Atle Melleby Kongsvold

Validation of the AX3 accelerometer for detection of common daily activties and postures

Master's thesis in Human Movement Science
Supervisor: Paul Jarle Mork
Trondheim, juni 2016

Norges teknisk-naturvitenskapelige universitet
Faculty of Medicine
Department of Neuroscience

NTNU
Norwegian University of
Science and Technology

Sammendrag

Introduksjon: Stillesittende atferd har blitt foreslått å være en risikofaktor for helseproblemer uavhengig av fysisk aktivitetsnivå, men den eksakte effekten ulike typer aktivitet som ligge, sitte, stå og gå, har på helsa, er uvisst. Å oppnå presise objektive målinger av ulike typer fysisk aktivitet og inaktivitet er derfor svært viktig for videre forskning på feltet. Målet i denne studien var å utvikle og validere et oppsett med to tri-aksiale akselerometre for å differensiere mellom dagligdagse aktiviteter og positurer. Aktivitetsklassifisererne ble utviklet ved bruk av maskinlæringsalgoritmer. Klassifisererne ble også sammenligner med den beste eksiterende aktivitetsklassifisereren, Acti4.

Metode: Tjueto voksne (9 menn, 13 kvinner) ble rekruttert til deltagelse i studien. To akselerometer ble festet på deltagerne, et på låret og et på øvre del av rygg. Protokollen for validering ble delt opp i to deler, en strukturert protokoll i laboratoriet for å emulere dagligdagse aktiviteter, og en semistrukturert ut-av-lab-protokoll. Deltagerne ble filmet med et videokamera under begge protokollene. Videoene ble senere annotert bilde-for-bilde og brukt som kriteria for validering. Akselerometer- og videodata ble synkronisert og to ulike aktivitetsklassifiserere ble utviklet, en lab-modell trent på den strukturerte delen og testet på den semistrukturerte delen (NTNU$^{\text{LAB-MODEL}}$), og en modell trent og testet på det komplette datasettet (NTNU$^{\text{ADUL}}$). Et rammeverk med definisjoner av aktiviteter, positurer og transisjoner ble også utviklet.

Resultat: IRR fra videoannotering var 0.96 ($p<0.0001$) mellom tre kodere. Vektet gjennomsnitt av sensitiviteten for de tre modellene var 91% for NTNU$^{\text{ADUL}}$, og 87% for NTNU$^{\text{LAB-MODEL}}$ og Acti4. Sensitiviteten var ≥92% for gå, løpe, stå, sitte, ligge og sykle i NTNU$^{\text{ADUL}}$, mens spesifisiteten var ≥97% og nøyaktigheten var ≥95%. NTNU$^{\text{LAB-MODEL}}$ hadde sensitivitet på ≥89% for løpe, gå, stå, sitte og ligge. Acti4 hadde en sensitivitet på ≥81% for de samme aktivitetene.

Konklusjon: Aktivitetklassifisererne utviklet i denne studien klarte å detektere og differensiere mellom dagligdagse aktiviteter og positurer med høy sensitivitet, spesifisitet og nøyaktighet.

Abstract

Introduction: Sedentary behavior has been suggested as an independent risk factor for ill-health, with detrimental effects independent of physical activity. However, the exact effect of different types of activity i.e. lying, sitting, standing and walking, on health is uncertain. To obtain precise objective measurements of different types of physical activity and sedentary behavior is therefore of great importance for further research in this field. This study aimed to develop and validate a setup with two tri-axial accelerometers to differentiate between common daily activities and postures. The activity classifiers were developed by use of machine learning algorithms. The classifiers were also compared with the existing benchmark activity classifier Acti4.

Methods: Twenty-two adults (9 males, 13 females) were recruited to the study. Two accelerometers were fixed to the participants, one on the thigh and one on the upper back. The protocol for validation was divided into two sessions, one structured in-lab session emulating common daily activities, and one semi-structured out-of-lab session. Participants were filmed with a video camera during both sessions. The videos were later annotated frame-by-frame and used as criterion for validation. Accelerometer data and video data were synchronized and two different activity classifiers were created, one lab model trained on the structured session ($\text{NTNU}^{\text{LAB-MODEL}}$), and one model trained and tested on the complete dataset ($\text{NTNU}^{\text{ADUL}}$). A framework with definitions of activities, postures and transitions were also developed.

Results: The IRR from video annotation were 0.96 ($p<0.0001$) between three raters. The overall weighted sensitivity of three models were 91% for $\text{NTNU}^{\text{ADUL}}$, and 87% for $\text{NTNU}^{\text{LAB-MODEL}}$ and Acti4. The sensitivity was ≥92% for walking, running, standing, sitting, lying down and cycling in $\text{NTNU}^{\text{ADUL}}$, while specificity was ≥97% and accuracy ≥95%. $\text{NTNU}^{\text{LAB-MODEL}}$ had a sensitivity of ≥89% for running, walking, standing, sitting and lying down. Acti4 had a sensitivity of ≥81% for the same activities.

Conclusion: The activity classifiers developed in this study were able to detect and differentiate between common daily activities and postures with high sensitivity, specificity and accuracy.

Acknowledgements

Many people have contributed in the conduction of this project. I would like to thank my main supervisor Paul Jarle Mork for the opportunity to take part in this project, his supervision during the whole period and his invaluable part in the writing process. My collaborator during this project, Hilde Bårdstu needs a huge thank you for the good cooperation we have had during the development and completion of the protocol and definitions, discussion partner and as partner in the annotation process. Astrid Ustad for cooperation with the definitions and annotation. Co-supervisor Espen A. Ihlen for good discussions through the whole period and feedback during the writing process. Without the MATLAB help from Alan Bourke, who sat hours and hours to help us, this project would not have come together. Astrid Tessem and Hans-Olav Hessen also need a thank you for the development and creation of the algorithms, and answering all my stupid questions about machine learning. Ellen Marie Bardal for the work to enable analysis in Acti4. At last I would like to thank my girlfriend Sunniva Austnes, and my parents May and Egil for always being there for me. Without all of you as support during this process, this master thesis would not have been completed.

Table of Contents

Introduction

The relation between regular physical activity and a reduced risk for several diseases is well established[1]. Accordingly, physical inactivity is associated with non-communicable diseases (NCDs) such as diabetes, obesity, cardiovascular disease (CVD) and several types of cancer[2-5]. Lee et al.[6] estimated that physical inactivity accounts for 9 % of global premature mortality, which translates to 5.3 million people each year.

The recommendations from the World Health Organisation (WHO)[7] regarding physical activity suggest that adults perform at least 30 min of moderate to vigorous intensity physical activity (MVPA) every day. Moderate to vigorous intensity is defined as 3.5 metabolic equivalents or above (>3.5 METs) corresponding to activities such as brisk walking and easy biking. It is also recommended that each bout of activity to last at least 10 min in order to have a positive effect on health. People that do not get enough MVPA according to the guidelines are often characterised as inactive[8]. The public health recommendations for physical activity are mainly derived from studies using self-reported physical activity by questionnaires and diaries as exposure measurement[9]. These tools put a low burden on the respondent and they are cost-effective and hence feasible in large population based studies. An alternative method for recording physical activity is by objective measurements which are not prone to recall bias.

The validity of the subjective measures are seen as acceptable for MVPA, since it often is structured and planned[10]; however, the validity for the lower end of the spectrum of physical activity and sedentary activities is low[11,12]. A review on the validity of the International Physical Activity Questionnaire Short-Form (IPAQ-SF) found that people overestimated physical activity with an average of 84% when compared to objective measures[13]. Correlation between self-reported sitting time per day derived from the international physical activity questionnaire (IPAQ) compared with objective measures (accelerometer) was low to moderate (r=0.14-0.51)[14,15].

The importance of being able to measure the range of the activity spectrum has lately been emphasised. Sedentary behaviour, such as long sitting periods with low energy expenditure (<1.5 METs), has emerged as a separate risk factor with detrimental effects on health independent of physical activity levels[16]. People who are sufficiently active according to guidelines, but apart from this living a sedentary lifestyle have been found to have elevated risk for diseases such as CVD[17,18].

With the technological developments the recent decades, new possibilities for objective measurements of physical activity has emerged. Pedometers, heart rate monitors, accelerometers, global positioning systems and smart phones are all examples of wearable devices that has been utilised to obtain measurements of physical activity[19]. Currently, there is no consensus regarding what should be the gold standard for measurements of physical activity monitoring and obviously, which device that is most appropriate to use depend on the research question. Physical activity is a complicated construct that can be described both by quantitative and qualitative measures and characteristics[20]. The technology driven development of measurement tools has mainly been focused on quantitative assessment, and use of wearable accelerometers to quantify physical activity has gained much popularity during the recent decade. Tri-axis accelerometers enables measurements in three planes, with a graded output signal (i.e. amplitude may be used to indicate intensity). Moreover, analysing the output in relation to gravity enable the identification the orientation of the accelerometer, and thereby the angle[21].

