenges of classifying complex PA as seen in many organized sports, e.g. soccer or handball. Children and adolescents has a high participation rate in team sports, especially in the U.S. and Europe (40). A next step in the development of the HAR model would therefore be to train the model on the dataset from the short bout protocol in the hope to be able to recognize and classify types of PA during intense and complex physical activity. This also includes training the model to recognize and classify jumping as a separate category.

4.6 Strength/Limitations

In addition to what has been discussed in the chapters above, there are some other strengths and limitations that is worth pointing out.

The annotation of video recordings on a frame-by-frame basis is a strength of this thesis. It is an accurate method of observation and gives the opportunity to playback the video after the data collection has ended. This is in contrast to other methods of observation, like manual observation or subject diaries, that is to a greater extend subject to recall bias or failing to properly record everything that was done during more complex protocols.

At the same time, the annotations are only as good as the person doing the annotating. It is not unlikely that classification errors have occurred during the manual labeling process. Especially during the more complex sessions in the semi-structured protocols, the probability of human errors increases. Further the annotation of video is a time-consuming process. In general, one 20-minute recording from the semi-structured protocol took about 2-3 hours to complete. A consequence of that was that only video recordings from only 16 participants performing the semi-structured protocol was annotated. Further analysis and/or development should try to include data from the remaining participants in the semi-structured protocol.

An Inter-Rater-Reliability (IRR) score was not calculated in this study, which is a big limitation. IRR is a measure of agreement between the observers (41). In this case between the raters that annotated/labeled the video recordings. Though most of the annotations was done by one person, some work had already been done prior to the start of this study by another person. As such, an IRR should have been calculated to check if there were any significant differences between the raters. However, due to unforeseen events this could not be achieved in time for this study.

The NTNU-HAR model is open source, using raw acceleration signals. This ensures full transparency, and that the method is available to other researchers.

The subject size for the TD group is another strength of this thesis. However, it could also have been even larger had there been more time to annotate the rest of the video recordings. Fredson, Pober and Janz recommended (in 2005) to include at least 10 participants for every other age group (e.g. age 6-7, 8-9 etc.) to obtain a representative sample in order to account for potential age differences when conducting calibration studies in children (42). This was related to accelerometry and energy expenditure. Prior to this study, a goal of at least 6 participants per age group was targeted. This was achieved to a certain extend in the long bout protocol. There were also some dropouts during the testing of the model due to other technical difficulties. Unfortunately, 6 participants were omitted due to failure in synchronizing the two accelerometers on the lower back and thigh.

The inclusion of children with CP is another strength of the study as it gives a broader assessment of the model's performance in different study populations. CP was chosen as it is the most common physical disability among children. Children with CP are less physically active than their typically developing (TD) peers (43). As with TD children, children with CP that are physically active during childhood and adolescence are more likely of being active as adults (44). The same benefits and effects of physical activity and inactivity also applies for children with CP. As such, it is concerning that the majority fails to meet the physical activity guidelines and spend a lot of their time sedentary (43). The development of good analytical methods for assessment of PA in this population is important as it will lead to more accurate and precise evaluation of the effect of treatment to increase PA, as well as measurements of PA levels for this population.

There were some problems occurring during the testing of the model. As several sensor pairs could not be synchronized with each other, data from several participants were omitted. Further the synchronization between the start time from the annotated data and the accelerometers was not always on point, as can be seen in figure 3. When that was the case the starting time between the targets and the prediction needed to be adjusted "manually" to find the best fit between them. It is possible that this is a cause of increased misclassification in some participants due to a displacement in the start time. This was especially a case in the CP group where the starting time for the heel-drops was not always recorded during the data collection.

4.7 Conclusion

Based on the results from the present study we can conclude that the NTNU-HAR is a valid tool for classifying physical activity types in TD children if activities are performed over longer periods of time. However, the length of the activity periods, and thus the activity pattern of children, have an effect on the performance of a HAR-model in predicting PA-types, with accuracy decreasing if the length of the activity bouts are short. This might indicate a poorer validity of the model during free play. There was no effect of age on the performance of the model, and the activities that are most often misclassified are shown to be a general problem of the model.