Several large epidemiological studies such as NHANES[22] and ALSPAC[23] has incorporated long-term objective measurements of physical activity by accelerometers. In these studies, three-axial accelerometers are worn by the participants in a belt around the waist. The level of physical activity is usually estimated by number of "counts" from the accelerometer signal. Counts are the result of summing number of times the acceleration signal bypass a given threshold, most of the presented as an aggregated number for 15-60sec epochs[24]. Counts are often used as estimates for energy expenditure (EE), and is categorized (based on number of counts for a given period), into low-, light-, moderate-, and vigorous physical activity. A higher number of counts per time unit would mean more intense physical activity. However, it has been questioned whether number of counts represent a good measure of the range of different types of physical activity[25]. Hamilton et al.[26] found that quiet standing and sitting is significantly different from each other physiologically, but both can be defined as sedentary behaviour when estimating EE based on counts (<1.5 Mets). Moreover, results from a prospective study[27] found that substituting one hour of sitting with one hour of standing, had positive effects on all-cause mortality. Based on findings like these, the focus has shifted to address different types of physical activity as walking, standing, sitting, lying etc. rather than EE.

Validation studies conducted with accelerometers to differentiate between different levels of EE, activities, and postures are often carried out in standardised laboratory settings[28-35], and have shown high accuracy/sensitivity (95.0-99.8% sensitivity/accuracy), but that does not necessarily reflect valid detection in real-life settings. When applied to less confined settings human activity recognition systems have shown a considerable decrease in accuracy and sensitivity (17-20%)[35-37].

The use of multiple accelerometers placed at different body locations to measure physical activity has been investigated by Cleland et al.[38]. They used from one and up to six tri-axial accelerometers on different parts of the body, and concluded that two accelerometers yielded small, but significantly better results than one in terms of accuracy in detecting common daily activities. Interestingly, they also found that adding a third accelerometer did not provide any further increase in accuracy. One accelerometer one the upper body and one on the lower body showed best results. This also supported earlier research[39] on the topic.

Skotte et al.[40] is one of the few studies that have implemented the use of two accelerometers. They developed a MATLAB (Math, Works, Inc., US) based software program called Acti4, to analyse accelerometer data recorded with Actigraph GT3X (Actigraph Corp., US). This activity classifier is able to synchronise and analyse data from two sensors simultaneously and distinguish between a range of static and dynamic activities. Its classification is rule based, i.e. the inclination of the thigh can decide if a person is standing or sitting, and it has been validated with high sensitivity and specificity (>95%) in standardised field trials. When validated in free-living with video recordings, the sensitivity was lower for many activities and was reported between 50-99%, while the specificity was >83%[41]. It is thought that the performance of analytic tools such as Acti4 can be improved based on advancements in machine learning software. In addition, Acti4 has only validated Actigraph, which is not available as open-source.

An alternative to Actigraph, is the AX3 (Axivity, York, UK) sensor, which is a tri-axial lightweight accelerometer. The Axivity system is based on open-source thereby giving the researcher the possibility to have full control of the signal processing (e.g., effect of re-sampling, filtering and so on). It can be combined with machine learning software, which have made it possible to analyse complex signals from several accelerometers simultaneously. Attal et al.[42] investigated the performance of different supervised and unsupervised machine learning algorithms to detect common daily activities with nine sensors. Supervised learning algorithms produced the best results and several algorithms had an overall accuracy of >95%.

The aim of this study is to validate a setup with two AX3 accelerometers for the detection of common daily activities such as walking, running, standing, sitting, lying and cycling, in adults. Two activity classifiers developed on the basis of machine learning will be evaluated. One classifier will be trained on data from a structured session mimicking daily activities, whereas the other also will be trained on data from a semi-structured session. Direct observation through video will be used as criterion for validation. A framework with definitions of activities, postures and related transitions will be presented, and will work as guidelines when analysing videos. In addition, a comparison of the new activity classifiers with the already existing Acti4, which is regarded as the benchmark activity classifier, will be conducted.

Methods

Participants

Twenty-two adult men and women participated in the study. All participants were recruited among staff at the university (NTNU) and the university hospital (St. Olavs Hospital). The characteristics of the study sample are presented in table 1. All participants received written information prior to participation (Appendix C) and all participants signed a written consent upon inclusion in the study. Anthropometric measurements were performed before sensors were attached. The regional ethical committee (REK) approved the study and the study was carried out according to the declaration of Helsinki.

Table 1. Characteristics of study sample. Values are mean±SD (range)

	Male (n=9)	Female (n=13)
Age, years	36.2± 6.4 (28-46)	42.0±5.6 (35-52)
Height, cm	188±9.0 (168-198)	169±6.5 (155-178)
Weight, kg	87.5±7.0 (74.0-102)	65.6± 8.2 (49.5-79.4)
Body mass index (kg/cm^2)	24.8±1.5 (22.6-26.8)	22.9±2.5 (19.2-29.8)

Instrumentation

The AX3 is a tri-axis wireless accelerometer, able to record acceleration within a dynamic range of ±2 g to ± 16 g (1 g = 9.81 m/s^2) and a precision of 13 bit. It has an internal memory of 512MB and can log data at 100 Hz for 14 days. The size of AX3 is 23x32.5x7.6mm, and the weight is 11g. The accelerometers were connected to a PC with a USB cable, and later initiated and set-up with the software Open Movement GUI (OMGUI, 1.0.0.29) provided by the manufacturer. The sensors were set to record for two hours at a sample rate of 200Hz.

To record activity, the two AX3 accelerometers were attached directly to the participants skin prior to the test session. One sensor was attached to the frontal part of the right thigh, 10 cm above the upper line of the patella. The second sensor was attached on the upper part of the back adjacent to the spinous process of T5. Sensors were capsuled with a finger condom and fixed to the body by tape (3M, Hair-set, double-sided adhesive tape and Fixomull, BSN medicalon). An adhesive waterproof bandage (Flexifix, Smith & Nephew) was wrapped around the accelerometers to ensure the position and fixation to the skin. The x-axis was orientated in the longitudinal plane, y-axis in the mediolateral plane and z-axis in the anteroposterior plane. The placements of the accelerometers are shown in figure 1.

The recorded data was downloaded from the accelerometer and the raw CWA-file was exported to a computer for further processing. The actual sample rate of the accelerometers deviated slightly from the sample rate predefined in the OMGUI software. Therefore, a MATLAB-based software was created to synchronise the sampling rate between accelerometers and the video annotation. Furthermore the signal was also resampled to 100 Hz in the MATLAB software.

Both sessions were filmed with a GoPro Hero 3+ (GoPro, INC., USA) video camera set to record at 30 frames per second (FPS) with a resolution of 720x360Pi. During the in-lab session the video camera was fixed on a tripod, while the camera was mounted on the chest with straps and pointing down towards the legs during the treadmill running, vigorous activity, and semi-structured out-of-lab session.

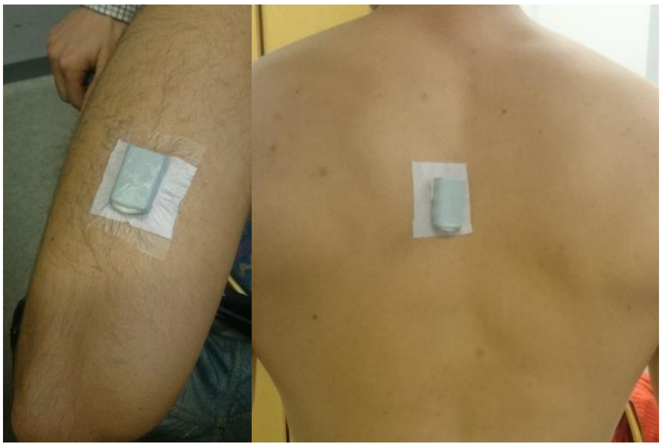

Figure 1. The accelerometers attached to the thigh and back.

Protocol

A flow chart of the methodological process is presented in Figure 2.

The participants performed two different sessions. At the start of the protocol, between the two session, and at the end of the out-of-lab session, three heel drops were performed. This was done to be able to recognise the shift between start and stop of the recording, the shift between the in-lab and out-of-lab session, and to synchronise the accelerometers and video data.

The first session was standardised, and performed in a laboratory designed to accommodate the activities in the protocol. Participants were instructed by a test leader to perform different activities. The order of the activities was fixed and equal for all participants, but the execution varied as participants were instructed to act normal. The activities included were walking, sitting, standing, lying (supine, prone, left, right), bending and picking, stair ascending, stair descending, vigorous activity with rapid changes of direction, running on a treadmill at 8km/h, walking on a treadmill at 2, 4 and 6km/h, 3^{o}, 6^{o} and 9^{o} incline at 4 km/h, and stationary biking. All activities were repeated 2-3 times, and each static posture were held for at least 8-10 sec during each repetition. The in-lab protocol lasted approximately 30 min.

The semi-structured session was performed subsequent the in-lab session. This part was conducted during free-living, mostly in a working environment. All participants received a list with activities, corresponding to the activities in the in-lab session except stationary biking, treadmill walking and vigorous activity with rapid changes in direction. Participants were free to choose the succession and pace of the activities to be performed, for the session to be as natural as possible. The length of this session varied from 30-60 min depending on time available. All participants completed both sessions, which resulted in approximately 25.2 hours of video. The complete standardised protocol is shown in the Appendix A and B for both the structured in-lab and semi-structured protocol.

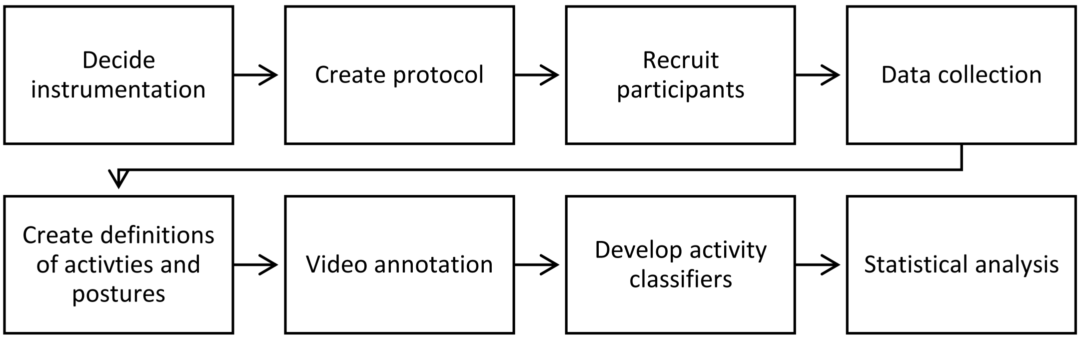

Figure 2. Flow chart of the methodological process

Activity definitions and annotation

Prior to analysing the video data recorded, activity definitions were set to provide guidelines to be used in the process of annotating the videos. This included running, walking, stair ascending, stair descending, shuffling, standing, sitting, lying down, bending and picking, cycling (sit and stand), other vigorous activity, unclassified and undefined. In addition, transitions between static and dynamic activities and between different static activities i.e. sitting-to-lying, were defined. The complete list of activities and transitions with descriptions is presented in Appendix D.

Analyses of the recorded video was used as criterion validity when testing the classifiers. Definitions of activities and related transitions gave guidelines for the video annotation. The videos were divided between two of the trained raters. All video annotation was computed after the video recordings was completed. Prior to annotating, all videos were downloaded to a PC and transferred to MPEG Streamclip 1.2 (Squared5, Italy) where the videos were converted from MP4 to AVI with the codec Apple Cinepak and resampled to 25 FPS. In addition, the soundtrack was removed and the frame size was reduced to 640x360Pi.

Video annotation was done frame-by-frame with the computer software program Anvil[43]. The accelerometer signal was not visible while annotating. When the annotation was finished, the frame-by-frame output and time labels were exported to a text document. The exported files were imported to MATLAB were the signal from the two accelerometers were synchronised. Thereafter, the synchronised files were transferred to The Waikato Environment for Knowledge Analysis (Weka) toolkit for analysis.

An inter-rater reliability (IRR) analysis was performed to assess the agreement between three different raters annotating the same videos. To establish the IRR, Fleiss'es (overall) kappa was determined. All videos from one subject were annotated by the three raters and used to calculate the IRR. In total approximately 70 min of video was used in this procedure. All defined activities and transitions were represented in the IRR-videos.

Activity classifiers

An open source machine learning software called Weka was used to produce the activity classifier. Feature generation and selection procedure was first conducted to find features of the acceleration signal that could separate between activities. In this procedure the Wrapper method and several filtering methods produced sets of features[44]. Different classification algorithms including decision tree, Random Forest, Support Vector Machine, artificial neural network and Naïve Bayes was then tested on the selected features. The features generated by the wrapper method combined with the Random Forest classifier provided the best results and was the method used to produce the two models in this study.

Two activity classifiers was produced, one where the whole data set was used in both training and testing, referred to as NTNU^{ADUL}. The other classifier was divided into an in-lab part, which was used in training, whereas the semi-structured session was used in testing, and is referred to as NTNU$^{LAB\text{-}MODEL}$. A 10-fold cross validation was used to calculate the confusion matrix for NTNU^{ADUL}, where the dataset was split into 10 equal subsets. Nine of the sets were used to train the classifier, and then tested on the tenth. This procedure was repeated 10 times. The resulting 10 classification models were averaged and used for the final model. A 1 sec window with 50 % overlap was used as window size for both models. Each segment was identified as one activity.

This study uses the preliminary result of the classifier, available in February 2016. The model will be further developed during 2016. The performance of the newly developed classifiers was compared to the already existing Acti4 classifier.

Acti4 was developed by Skotte and co-workers and is described in details elsewhere[40]. It is a rule based classifier that can detect the activities walking, running, shuffling, stair walking, standing, sitting, lying down and cycling. The activity shuffling used in the current study is equivalent to "move" used in Acti4. Minimum bout length for the different activities were set at 2 sec for running, walking, shuffling and standing, 5sec for stair walking, sitting and lying down, 15 sec for cycling. Threshold for differentiation of standing/shuffling were 0.1 g, walking/running 0.72 g, sitting/standing 45^{o} and 40^{o} for stair/cycling. These were the default settings. The raw CWA-files from the AX3 sensors were converted to CSV, and resampled to 30 Hz, to enable analyses in Acti4. The output from Acti4 were 1Hz, which was resampled to 25 Hz, to match video annotation. Acti4 has originally validated a setup with one sensor on the thigh and one on the hip, but has later also used a sensor fixed on the upper back[45], adjacent to T1-T2, which enabled a comparison with the setup used in this study.

Statistics

Sensitivity, specificity and accuracy were calculated from confusion matrices to evaluate the validity of NTNU$^{LAB\text{-}MODEL}$, NTNU^{ADUL} and Acti4, respectively. Confusion matrices were computed directly from Weka in the NTNU models, while MATLAB was used to compute the confusion matrix for Acti4. The confusion matrices makes it possible to identify how activities and postures get misclassified. Sensitivity is the probability of the classifier to classify the activity when the activity is present, also referred to as recall. Specificity is the probability of the classifier to reject the activity when it is not present. The calculation of accuracy is presented in equation 1. The overall accuracy equals the weighted average of the sensitivity. Only activities that Acti4 could classify were included in the calculations.

$$\frac{(\text{true positives} + \text{true negatives})}{(\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives})} \qquad (1)$$

IRR of the video annotation was calculated from Fleiss (overall) kappa. All statistical analyses were done using Microsoft Excel (2007) and MATLAB (2015b). Descriptive statistics were calculated and reported as mean ± SD of the mean for the variables. Statistical significance was set at $p<0.05$.

Results

Table 2 presents the number of epochs that was annotated for each activity by the three raters and time for each activity. The result from the IRR between the three raters was r=0.96 and p<0.0001. A r above 0.82 is seen as almost perfect agreement according to Landis and Koch [46]. A discrepancy of 20% in number of epochs was seen between the raters in standing, while shuffling had 17 %. Walking had a variation of six epochs and 28 sec between the raters. All other activities were rated with a discrepancy of three epochs or less.

Table 2. Classification of the different activities by the three raters. Values are number of epochs annotated. Numbers in parenthesis is the time.

Annotated activities	Rater 1	Rater 2	Rater 3
Walking	77 (838)	83 (860)	78 (832)
Running	4 (77)	4 (75)	4 (77)
Shuffling	146 (217)	122 (212)	140 (233)
Stairs up	18 (112)	18 (113)	18 (114)
Stairs down	10 (47)	10 (48)	10 (48)
Standing	177 (629)	142 (616)	159 (640)
Sitting	21 (1552)	21 (1554)	19 (1545)
Lying down	23 (297)	23 (310)	23 (295)
Transition	57 (151)	57 (128)	56 (152)
Bending/picking	66 (62)	66 (57)	66 (60)
Cycling (sit)	6 (58)	6 (58)	6 (58)
Cycling (stand)	3 (28)	3 (28)	3 (27)
Other vigorous activity	4 (46)	6 (51)	3 (44)
Other unclassified activities	13 (21)	15 (25)	13 (28)
Total number of epochs (sec)	625 (4135)	576 (4133)	598 (4153)

Sensitivity, specificity and accuracy for the two NTNU models and Acti4 are presented in table 3. One subject had missing data when processed in Acti4, and is only included in the NTNU-models. The overall weighted sensitivity for the different classifiers was 91% ($NTNU^{ADUL}$) and 87% ($NTNU^{LAB\text{-}MODEL}$ and Acti4). The sensitivity was high in $NTNU^{ADUL}$ for running, walking, standing, sitting and lying down ($\geq$92%). In $NTNU^{LAB\text{-}MODEL}$ the sensitivity for the same activities were $\geq$89%, while Acti4 had a sensitivity of $\geq$89% for these activities, except standing (81%). Shuffling was identified with low sensitivity in all three models: 41% ($NTNU^{ADUL}$), 15% ($NTNU^{LAB\text{-}MODEL}$) and 51% (Acti4). The sensitivity of stair walking was higher for $NTNU^{ADUL}$ (82%) compared to $NTNU^{LAB\text{-}MODEL}$ and Acti4 (61%).