Future research should seek to train the model on the whole dataset that was available for this study, as well as assessing if a shorter window size would further improve the results.

For children with CP, promising results were found, but differences in protocol both compared to the TD group as well as between the participants makes it hard to draw conclusions whether a new ML-model needs to be developed exclusively for this group, or only trained on short/complex physical activities. This should be further investigated in future studies.

5. References

1. Gonzalez K, Fuentes J, Marquez JL. Physical Inactivity, Sedentary Behavior and Chronic Diseases. Korean journal of family medicine. 2017;38(3):111-5.
2. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet (London, England). 2012;380(9838):219-29.
3. Reiner M, Niermann C, Jekauc D, Woll A. Long-term health benefits of physical activity–a systematic review of longitudinal studies. BMC public health. 2013;13(1):813.
4. Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. Journal of applied physiology. 2008;105(3):977-87.
5. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland2010.
6. Mintjens S, Menting MD, Daams JG, van Poppel MNM, Roseboom TJ, Gemke R. Cardiorespiratory Fitness in Childhood and Adolescence Affects Future Cardiovascular Risk Factors: A Systematic Review of Longitudinal Studies. Sports Med. 2018;48(11):2577-605.
7. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts. 2009;2(3):187-95.
8. Rangul V, Bauman A, Holmen TL, Midthjell K. Is physical activity maintenance from adolescence to young adulthood associated with reduced CVD risk factors, improved mental health and satisfaction with life: the HUNT Study, Norway. The international journal of behavioral nutrition and physical activity. 2012;9:144.
9. Bringolf-Isler B, Schindler C, Kayser B, Suggs LS, Probst-Hensch N. Objectively measured physical activity in population-representative parent-child pairs: parental modelling matters and is context-specific. BMC Public Health. 2018;18(1):1024.
10. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. Lancet Child Adolesc Health. 2020;4(1):23-35.
11. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput JP, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. Appl Physiol Nutr Metab. 2016;41(6 Suppl 3):S240-65.
12. Liu M, Wu L, Yao S. Dose-response association of screen time-based sedentary behaviour in children and adolescents and depression: a meta-analysis of observational studies. British journal of sports medicine. 2016;50(20):1252-8.

13. Sedentary Behaviour Research N. Letter to the Editor: Standardized use of the terms "sedentary" and "sedentary behaviours". Applied Physiology, Nutrition, and Metabolism. 2012;37(3):540-2.
14. Keane E, Li X, Harrington JM, Fitzgerald AP, Perry IJ, Kearney PM. Physical Activity, Sedentary Behavior and the Risk of Overweight and Obesity in School-Aged Children. Pediatric exercise science. 2017;29(3):408-18.
15. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. British journal of sports medicine. 2014;48(13):1019-23.
16. Bassett DR, Jr., Rowlands A, Trost SG. Calibration and validation of wearable monitors. Medicine and science in sports and exercise. 2012;44(1 Suppl 1):S32-8.
17. Rowlands AV, Olds TS, Hillsdon M, Pulsford R, Hurst TL, Eston RG, et al. Assessing sedentary behavior with the GENEActiv: introducing the sedentary sphere. Medicine and science in sports and exercise. 2014;46(6):1235-47.
18. SAS Institute. Machine Learning: What it is and why it matters [cited 2020 Mar 25]. Available from: https://www.sas.com/en_us/insights/analytics/machine-learning.html.
19. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. Physiological measurement. 2009;30(4):R1-33.
20. Norges teknisk-naturvitenskapelige universitet (NTNU). HUNT4 [cited 2020 Mar 25]. Available from: https://www.ntnu.no/hunt/hunt4.
21. Reinsve Ø. Data analytics for hunt: Recognition of physical activity on sensor data streams: NTNU; 2018.
22. Trost SG, Zheng Y, Wong WK. Machine learning for activity recognition: hip versus wrist data. Physiological measurement. 2014;35(11):2183-9.
23. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Research quarterly for exercise and sport. 2000;71(sup2):59-73.
24. Williams CA, Wood L, De Ste Croix M. Growth and maturation during childhood. Paediatric Biomechanics and Motor Control Theory and Application Abingdon, Oxon: Routledge. 2013.
25. Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM. The level and tempo of children's physical activities: an observational study. Medicine and science in sports and exercise. 1995;27(7):1033-41.
26. Berman N, Bailey R, Barstow TJ, Cooper DM. Spectral and bout detection analysis of physical activity patterns in healthy, prepubertal boys and girls. Am J Hum Biol. 1998;10(3):289-97.