Cycling had a sensitivity of 95% in NTNU$^{\text{ADUL}}$, while it was low in Acti4 (37%). The specificity and accuracy were generally high in all activities in all three models. NTNU$^{\text{ADUL}}$ had specificity of ≥97% and accuracy ≥95%, NTNU$^{\text{LAB-MODEL}}$ ≥94% and ≥94%, and Acti4 ≥93% and ≥91%, respectively. The largest difference in specificity and accuracy were in shuffling with 6% discrepancy in specificity and 4% discrepancy in accuracy.

Table 4 presents the confusion matrix for NTNU$^{\text{LAB-MODEL}}$, while table 5 presents the confusion matrix for NTNU$^{\text{ADUL}}$. Walking, shuffling and standing was primarily misclassified as the other two. Shuffling was misclassified as walking 26% and as standing 54% of the time in NTNULAB-MODEL. The same misclassification appeared in NTNU$^{\text{ADUL}}$, but here the proportion of misclassification as standing was 29%. The misclassification of walking as standing was reduced from 5.9% in NTNU$^{\text{LAB-MODEL}}$ to 2.2% in NTNU$^{\text{ADUL}}$. Standing had an increase in misclassification as shuffling from 1.6% in NTNU$^{\text{LAB-MODEL}}$ to 4.1% in NTNU$^{\text{ADUL}}$.

Stair walking was primarily misclassified as walking in both models, but the proportion of walking upstairs was reduced from 51% in NTNU$^{\text{LAB-MODEL}}$ to 16% in NTNU$^{\text{ADUL}}$. The latter was reflected in the sensitivity (40% in NTNU$^{\text{LAB-MODEL}}$ and 82% in NTNU$^{\text{ADUL}}$). Lying down and transition was identified as the other, or sitting, when classified wrongly in NTNU$^{\text{LAB-MODEL}}$. The misclassification of lying down as sitting was reduced from 2.6% to 0.1% in NTNU$^{\text{ADUL}}$. Other vigorous activity was misclassified as running (33%), standing (11%) and walking (5.0%) in NTNU$^{\text{LAB-MODEL}}$, and 4.0%, 3.2% and 7.0% in NTNU$^{\text{ADUL}}$.

Table 6 presents the confusion matrix for Acti4. In addition to showing how the software classified the activities, the table shows how it classified activities it cannot detect (transition, bending/picking, other vigorous activities and unclassified). In Acti4, walking was misclassified as shuffling (5.8%), standing (1.6%), stair walking (1.6%), and sitting (1.1%). Cycling was misclassified as lying down (21%) and stair walking (35%).

Table 3. Sensitivity, specificity and accuracy for the three models. Empty cells in NTNU$^{LAB\text{-}MODEL}$ means no data available. Acti4 does not classify the activities; transition, bending/picking, other vigorous activities and other unclassified activities. Thus, these cells are empty.

	Sensitivity (%)			Specificity (%)			Accuracy (%)		
Activity	NTNU^{ADUL}	NTNU$^{LAB\text{-}MODEL}$	Acti4	NTNU^{ADUL}	NTNU$^{LAB\text{-}MODEL}$	Acti4	NTNU^{ADUL}	NTNU$^{LAB\text{-}MODEL}$	Acti4
Walking	94	89	89	97	95	95	96	94	94
Running	96	89	89	100	100	100	100	99	100
Shuffling	41	15	51	98	99	95	95	94	91
Stair walking*	82	47	61	100	100	98	99	98	97
Standing	92	91	81	97	94	97	96	94	94
Sitting	99	98	98	99	97	98	99	98	98
Lying down	97	93	99	100	99	97	99	99	97
Transition	67	66	-	99	99	-	98	98	-
Bending/picking	76	38	-	100	100	-	99	99	-
Cycling**	95	-	37	100		100	100	-	98
Other vigorous activities	79	48	-	100	100	-	100	100	-
Unclassified	63	41	-	100	100	-	100	100	-

*Stairs up and down merged together in NTNU^{ADUL} and NTNU$^{LAB\text{-}MODEL}$.

**Sit cycling and stand cycling merged together in NTNU^{ADUL}.

Table 4. Confusion matrix from NTNU$^{LAB\text{-}MODEL}$. Rows represent the activities annotated from the video recordings, while columns represents the activities classified by the model. Numbers in bold are the sensitivity and the number of sec the sensitivity represents. All other values are normalised and presented as %.

	Activities classified by NTNU$^{LAB\text{-}MODEL}$											
Annotated Activities	Walking	Running	Shuffling	Stairs up	Stairs down	Standing	Sitting	Lying down	Transition	Bending/ picking	Other vigorous activity	Unclassified
Walking	**5350 89**	0.3	3.0	0.3	0.4	5.9	0.2	0	1.0	0.1	0.2	0
Running	2.3	**520 89**	0.2	0.3	0.9	0.3	0	0	0	0	6.9	0
Shuffling	26	0.1	**484 15**	0.1	0.3	54	0.9	0	3.7	0.5	0.4	0
Stairs up	51	0	0.6	**668 40**	1.6	1.3	0	0	2.3	0	3.6	0
Stairs down	23	1.1	0.8	0.4	**311 67**	1.4	0	0	0	0	7.1	0
Standing	2.9	0	1.6	0.1	0.1	**10533 91**	3.3	0	0.6	0.4	0	0
Sitting	0.1	0	0	0	0	0.1	**26468 98**	0.8	0.9	0	0	0
Lying down	0	0	0	0	0	0	2.6	**1205 93**	4.7	0	0	0
Transition	7.6	0	2.1	0.2	0	2.3	9.4	11	**450 66**	1.4	0.4	0
Bending/picking	5.5	0	2.7	0	0	6.4	16	2.4	29	**118 38**	0.5	0
Other vigorous activity	5.0	33	1.3	0	2.5	11	0	0	0	0	**19 48**	0
Unclassified	6.8	0	0.7	0	0.2	34	11	0.9	3.5	1.2	0	**88 41**

Table 5. Confusion matrix from NTNU^{ADU_L}. Rows represent the activities annotated from the video recordings while columns represent the activities classified by NTNU^{ADU_L}. Numbers in bold are the sensitivity and the number of sec the sensitivity represents. All other values are normalised and presented as %.

Annotated activities	Activities classified by NTNU^{ADU_L}													
	Walking	Running	Shuffling	Stairs up	Stairs down	Standing	Sitting	Lying down	Transition	Bending/ picking	Cycling sit	Cycling Stand	Other vigorous activity	Unclassified
Walking	**17316 94**	0.3	2.8	0.4	0.2	2.2	0	0	0.4	0.1	0	0	0	0
Running	2.2	**1862 96**	0	0.2	0.5	0.1	0	0	0	0	0	0	1.0	0
Shuffling	26	0.1	**1949 41**	0.6	0.6	29	0.3	0	1.6	0.4	0	0	0.2	0
Stairs up	16	0.4	0.4	**1734 82**	0.3	0.3	0	0	0.2	0	0	0	0.5	0
Stairs down	17	0.5	0.3	1.4	**863 79**	0.3	0	0	0	0	0	0	1.1	0
Standing	2.4	0	4.1	0	0.1	**18436 92**	0.2	0	0.2	0.7	0	0	0.1	0.1
Sitting	0.1	0	0	0	0	0.1	**28954 99**	0	0.5	0	0.1	0	0	0
Lying down	0	0	0	0	0	0	0.1	**5584 96**	3.2	0	0	0	0	0
Transition	4.2	0	3.7	0.4	0	3.6	8.2	9.7	**1956 67**	2.1	0.3	0.3	0.1	0
Bending/picking	1.4	0	1.0	0	0	14	2.2	0	4.6	**1231 77**	0.1	0.1	0.2	0
Cycling sit	0	0	0	0	0	0.1	6.7	0	0.8	0	**887 90**	1.9	0	0
Cycling stand	0	0	0	0	0	0	0.5	0	0.3	0	3.8	**675 95**	0	0
Other vigorous activity	7.0	4.0	0.6	2.4	2.5	3.2	0	0	0.9	0	0	0.2	**587 79**	0
Unclassified	5.1	0	2.0	0	0.3	25	2.3	0	1.1	0.7	0	0.1	0.3	**256 63**

Table 6. Confusion matrix from Acti4. Rows represent the activities annotated from the video recordings while columns represent the activities classified by the model. Numbers in bold are the sensitivity and the numbers of sec the sensitivity represents. All other values are normalised and presented as %.

	Activities classified by Acti4							
Annotated activities	Walking	Running	Shuffling	Stair walking	Standing	Sitting	Lying down	Cycling
Walking	**16196 89**	0.6	5.8	1.6	1.6	1.1	0.1	0
Running	5.5	**1716 89**	5.1	0	0.1	0	0	0
Shuffling	21	0.2	**2377 51**	3.7	23	1.7	0.3	0
Stair walking	18	0	20	**1925 61**	0.7	0	0	0
Standing	5.7	0.1	11	0.6	**15646 81**	1.0	0.2	0
Sitting	0.1	0	0.2	0.2	0.7	**27283 98**	0.6	0.3
Lying down	0	0	0	0	0.4	0.6	**5613 99**	0
Transition	2.7	0	8.2	0.2	12	21	56	0
Bending/picking	0.8	0	68	5.6	21	4.1	0.5	0
Cycling	0	0	1.3	35	0.6	5.6	21	**616 37**
Other vigorous activity	14	0	30	52	4.4	0	0	0
Unclassified	61	2.4	10	0.6	20	6.2	0.2	0

Discussion

The aim of this study was to validate a setup with two AX3-sensors for the detection of common daily activities. The two activity classifiers were developed by means of machine learning algorithms. The main finding was that the activity classifier is usable to discriminate between common daily activities with high sensitivity, specificity and accuracy. The discussion section is divided into a study specific discussion, which discusses how the models correctly classified and misclassified activities and postures, and a general discussion that assess important methodological aspects to consider when conducting research on objective measures of physical activity. In addition, limitations of the current study is discussed.

Study specific discussion

Running, walking, standing, sitting, lying down and cycling was identified with a high sensitivity ≥92%, specificity ≥97% and accuracy ≥96% in NTNU$^{\text{ADUL}}$, while the overall weighted sensitivity was 91% for all activities included in the analyses. These results are comparable with what Stemland et al.[41] found when Acti4 was used to validate a similar setup with Actigraph. The NTNU$^{\text{LAB-MODEL}}$ had an overall weighted sensitivity of 87%, which was 4% lower than NTNU$^{\text{ADUL}}$. Since NTNU$^{\text{ADUL}}$ was trained and tested on the same dataset, while NTNU$^{\text{LAB-MODEL}}$ was trained and tested on two different sessions, it was expected that NTNU$^{\text{ADUL}}$ would perform better. A larger difference between NTNU$^{\text{ADUL}}$ and NTNU$^{\text{LAB-MODEL}}$ than the observed 4% was expected, based on what other studies have found (17-20%)[35,47]. This can be an indication that the machine learning algorithm used in this study found a signature in the features used to detect the different activities in lab, that is generalizable to out-of-lab. It is also a possibility that the structured protocol was less standardised than what has been used in other studies, and more similar to free-living. Future studies could include a session in "true" free-living with other participants, to test all the models on, and to conduct an external validation of the models developed in this study.

Running, walking, standing, sitting and lying down had 1-7% lower sensitivity score in NTNU$^{\text{LAB-MODEL}}$. The activities bending and picking, other vigorous activity and unclassified had larger discrepancies between the two models, which can be because of the large diversities in how these activities can be conducted. Since specificity was very high for all activities, the sensitivity separates the performance of the models better. In other words, all three models performed well in identifying periods when a certain activity was not present. Both other vigorous activity and unclassified are categories consisting of many different activities that cannot be defined as any of the other activities. The majority of other vigorous activity was conducted in the structured session, and since participants were not asked to conduct any other vigorous activity in the semi-structured session, this resulted in few out-of-lab samples to test NTNU$^{\text{LAB-MODEL}}$ on. These categories probably need a large sample in order to be identified correctly by an activity classifier. Whether it was the large number of possible variations these activities can be conducted in, or if it was that sample number of the activity were too low, is uncertain. Further studies are needed to investigate if individual thresholds exist and how many samples that is needed in the training data set for the activity classifier to converge to a numerical stable estimate.

NTNU$^{\text{ADUL}}$ had a somewhat lower proportion of misclassification than NTNU$^{\text{LAB-MODEL}}$ i.e. lying down was misclassified as sitting (2.6%) and transition (4.7%) in NTNU$^{\text{LAB-MODEL}}$, while this proportion was reduced in NTNU$^{\text{ADUL}}$ to 0.1% and 3.2%, respectively. NTNU$^{\text{LAB-MODEL}}$ confused 51% of stairs up with walking, while this confusion was reduced to 16% in NTNU$^{\text{ADUL}}$. This can be an indication that activities need a large sample size for it to be correctly trained and thereby classified by the models. Transition is one activity that had little difference in terms of sensitivity, specificity and accuracy. Investigating the confusion matrices reveals that transition was misclassified as many of the other activities. Since transition is a combination of many different activities such as siting-to-lying, standing-to-sitting, walking-to-sitting, etc. it is logical that the models had problems identifying the transition correctly in every event. Further studies should perhaps split the category transitions into several sub-categories that could be to be identified more correctly.

NTNU$^{\text{ADUL}}$ had higher sensitivity, specificity and accuracy than the benchmark activity classifier Acti4 (91% vs. 87% overall weighted sensitivity), but since it was an external validation of Acti4, and NTNU$^{\text{ADUL}}$ was trained and tested on the same dataset, NTNU$^{\text{LAB-MODEL}}$ is a more justifiable comparison. Cycling was only performed in lab and was therefore only available in NTNU$^{\text{ADUL}}$ and Acti4. The sensitivity of cycling was 95% in NTNU$^{\text{ADUL}}$, and only 37% in Acti4, which misclassified cycling as stair walking (35%) and lying down (21%). It was unanticipated that Acti4 misclassified cycling as lying down, which are very unlike in the movement pattern, and has been identified with high sensitivity previously[40]. The reason could be that cycling only was done inside on a stationary bike, since the acceleration signal is different outside, due to movements of the bike. There is also a possibility that pre-processing of the acceleration signal and video annotation to enable analyses in Acti4, influenced the results. To process the acceleration and video annotations in Acti4, several procedures in MATLAB was done. The signal was first resampled from 100Hz, down to 30Hz and analysed in Acti4. The output file from Acti4 was 1Hz, which were resampled up to 25Hz, to match the signal with the video annotation. This processing may introduce information loss due to down-sampling and can therefore have influenced the final result. Future studies could include a session on a real bike outside, and include a simplified procedure of the pre-processing i.e. down-sample the video annotation instead of up-sample the acceleration signal.

This was an external validation of Acti4, which previously has been used to validate a similar set-up with Actigraph, as used in the current study. Since NTNU$^{\text{LAB-MODEL}}$ used the exactly same sensors, the same participants in both training and testing, the external validity of the testing can be limited, and it was therefore expected that NTNU$^{\text{LAB-MODEL}}$ would perform better than Acti4. However, this was not reflected in the results as both NTNU$^{\text{LAB-MODEL}}$ and Acti4 had 87% overall weighted sensitivity. Several activities such as Running, walking and sitting were classified with same sensitivity, while standing had 91% sensitivity in NTNU$^{\text{LAB-MODEL}}$, and Acti4 had 81%. Stair walking had higher sensitivity in Acti4 (61%), than NTNU$^{\text{LAB-MODEL}}$ (47%), and was mostly misclassified as walking and shuffling by both Acti4 and NTNU$^{\text{LAB-MODEL}}$. Specificity and accuracy differed 3% or less between Acti4 and NTNU$^{\text{LAB-MODEL}}$, except for shuffling, which had a specificity of 95% with Acti4 and 99% with NTNU$^{\text{LAB-MODEL}}$. Shuffling were originally named “move” in Acti4, and might have been defined different, since many of the activities Acti4 could not detect were misclassified as shuffling. For example 68% of all bending and picking, and 30% of other vigorous activity (see table 6) were classified as shuffling, which can be an indication that “move” is an activity comprised of several different activities. Future validation studies should try to compare activity classes with similar definitions even though this might be difficult at present, because of the lack of an international consensus on definition of activity classes.