27. Hollung SJ, Vik T, Lydersen S, Bakken IJ, Andersen GL. Decreasing prevalence and severity of cerebral palsy in Norway among children born 1999 to 2010 concomitant with improvements in perinatal health. Eur J Paediatr Neurol. 2018;22(5):814-21.
28. Oskoui M, Coutinho F, Dykeman J, Jette N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Dev Med Child Neurol. 2013;55(6):509-19.
29. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, et al. A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl. 2007;109(suppl 109):8-14.
30. Andersen GL, Irgens LM, Haagaas I, Skranes JS, Meberg AE, Vik T. Cerebral palsy in Norway: prevalence, subtypes and severity. Eur J Paediatr Neurol. 2008;12(1):4-13.
31. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Gross motor function classification system for cerebral palsy. Dev Med Child Neurol. 1997;39(4):214-23.
32. Axivity Ltd. AX3 [cited 2020 Feb 28]. Available from: https://axivity.com/products/ax3.
33. Breiman L. Random forests. Machine learning. 2001;45(1):5-32.
34. Kuhnhausen J, Dirk J, Schmiedek F. Individual classification of elementary school children's physical activity: A time-efficient, group-based approach to reference measurements. Behavior research methods. 2017;49(2):685-97.
35. Mannini A, Rosenberger M, Haskell WL, Sabatini AM, Intille SS. Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle. Medicine and science in sports and exercise. 2017;49(4):801-12.
36. Trost SG, Wong WK, Pfeiffer KA, Zheng Y. Artificial neural networks to predict activity type and energy expenditure in youth. Medicine and science in sports and exercise. 2012;44(9):1801-9.
37. Farrahi V, Niemela M, Kangas M, Korpelainen R, Jamsa T. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. Gait & posture. 2019;68:285-99.
38. Ahmadi M, O'Neil M, Fragala-Pinkham M, Lennon N, Trost S. Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. Journal of neuroengineering and rehabilitation. 2018;15(1):105.
39. Hegde N, Zhang T, Uswatte G, Taub E, Barman J, McKay S, et al. The Pediatric SmartShoe: Wearable Sensor System for Ambulatory Monitoring of Physical Activity and Gait. IEEE Trans Neural Syst Rehabil Eng. 2018;26(2):477-86.

40. Hulteen RM, Smith JJ, Morgan PJ, Barnett LM, Hallal PC, Colyvas K, et al. Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis. Preventive medicine. 2017;95:14-25.
41. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica: Biochemia medica. 2012;22(3):276-82.
42. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. Medicine & Science in Sports & Exercise. 2005;37(11):S523-S30.
43. Carlon SL, Taylor NF, Dodd KJ, Shields N. Differences in habitual physical activity levels of young people with cerebral palsy and their typically developing peers: a systematic review. Disabil Rehabil. 2013;35(8):647-55.
44. Waltersson L, Rodby-Bousquet E. Physical Activity in Adolescents and Young Adults with Cerebral Palsy. Biomed Res Int. 2017;2017:8080473.

Appendix 1: Protocol

Protokoll

Skrive ned tidspunkt for alle oppgaver i registreringsskjema. Alle aktiviteter unntatt bakgrunnsmålinger skal filmes (ikke stå i veien for filmkamera!).