Shuffling had low sensitivity (15-51%) and was misclassified as standing and walking in all three models. There was some misclassification between standing and walking as well, but the proportion was much lower. In NTNU$^{\text{ADUL}}$ shuffling was misclassified as standing 29% and walking 26%. Results from the IRR showed an agreement of 0.96 ($p<0.0001$), which is close to perfect agreement, but there was a 24 epoch discrepancy between the three raters in shuffling and 35 in standing, while walking had six. It seems that there was some overlap especially between shuffling and standing, which can come from the definitions of the activities used in this study. Walking was defined as one stride (one step with each foot) and shuffling was everything less than stride to the smallest foot movement, while standing was classified when there was no movement of the feet. The activity classifier might have missed many of the small movements that was defined as shuffling, and identified the period as standing. On the contrary, when participants moved for almost two steps i.e. one and a half the algorithm probably identified the period as walking. A solution to this challenge can be to post-process the output, for example to define shuffling by a time parameter. All walking periods shorter than 3 sec can be redefined as shuffling, and a merging with standing is also feasible from a health perspective standpoint. From a pragmatic standpoint, the health difference of standing still, and standing while moving the feet is probably minimal.

In summary NTNU$^{\text{ADUL}}$ had high overall weighted sensitivity (91%), while NTNU$^{\text{LAB-MODEL}}$ and Acti4 had 87%. All three models had high specificity and accuracy. The observed difference in weighted overall sensitivity of 4% between NTNU$^{\text{ADUL}}$ NTNU$^{\text{LAB-MODEL}}$ was lower than expected. The proportion of misclassification was somewhat lower in NTNU$^{\text{ADUL}}$ than NTNU$^{\text{LAB-MODEL}}$. It was unanticipated that Acti4 would perform similar to NTNU$^{\text{LAB-MODEL}}$, since it was an external validation of Acti4. Shuffling had low sensitivity in all models (15-51%), which can be because of the definitions used in the video annotation, since the results from the IRR showed some confusion in standing, shuffling and walking between the three raters.

General discussion

Currently, there does not exist a consensus on how to define activities and postures in objective measurements of physical activity. Ainsworth et al.[48] developed a comprehensive compendium of physical activity, with definitions of 821 activities by their rate of energy expenditure, to enhance comparability of results across studies using subjective measures. Using energy expenditure to define activities limits the transferability to studies taking the new approach and try to differentiate between different postures and activities. Consequently, studies develop own definitions when assessing physical activity, by activity type. The result is often technical definitions adapted to the ongoing study, i.e. features of the current technology that enables the researchers to distinguish between activities and postures. Defining postures by angles of the thigh and back, as Skotte et al.[40] did, only enables comparisons with studies that have the technology to measures this available, and is not transferable to other settings, i.e. when using direct observation. Other studies[32,49] did not provide any information about definitions of activities and postures, only a figure depicting how the acceleration signal reacted during different activities, which makes it difficult to compare findings across studies. On the background of this, it is clear that a consensus on definitions of activities and postures that is universal and independent of technology, is needed.

Prior to annotating the videos in this study, definitions with common daily activities were set. The goal with these definitions was that they should be universal and not dependent of technology. It is proposed to set definitions based on a hierarchical structure, instead of the flat structure used by Ainsworth et al.[50] who defined activities such as “walking the dog”, “walking to a neighbour”, and “walking slowly/strolling” as three separate activities, and not as the same activity “walking” in different activity contexts. With the approach of Ainsworth et al. it is not possible to define all activities specific to the activity context because the number of contexts which an activity like walking can appear, is very high. A hierarchical structure on the other hand, which was used in this study, enables researchers to account for all types and levels of activity. Activities can be divided into static and dynamic, and further sub-divided into activities to increase the detail level and correspond to the aim and population in the study. An example of a hierarchical structure is presented in figure 3.

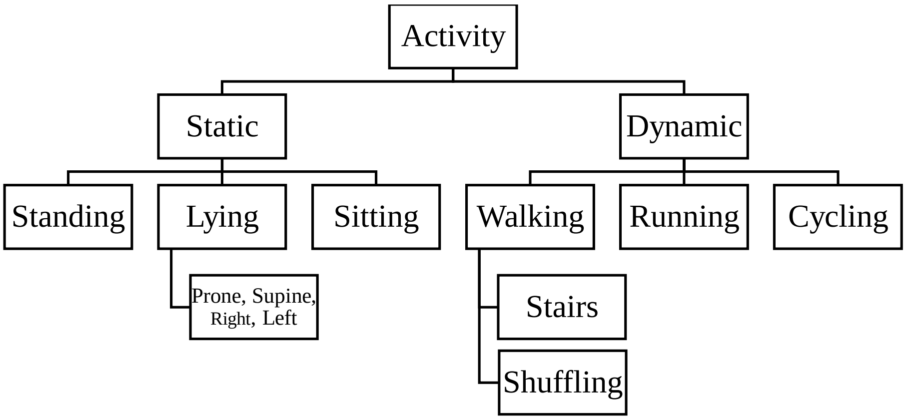

Figure 3. An activity tree, an example of a hierarchical structure.

The study population is important when choosing how detailed the recognition should be, since activity patterns differs between sub-groups such as children, adults, elderly and patient groups[51,52]. When assessing activity patterns of the elderly, it might be sufficient to assess walking, standing, sitting and lying down, while children have a significantly larger diversity in activities they conduct, and will require a more detailed activity tree with several sub-divisions to account for the daily activities children take part in. In this study, adults were the study population, and activities and postures were therefore divided into categories that were thought to comprise most daily activity that adults conduct, thereby also representing potentially important activities in a health perspective. All activity was accounted for and several activities in addition to the most common were included (i.e. bending and picking, other vigorous activity and unclassified), since it is important to don't have uncertainty about identification of activities in a validation study. However, when applied in a population study, the focus will probably be on the large activity categories, and in addition have the possibility to investigate sub-groups in more detail.

In a population based study, the requirements for accuracy is not as high as if the goal is to provide individuals with detailed information about their activity pattern. Nevertheless, the results of this study indicates that the NTNU^{ADUL} classifier is useful for providing individual feedback and is substantially more valid than self-reported measurements of activity. In a review by Prince et al.[53] the correlation between self-reported physical activity and objective measurements varied between -0.71 to 0.96, which can a problem when correcting for differences observed between objective and self-reported measures, because the reported error is not constant. However, self-reported measurements can provide important information about the context the activity is performed in, which can contribute to more detailed information about activity patterns. According to Rowe[54] physical activity and sedentary behaviour is a complex construct comprised of frequency, intensity, time, mode, context, volume and energy expenditure. The setup used in this study can assess most of these dimensions to some extent, however it cannot assess the context the activity is conducted in. This is of importance because the context describes the surroundings, which can help researchers understand why people are active. A combination of self-reported measures and objective measures is therefore suggested.

A main strength of this study is that the activity classifier and sensor all are based on open-source. Likewise, the definitions used to annotate the videos are available to everyone. To the authors knowledge, direct observation of activity has not previously been analysed frame-by-frame, which has enabled detailed classification of many different activities and postures. Based on this, the possibilities for future development of activity classifiers with the use of this dataset is large.

Study limitations

There was no confidence interval available from the results, but since the dataset was large with 25 hours of video, there is a possibility of statistical significance differences between the models. Especially the large activity classes with variations in sensitivity, as standing (91% in NTNU$^{LAB-MODEL}$ vs. 81% in Acti4), is nearby to assume would be significantly different from each other. However, the difference of 4% between NTNU^{ADUL} and the two other models in this study, probably have little practical value, at least when the applied to population based studies, where small differences disappear, because of the large amount of data.

This protocol did not include a session with "true" free-living. The semi-structured session was designed to emulate free-living, but since this session was conducted during working hours in an office area and only lasted 30-60 min, there was a lack in diversity of the activities conducted by the participants. The structured session can have influenced how participants conducted the semi-structured protocol, since they received a list with different activities corresponding to activities conducted during the structured session. A solution could have been to provide participants with tasks with a goal, e.g., get a cup of coffee, instead of requests to lie down, sit in a chair etc. A larger diversity in activity patterns, equal to free-living would probably have been observed.

This study did not include heart rate measurements. Intensity is therefore only possible to describe by cadence, or the counts from the raw acceleration output. To include a heart rate sensor and synchronise it with the accelerometers data would provide a more precise measure of intensity. An absolute intensity of 6 km/h might be perceived high for one person, but low for another, because of large differences in physical capacity. The structured protocol contained a short session on a treadmill were participants were walking at different speeds and inclinations. This enables the possibility to post-process the acceleration signal, identify differences in gait between the different speeds and inclines, and in addition incorporate heart rate estimation.

This study did not assess time spent in passive transportation, which others have had problems to detect. Bastian et al.[37] misclassified a large part of time spent sitting in a car or bus, which led to large underestimations of sitting time. On the other hand, a study by Kerr et al.[55] managed to detect time spent in vehicle with a sensitivity of 87%. Assessing this is something to consider in future validation studies in free-living. Cycling was not assessed on a real bike outdoor, therefore, it is unknown whether NTNU^{ADUL} can measure cycling on a bike outside.

Conclusion

In conclusion, a setup with two AX3 sensors placed on the thigh and back, combined with machine learning algorithms can detect and differentiate between common daily activities and postures with high sensitivity, specificity and accuracy. NTNU^{ADUL} had higher sensitivity for all activities than NTNU$^{\text{LAB-MODEL}}$, and hence a lower proportion of misclassification. Acti4 had a somewhat lower overall weighted sensitivity than NTNU$^{\text{ADUL}}$ (4%), and equal overall weighted sensitivity as NTNU$^{\text{LAB-MODEL}}$, which was unexpected since it was an external validation of Acti4. All models had high specificity and accuracy. Shuffling had low sensitivity in all models (15-51%), which can be due to the definitions used in the video annotation.

NTNU$^{\text{ADUL}}$ was not assessed in “true” free-living, passive transportation was not assessed, confidence intervals were not available, neither was heart rate measurements, and it is recommended that future studies address this.

References

1. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 2006;174:801-9.

2. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS, Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. The New England journal of medicine 1991;325:147-52.

3. Keith SW, Redden DT, Katzmarzyk PT, et al. Putative contributors to the secular increase in obesity: exploring the roads less traveled. International journal of obesity (2005) 2006;30:1585-94.

4. Morris JN, Clayton DG, Everitt MG, Semmence AM, Burgess EH. Exercise in leisure time: coronary attack and death rates. British Heart Journal 1990;63:325-34.

5. Friedenreich CM. Physical activity and cancer prevention from observational to intervention research. Cancer Epidemiology Biomarkers & Prevention 2001;10:287-301.

6. 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.

7. Global Recommendations on Physical Activity for Health. World Health Organization, 2010. at http://www.ncbi.nlm.nih.gov/books/NBK305060/.)

8. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010;38.

9. Warburton DER, Charlesworth S, Ivey A, Nettlefold L, Bredin SSD. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. The International Journal of Behavioral Nutrition and Physical Activity 2010;7:39-.

10. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? International journal of obesity and related metabolic disorders : Journal of the International Association for the Study of Obesity 1998;22 Suppl 2:S30-8.

11. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the Global Physical Activity Questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. BMC Public Health 2014;14:1-11.

12. Lagersted-Olsen J, Korshoj M, Skotte J, Carneiro IG, Sogaard K, Holtermann A. Comparison of objectively measured and self-reported time spent sitting. International journal of sports medicine 2014;35:534-40.
13. Lee PH, Macfarlane DJ, Lam T, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. International Journal of Behavioral Nutrition and Physical Activity 2011;8:115.
14. Rosenberg DE, Bull FC, Marshall AL, Sallis JF, Bauman AE. Assessment of sedentary behavior with the International Physical Activity Questionnaire. Journal of physical activity & health 2008;5 Suppl 1:S30-44.
15. Booth ML, Ainsworth BE, Pratt M, et al. International physical activity questionnaire: 12-country reliability and validity. Medicine and science in sports and exercise 2003;195:3508-1381.
16. Yates T, Wilmot EG, Khunti K, Biddle S, Gorely T, Davies MJ. Stand up for your health: Is it time to rethink the physical activity paradigm? Diabetes Research and Clinical Practice 2011;93:292-4.
17. Warren TY, Barry V, Hooker SP, Sui X, Church TS, Blair SN. Sedentary behaviors increase risk of cardiovascular disease mortality in men. Medicine and science in sports and exercise 2010;42:879-85.
18. Matthews CE, George SM, Moore SC, et al. Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. Am J Clin Nutr 2012;95:437-45.
19. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57:387-95.
20. Butte NF, Ekelund U, Westerterp KR. Assessing physical activity using wearable monitors: measures of physical activity. Medicine and science in sports and exercise 2012;44:S5-12.
21. Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. Medicine and science in sports and exercise 2005;37:S490.
22. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Medicine and science in sports and exercise 2008;40:181.
23. Riddoch CJ, Mattocks C, Deere K, et al. Objective measurement of levels and patterns of physical activity. Archives of disease in childhood 2007;92:963-9.

24. Santos-Lozano A, Marín PJ, Torres-Luque G, Ruiz JR, Lucía A, Garatachea N. Technical variability of the GT3X accelerometer. Medical Engineering & Physics 2012;34:787-90.

25. Crouter SE, Churilla JR, Bassett DR. Estimating energy expenditure using accelerometers. European journal of applied physiology 2006;98:601-12.

26. Hamilton MT, Hamilton DG, Zderic TW. Role of Low Energy Expenditure and Sitting in Obesity, Metabolic Syndrome, Type 2 Diabetes, and Cardiovascular Disease. Diabetes 2007;56:2655-67.

27. Stamatakis E, Rogers K, Ding D, et al. All-cause mortality effects of replacing sedentary time with physical activity and sleeping using an isotemporal substitution model: a prospective study of 201,129 mid-aged and older adults. International Journal of Behavioral Nutrition and Physical Activity 2015;12:1-10.

28. Wang J, Chen R, Sun X, She MFH, Wu Y. Recognizing Human Daily Activities From Accelerometer Signal. Procedia Engineering 2011;15:1780-6.

29. Ravi N, Dandekar N, Mysore P, Littman ML. Activity recognition from accelerometer data. AAAI; 2005. p. 1541-6.

30. Casale P, Pujol O, Radeva P. Human activity recognition from accelerometer data using a wearable device. Pattern Recognition and Image Analysis: Springer; 2011:289-96.

31. Mathie M, Celler BG, Lovell NH, Coster A. Classification of basic daily movements using a triaxial accelerometer. Medical and Biological Engineering and Computing 2004;42:679-87.

32. Khan AM, Lee Y-K, Lee SY, Kim T-S. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. Information Technology in Biomedicine, IEEE Transactions on 2010;14:1166-72.

33. Curone D, Bertolotti GM, Cristiani A, Secco EL, Magenes G. A Real-Time and Self-Calibrating Algorithm Based on Triaxial Accelerometer Signals for the Detection of Human Posture and Activity. Ieee Transactions on Information Technology in Biomedicine 2010;14:1098-105.

34. Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. Information Technology in Biomedicine, IEEE Transactions on 2006;10:156-67.

35. Gyllensten IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. IEEE transactions on bio-medical engineering 2011;58:2656-63.

36. van Hees VT, Golubic R, Ekelund U, Brage S. Impact of study design on development and evaluation of an activity-type classifier. Journal of Applied Physiology 2013;114:1042-51.
37. Bastian T, Maire A, Dugas J, et al. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. Journal of Applied Physiology 2015;118:716-22.
38. Cleland I, Kikhia B, Nugent C, et al. Optimal Placement of Accelerometers for the Detection of Everyday Activities. Sensors 2013;13:9183.
39. Atallah L, Lo B, King R, Yang GZ. Sensor Positioning for Activity Recognition Using Wearable Accelerometers. IEEE Transactions on Biomedical Circuits and Systems 2011;5:320-9.
40. Skotte J, Korshoj M, Kristiansen J, Hanisch C, Holtermann A. Detection of Physical Activity Types Using Triaxial Accelerometers. Journal of physical activity & health 2014;11:76-84.
41. Stemland I, Ingebrigtsen J, Christiansen CS, et al. Validity of the Acti4 method for detection of physical activity types in free-living settings: comparison with video analysis. Ergonomics 2015;58:953-65.
42. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical Human Activity Recognition Using Wearable Sensors. Sensors (Basel) 2015;15:31314-38.
43. Kipp M. Graduate College for Cognitive Sciences University of the Saarland, Germany. 2001.
44. Guyon I, Elisseeff A. An introduction to feature extraction. Feature extraction: Springer; 2006:1-25.
45. Korshoj M, Skotte JH, Christiansen CS, et al. Validity of the Acti4 software using ActiGraph GT3X+accelerometer for recording of arm and upper body inclination in simulated work tasks. Ergonomics 2014;57:247-53.
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
47. Ermes M, Parkka J, Mantyjarvi J, Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. Information Technology in Biomedicine, IEEE Transactions on 2008;12:20-6.
48. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of physical activities: classification of energy costs of human physical activities. Medicine and science in sports and exercise 1993;25:71-80.

49. Bonomi AG, Goris AHC, Yin B, Westerterp KR. Detection of Type, Duration, and Intensity of Physical Activity Using an Accelerometer. Medicine and science in sports and exercise 2009;41:1770-7.

50. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and science in sports and exercise 2000;32:S498-S504.

51. Hoos M, Kuipers H, Gerver W-J, Westerterp K. Physical activity pattern of children assessed by triaxial accelerometry. European journal of clinical nutrition 2004;58:1425-8.

52. Motl RW, Snook EM, Agiovlasitis S. Does an accelerometer accurately measure steps taken under controlled conditions in adults with mild multiple sclerosis? Disability and health journal 2011;4:52-7.

53. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5.

54. Rowe D. Back to the future? Algorithms and equipment vs. simplicity and common sense. International Journal of Human Movement Science 2011;5.

55. Kerr J, Patterson RE, Ellis K, et al. Objective Assessment of Physical Activity: Classifiers for Public Health. Medicine and science in sports and exercise 2016;48:951-7.

Appendix A

The in-lab protocol:

	Activity	Repetitions
1	Stand – heel drop– stand	3
2	Stand – sit – stand	3
3	Stand – sit at table - stand	3
4	Stand - lie down on the back - turn to right/left/stomach - stand	3
5	Stand – bend with straight legs - pick object from floor forward/left/right - stand	3
6	Stand - lie - sit - lie - stand	3
7	Stand - bend with bent legs – pick object from floor forward/left/right - stand	3
8	Sit still - sit cycling - stand cycling - sit cycling - sit still	3
9	Stand - heel drop - stand	3
10	Stand - walk at preferred pace - stand	2
11	Stand - Agility drill (forward/backwards/left/right) - stand	3
12	Stand - climb stairs (right foot first) - stand - descend stair (right foot first)	2
13	Stand - climb stairs (left foot first) - stand - descend stair (left foot first)	2
14	Stand - walk (slow speed) flat - stand	1
15	Stand - walk (normal speed) flat - walk (normal speed) 3 - walk (normal speed) 6 - walk (normal speed) 9 - walk (normal speed) 6 - walk (normal speed) 3 - walk (normal speed) flat - Stand	1
16	Stand - walk (fast speed) flat – stand	1
17	Stand – walk flat – run flat - walk flat – stand	3
18	Stand – heel drop – stand	3

Appendix B

Semi-standardisert protokoll: I hverdagen - Voksne

I løpet av den neste timen ønsker vi at du gjennomfører aktivitetene som er listet opp under. Når en aktivitet er gjennomført setter du et kryss under "gjennomført". Etter at du har fullført hele listen, går du tilbake til lab for å levere utstyret.

NB: ikke ta av akselerometer eller kamera før du er tilbake i lab!!

Subjekt ID: _______________

Aktiviteter	**Gjennomført**
Sitte - helst i to ulike stoltyper	
Stå stille	
Stå - shuffle (Beveg beina på stedet)	
Stå - løft en gjenstand fra bakken med bøyde bein	
Gå flatt	
Gå trapp opp og ned	
Ligge på rygg, høyre side og venstre side	
Løpe (varighet over 10sek)	
Tilbake i lab: Heel drop: Stå på tå - slipp hælene i bakken så hardt du kan. Gjennomføres tre ganger.	

Appendix C

Forespørsel om deltakelse i forskningsprosjektet

"Kvalitetssikring av objektive målinger av fysisk aktivitet"

Bakgrunn og hensikt

Dette er et spørsmål til deg om å delta i en forskningsstudie det målsettingen er å utvikle og teste en ny metode for måling av daglig fysisk aktivitet blant voksne. Ansvarlig for prosjektet er Institutt for Samfunnsmedisin ved Norges teknisk-naturvitenskapelige universitet (NTNU).

Hva innebærer studien?

Studien inkluderer to testsekvenser. Den første sekvensen tar ca. 30 min og vil foregå i et laboratorium ved NTNU på Øya. Først vil vi måle høyde og vekt og notere alder og kjønn. Deretter vil vi feste to aktivitetsmålere på kroppen din med tape – den ene festes bak på ryggen og den andre foran på låret rett over kneet. Målerne vil være skjult under klærne. Målerne er svært små (23x32.5x7.6 mm, vekt 11 g) og du vil i liten grad merke at sensorene er festet på huden (se bilde). Etter at målerne er festet på kroppen vil du bli bedt om å gjennomføre ulike aktiviteter i en bestemt rekkefølge (ligge, sitte, stå, gå, løpe på tredemølle og sykle på ergometersykkel). Gjennomføring av aktivitetene vil ikke være fysisk anstrengende. Etter at den første delen er ferdig ønsker vi at du fortsetter å gå med aktivitets-målerne i ca. 1 time der du gjennomfører dine vanlige arbeidsrutiner. I tillegg til aktivitets-målerne vil et lite bærbart kamera bli festet på brystet ditt. Dette kameraet vil kun filme beina dine slik at vi i ettertid kan gjenkjenne aktiviteten du har gjennomført (sitte, stå, gå osv.).

Mulige fordeler og ulemper

Ved å være med i studien vil du gi et viktig bidrag for å utvikle en ny metode for måling av fysisk aktivitet. Denne kunnskapen vil senere benyttes i ulike forskningsprosjekter der målsettingen er 1) å forstå hva som påvirker fysisk aktivitetsnivå i hverdagen, og 2) hvordan fysisk aktivitet over tid påvirker helsen. I studien skal du bære aktivitetsmålere festet til kroppen i ca. 1,5 time samt et kamera foran på brystet i ca. 1 time. Du vil i liten grad merke at du bærer målerne på kroppen og hverken målerne eller kameraet vil være til hinder for å gjennomføre dine vanlige arbeidsrutiner. Alle som deltar i studien vil være med i trekningen av en iPad mini (verdi ca. kr 2500,-).

Hva skjer med prøvene og informasjonen om deg?

Informasjonen som registreres om deg av aktivitetsmålerne og videokamera skal kun brukes slik som beskrevet i hensikten med studien. Alle opplysningene vil bli behandlet uten navn og fødselsnummer eller opplysninger som kan knyttes til deg personlig. En kode knytter deg til dine opplysninger gjennom en navneliste. Det er kun autorisert personell knyttet til prosjektet som har adgang til navnelisten og som kan finne tilbake til deg. Det vil ikke være mulig å identifisere deg i resultatene av studien når disse publiseres.

Frivillig deltakelse

Det er frivillig å delta i studien. Du kan når som helst og uten å oppgi noen grunn trekke ditt samtykke til å delta i studien. Dette vil ikke få konsekvenser for din videre behandling. Dersom du ønsker å delta, undertegner du samtykkeerklæringen på siste side. Om du nå sier ja til å delta, kan du senere trekke tilbake ditt samtykke uten at det påvirker din øvrige behandling. Dersom du senere ønsker å trekke deg eller har spørsmål til studien, kan du kontakte prosjektleder Paul Jarle Mork.

Appendix D

Definitions of activities

Activity	Description
Sitting	When the person's buttocks is on the seat of the chair, bed or floor. 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.
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 one stride or more, (one step with both feet, where one foot is placed at the other side of the other). Walking could occur in all directions. Walking along a curved line is allowed.
Shuffling	Stepping in place by non-cyclical and non-directional movement of the feet. Includes 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 non-directional 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 down	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 positons. 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.
Sit cycling	Pedalling while the buttocks is placed at the seat. Cycling starts on first pedalling and finishes when pedalling ends. **For outdoor bicycling:** Cycling starts at first pedalling, or when both feet have left the ground. Cycling ends when the first foot is in contact with the ground. **Not pedalling:** Sitting without pedalling should be tagged separate as sitting.
Stand cycling	Pedalling while standing. Cycling starts on first pedalling and finishes when pedalling ends. Standing without pedalling should be tagged separate as standing.
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.

Appendix D cont.

Bending	While standing/sitting, bending towards an object placed below knee-height is 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 person starts to rise it’s trunk, picking finishes, and bending begins.
Other vigorous activities	All non-cyclic rapid leg movements that do not classify as running. This includes sport like activities such as rapid change in direction and jumping. Can occur in all directions.
Unclassified	All non-cyclic movements that do not classify according to the definitions. Can occur in all directions.
Undefined	Until all the sensors are attached, or final adjustment made to position the video can be tagged as undefined. All postures/movements that can not be clearly identified should be tagged as undefined.

Appendix D cont.

Definitions of transitions

Transitions	Description
Bending to picking from standing/walking/sitting	As soon as forward/sideways trunk tilt occurs, bending has started. Bending finishes when the person has reached the lowest point of the movement and picking occurs. When the person starts to rise up, picking finishes and bending begins. When the trunk is in an upright and stable position, bending finishes. This should be tagged as "bending-picking-bending". 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 trunk, the transition has started. 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's buttocks leave the chair and ends when the trunk has reached its upright position. Steps and turning can occur during the transition from sitting to upright.
Standing/walking/sitting to lying	When the trunk flexion begins, or a lowering of the centre of mass, the transition has started. Transition finishes when the person is lying flat with the trunk in a stable position.
Lying to standing/walking/sitting	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. The trunk angle should be in a steady posture for the transition to finish. Steps can occur during the transition.
Standing to walking	As soon as heel-off occurs, walking has started.
Standing to shuffling	As soon as one foot moves, shuffling has started.
Shuffling/walking 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 is placed at the seat, sit cycling can be inferred.