DEL 1 – Aktivitetsgjenkjenning	
Bakgrunnsmålinger	Montert utstyr: *Ingen* Utstyr i lokalet: *vekt, høydemåler, registreringskjema*

- Fylle ut bakgrunnsinformasjon i registreringsskjema
- Måle høyde og vekt

Aktivitetsgjenkjenning- strukturert del - Utføres innendørs	Montert utstyr: *Axivity på lår, L3, og tå. Polar HR monitor* Utstyr i lokalet: *GoPro på stativ, stoppeklokke, registreringsskjema, kjegler, erteposer, hoppetau, lav benk, god stol, madrass/seng.*

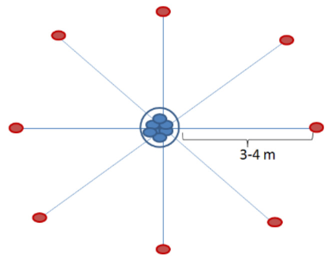
1. Hælslipp
   - Utfør 3 hælslipp eller knips på akselerometrene 3 gang
   - Sett markør i HR rate målingene (‘lap’ på polarklokka) ved første hælslipp
   - Skriv ned tidspunkt for første hælslipp
2. Finne hastighet for ulike øvelser:
   - Gå oppmerket distanse (ca 30 m) i normal, sakte og rask hastighet
   - Jogg og løp oppmerket distanse.
   - Ta tiden på hver oppgave.

Verbal kommunikasjon ”gå helt vanlig”, ”gå saktere enn du går til vanlig” ”gå fortere enn du går til vanlig”, ”jogg vanlig” ”løp vanlig”.

3. Agility
   - Barnet skal stå og høre på at du forklarer oppgaven
   - Løpe sikk-sakk mellom kjegler: sett opp banen som vist i figuren. Lengden på banen er 10 m.

Barnet skal forflytte seg så fort de klarer fra første til siste kjegle (se figur), stå i ro i 5 sekund og ta samme vei tilbake. Utfør 3 runder.

- Løp og plukk opp erteposer: Sett opp banen som vist i figuren. Radius på banen er 3 m for de minste barna og 4 m for større barn/ungdommer. Barnet skal bytte ut alle de røde erteposene med de blå erteposene som ligger i midten av banen. Det er bare lov til å ta med seg 1 ertepose om gangen. Barnet får beskjed om at de må gjøre det som rask som mulig, men de får ikke lov til å kaste posene. Start og mål er i midten av banen. Utfør 2 runder.



4. <u>Sitte på benk/gulv/stol.</u>
   - Sitte på lav benk: Gå bort til benken, stå der i ca 5 sekund før dere setter dere på benken. Gi barnet et blad eller bok. Se litt i bladet sammen og be så barnet om å ta med seg bladet bort til godstolen og sette seg å se litt i det. Total sittetid på benk ca 1 minutt
   - Sitte i godstol/sofa: La barnet sitte i godstolen i ca 1 minutt.
   - Sitte på gulv: For mindre barn – samle sammen erteposene i en haug. Sett deg selv ned på gulvet. Be barnet komme bort til deg og «rydde» erteposene opp i en pose/bøtte. For større barn - be dem ta med seg bladet og sette seg på gulvet å lese videre. Total sittetid på gulv ca 1 minutt
   - Måle bakgrunnsvariabler (høyde og vekt)
5. <u>Hoppe</u>
   - Hoppe på trampoline – 2-3 min
   - Hoppe med tau – 2-3 min
6. <u>Ligge</u>
   - Ligge på ryggen –ca 30 sekund
   - Ligge på magen –ca 30 sekund
   - Ligge på venstre og høyre side – ca 30 sekund på hver side
   - Ligge som om du skal sove (ta med teppe) – ca 1 minutt
   - Ligge hvordan de selv ønsker med blad/mobil/nettbrett 1 minutt.
   - Måle lengde på ben (hoftekam til gulv) og lengde fra akserelometer på lår (nederst på akserelometer) til gulv

Aktivitetsgjenkjenning- fri del -utendørs og innendørselmenter hvis mulig	Montert utstyr: *Axivity på lår, L3, og tå. Polar HR monitor, GoPro med brystsele* Utstyr i lokalet: *Erteposer, stoppeklokke, registreringsskjema*

Instruksjonen i denne delen skal være oppgaverettet (f.eks finn og bring tilbake). Det skal <u>ikke</u> spesifiseres hvilke type bevegelser de skal gjøre (f.eks gå og plukk opp). Presiser at de har god tid og at det ikke er noen konkurranse, men ikke fortell dem hvordan de skal utføre oppgaven (f.eks. gå eller løpe).

1. Skattejakt for å finne 5 puslespillbrikker ute i klatrestativet (inne i gymsal om det er dårlig vær). Gi dem 1 brikke før de starter. Finn brikkene og ta dem med opp til sofaen utenfor kontoret og legg ferdig puslespillet der. Når puslespillet er ferdig går de ned og henter en av oss for så å ta oss med opp og vise at de er ferdig.

2. Fotballkamp: 3 mot 3 i hall – 5 min.

3. <u>Hælslipp</u>
   - Utfør 3 hælslipp eller knips på akselerometrene 3 gang

- Sett markør i HR rate målingene ('lap' på polarklokka) ved første hælslipp
- Skriv ned tidspunkt for første hælslipp

DEL 2 - Energiforbruk	
Energiforbruk -Gjerne utendørs hvis mulig	Montert utstyr: *Axivity på lår, L3, og tå. Polar HR monitor, GoPro med brystsele, Metamax* Utstyr i lokalet: *Stoppeklokke, registreringsskjema, skolestol, mobil (gjerne barnets egen)/nettbrett*

Lengde på aktivitetene tilpasses barnas alder og fysiske form. Minimum lengde på aktivitetene er 3 minutt. Lengden på gangbanen/runden må måles og noteres ned. Antall tilbakelagte lengder/runder registreres som et mål på hastighet. Barna bestemmer selv om de ønsker pause og hvor lange pausene skal være. Pausene tilbringes på skolestolen og/eller stående. Klikk gjerne av maska i pausene hvis barnet ønsker dette. Sett markør i metamax og polarmålinger ved start og stopp av hver aktivitet. Under måling kan barnet gi signal på om det går bra eller dårlig ved å gi tommel opp eller tommel ned. Sjekk inn med hyppig med barnet (minste en gang i minuttet).

<u>Hælslipp</u>

- Utfør 3 hælslipp eller knips på akselerometrene 3 gang
- Sett markør i HR rate målingene ('lap' på polarklokka) ved første hælslipp
- Skriv ned tidspunkt for første hælslipp

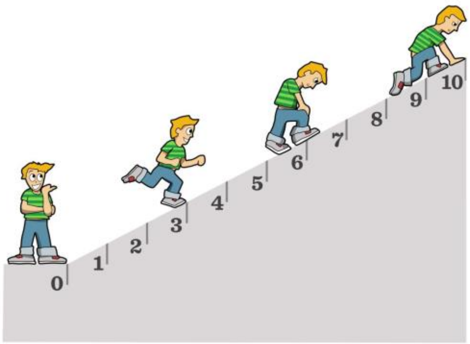
1. Gå helt rolig 3-5 minutt
2. Sitt i ro i 3 min, se på telefonen eller nettbrett
3. Stå og vent i 3 minutt, se på telefon eller nettbrett
4. Gå normalt 5 minutt – HUSK **OMNI** før og etter (NB: lik verbal kommunikasjon!).
5. Gå fort 5 minutter
6. Jogge 3-5 minutter
7. Løpe 3-5 minutter
8. Sykle 5 minutt

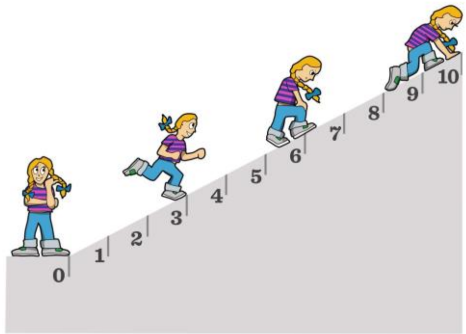
<u>Hælslipp</u>

- Utfør 3 hælslipp eller knips på akselerometrene 3 gang
- Sett markør i HR rate målingene ('lap' på polarklokka) ved første hælslipp
- Skriv ned tidspunkt for første hælslipp.

Prosedyre for bruk av OMNI Walk/ Run Rating of Perceived Exertion (OMNI- RPE)

Vis bilde av OMNI- RPE skalaen, les opp standardskriv om skalaen og vurdering: «*Nå kommer vi til å spørre deg om hvor sliten du føler deg mens du går. Vær vennlig og bruk numrene på tegningen for å forklare oss hvordan du kjenner deg etter gangtesten. Se på Emil/Hannah nederst i bakken. Hvis du kjenner deg som han/henne, er du ikke sliten i det hele tatt, da bør du peke på null-tallet. Så kan du se når Emil/Hannah står på toppen av bakken. Hvis du føler deg som han/henne, er du veldig, veldig sliten og du bør peke på tall nummer ti. Dersom du* *følerer at du er et sted imellom, så peker du på et tall mellom null og ti. Vi vil at du forteller oss hvordan hele kroppen kjennes, og husk at det verken finnes rette og eller gale svar. Bruk både bildet og ord for å velge passende tall»*





Appendix 2: Definition of activities

DEFINITION OF ACTIVITIES	
Sitting	When the person's buttocks is on the seat of the chair, bed or floor (children). Sitting can include some movement in the upper body and legs; this should not be tagged as a separate transition. Adjustment of sitting position is allowed, as long as it does not include change in posture and should not be tagged as sit-transition-sit.
Standing	Upright, feet supporting the person's body weight, with no feet movement, otherwise this could be shuffling/walking. Movement of upper body and arms is allowed until forward tilt and arm movement occurs below knee height. Then this should be inferred as bending. **For chest mounted camera**: If feet position is equal before and after upper body movement, standing can be inferred. Without being able to see the feet, if upper body and surroundings indicate no feet movement, standing can be inferred.
Walking	Locomotion towards a destination with 1 stride or more, (1 step with both feet, where 1 foot is placed at the other side of the other), walking could occur in all directions e.g. forward, backwards, sideways. Walking along a curved line is allowed. From standing walking starts when walking direction is set and **heel off** occurs on the first foot. From transition or shuffling walking starts when walking direction is set and **heel off** occurs on the latter foot.
Shuffling	Stepping in place, by non-cyclical movement of the feet, turning on the spot with feet movement not as part of walking bout. **For chest mounted camera:** Without being able to see the feet, if movement of the upper body and surroundings indicate feet movement, shuffling can be inferred.
Stair ascending/descending	**Start: Heel-off** of the foot that will land on the first step of the stairs. **End:** When the heel-strike of the last foot is placed on flat ground. If both feet rests at the same step with no feet movement, standing should be inferred.
Lying	The person lies down. Adjustment after lying down is allowed if it does not lead to a change between the prone, supine, right and left lying positions. Movement of arms and head is allowed. Movement of the feet is allowed as long as it does not lead to change in posture. **Prone:** On the stomach. **Supine:** On the back. **Right side:** On right shoulder. **Left side:** On left shoulder.
Cycling	Riding a bicycle. Can be either sitting or standing. **Sit:** Pedaling while the buttocks is placed at the seat. Cycling starts at first pedaling, or when the bike is moving while one/both feet are on the pedal(s). Cycling ends when the first foot is in contact with the ground. If one/both feet are placed on the pedal(s), the buttocks are placed at the seat, with no pedaling and the bike is standing still, this should be tagged as sitting. **Stand:** Standing with both feet on the pedals, while riding a bike. Stand cycling starts when the buttocks leave the seat, and ends when the buttocks are placed on the seat.

	Not pedaling: Sitting or standing without pedaling should be tagged separate as sitting/standing if lasting for more than 5 seconds.
Running	Locomotion towards a destination, with at least two steps where both feet leave the ground during each stride. **For chest mounted camera:** Running can be inferred when trunk moves forward is in a constant upward-downward motion with at least two steps. Running along a curved line is allowed.
Jumping	To move your body upward from the ground and often forward, backward, or sideways through the air by pushing with your legs. Jumping starts as soon as lowering of center of mass occurs and finishes when the person is in a stable and upright position.
Bending	While standing/sitting, bending towards something below knee-height is tagged as bending. Steps can occur during bending.
Picking	This refers to picking/placing/touching an object from below knee height. Picking occurs when the trunk is at its lowest point and the person has touched/placed/picked an object. When the subject starts to rise the trunk, picking finishes, and bending begins. Adjustment of position while picking is allowed.
Non-vigorous activity	All non-cyclic movements that are recognizable, but do not classify according to the definitions. Can occur in all directions. Can be crawling, rolling, cleaning the floor, falling etc.
Undefined	Until all the sensors are attached, or final adjustment made to position the video can be tagged as undefined. All postures/movements that do not classify according to definitions or can not be clearly identified should be tagged as undefined.
DEFINITIONS OF TRANSITIONS	
Stand/walk/sit to bending to picking	Bending begins as soon as forward/sideways trunk tilt occurs, or at first bending of the knee(s). Bending finishes when they have reached the lowest point of the movement and picking occurs. When the subject starts to rise up, picking finishes and bending begins. When the trunk is in an upright and stable position, bending finishes. Some steps can occur during bending. This should be tagged as "bending-picking-bending".
Picking to bending to stand/walk/sit	Starting with the trunk at its lowest point, bending begins when the subject starts to rise up, and finishes when the trunk is in an upright and stable position. Some steps can occur during bending.
Walking to posture	Walking ends when both feet are at rest, or at first evident forward tilt of upper body. Steps can occur during the transition from walking to posture
Upright to sitting	Can be from walking or standing, as soon as forward trunk tilt occurs, or a lowering of the center-of-mass or trunk, the transition has started. Some steps can occur during the transition for positioning. Transition ends when buttocks are in contact with the seat of the chair, bed or floor.
Sitting to upright	Transition starts when the person starts to raise their center of mass and ends when the trunk has reached its upright angle or if upright posture happens later from when the feet pass each other. Some steps and turning can occur during the transition.

Upright (stand/walking) to lying	When the trunk tilt begins, or a lowering of the center of mass or trunk, the transition has started. Transition finishes when the person is lying flat with the upper body/trunk in a stable position.
Lying to upright (stand/walking)	While lying, the transition begins with an upward movement of the trunk or leg movement that leads to a stable upright position or continuous walking. Transition finishes when the trunk is in a stable, upright position. Steps can occur during the transition.
Sit-lie / lie-sit	Between the static sitting and lying posture, the transition is the dynamic movement between them, including movement of the legs and trunk but not arms or head.
Standing to walking	As soon as heel-off occurs, walking has started.
Standing to shuffle	As soon as the feet move, or leg-movement that leads to feet movement, shuffling has started.
Shuffling/walk to standing	As soon as the feet stop moving, walking/shuffling has finished, and standing has started.
Shuffling to walking	As soon as walking direction is set and heel-off occurs, shuffling has ended, and walking starts.
Walking to shuffling	When walking is interrupted by stepping in place, non-cyclical, non-directional movement of the feet or turning on the spot, this should be tagged as shuffling.
Sit cycling to stand cycling / stand cycling to sit cycling	When the buttocks leave the seat, stand cycling can be inferred. When the buttocks are placed at the seat, sit cycling can be inferred.

NTNU
Norwegian University of
Science and Technology