Onderdeel	Datum	Naam
Introduction V2	7/12/2020	Ali
Recommendation PA V1	15/12/2020	Ali
MET regression models	15/12/2020 (Middag anders avond)	Colin
Activity classification	8/12/2020	Adnan
Results - Activity classification	8/12/2020	Adnan
V2 Activity classification	15/12/2020	Adnan
V2 Results - Activity classification	15/12/2020	Adnan
V1 Train/validation/test split	15/12/2020	Adnan
Results - MET regression models	17/12/2020 (Middag anders avond)	Colin

Theory v2	16/12/2020	Mark
V2 Train/validation/test split	18/12/2020	Mark
Discussion	Na alle results	Matthew
Study design v3	14/12/2020	Matthew
Introduction V3	14/12/2020	Matthew
Conclusion	18/12/2020 of 18/12/2020	Colin-Adnan

Title

Colin Werkhoven[1] , Adnan Akbaş[1] ,Ali Safari[1], Matthew Turkenburg[1] and Mark Boon [1]
[1] The Hague University of Applied Sciences, Johanna Westerdijkplein 75, 2521 EN Den Haag
Whois email adress?

Abstract

Introduction V1 - Ali

- State why the problem you address is important
- State what is lacking in the current knowledge
- State the objectives of your study or the research question

/// = later nog bronvermelding toevoegen

Adults must spend a minimum of 150 minutes of moderately intense physical activity per week to adhere to the international physical activity recommendation ///. In the past, CBS (Statistics Netherlands) has simply asked if people adhere to these norms. This is when the problem occurs, people are not good at keeping track of how many minutes they have spent moving or doing sports. A precise number of minutes people spend doing activities and the intensity of those activities are missing in the data. This inspired CBS to research how accelerometers could be useful to fill in this gap in the data.

In preliminary research, CBS compared different accelerometers to find the best individual sensor for activity classification and intensity measurement. The accelerometers taken into consideration were UKK, activPal, Hexoskin and IMU. CBS concluded that the activPAL accelerometer is the best individual sensor for this goal ///. The conclusion also states: "ActivPAL registers the approximate intensity of cycling well, but is not accurate enough to distinguish between light and heavy cycling. ActivePAL also overestimates the intensity of low intensity activities. " ///..

The research goal is to try and answer with accuracy if a person adheres to the international physical activity recommendation by answering the following research questions:

1. How can Machine Learning be used to predict the intensity of activities performed in a lab situation by a person, who is being monitored with Vyntus One and wearing ActivPal accelerometer?
    a. What measurement does ActivPal use for intensity and why?

Commented [T(1]: In this study PA was measured during daily life in a population of healthy adults using a single accelerometer. Simultaneously, TEE was assessed using 63 Improving assessment of energy expenditure the gold standard technique of doubly-labeled water. The aim was to investigate whether the identification of activity type combined with a simple methodology to define activity type intensity could improve the estimation of TEE, AEE, and PAL as compared to daily activity counts.

Commented [T(2R1]: Bovenstaand is een voorbeeld hoe je een introductie kan afsluiten. Door een soort samenvatting te geven welke research questions zijn onderzocht.

Commented [T(3]: Persoonlijk vind ik dit een beetje vreemde manier. (zie dit ook niet op deze manier terugkomen in de papers die ik heb gelezen). Misschien kunnen we de research questions meer verwerken in de introductie?

   b. Is it possible to extract this intensity measurement values from just Vyntus One data, if so, how?
2. How can Machine Learning be used to predict the intensity of activities performed by a person wearing only the ActivPal accelerometer, based on the data gathered from Vyntus One and ActivPal accelerometer in the lab situation?
   a. What machine learning model can best be used to measure the intensity for each activity?
3. How can Machine Learning be used to determine whether people did their 150 minutes of moderate activity in ActivPal accelerometer data of an entire week?
   a. How can Machine Learning be used to recognize the activities, performed in the lab situation, in the ActivPal accelerometer data?

Commented [T(4]: Navragen hoe we de research questions kunnen verwerken in de introductie.

Introduction V2 - Ali

· State why the problem you address is important

· State what is lacking in the current knowledge

· State the objectives of your study or the research question

It is general knowledge that performing moderately intense physical activities are good for your health. However, it is difficult for the general person to know if they have performed enough moderately intense physical activities to get the benefits. This could be because people are busy, they do not track it well or do not maintain a fixed schedule. According to the international physical activity recommendation, adults must spend a minimum of 150 minutes of moderately intense physical activity per week ///. CBS (Statistics Netherlands) has an interest in knowing whether people adhere to this norm. In the past, CBS has simply asked if people adhere to these norms in the form of SQUASH (Short Questionnaire to Assess Health-enhancing physical activity) ///. The problem is that people are not good at keeping track of how many minutes they have spent performing physical activities. A precise number of minutes people spend doing activities and the intensity of those activities are missing in the data. This inspired CBS to research how accelerometers could be useful to fill in this gap.

Commented [T(5]: Wat is moderately?

Commented [T(6R5]: En waarom kies je voor moderately

Commented [A(7]: Onderbouw "general knowdedge" anders is het een aaname . Volgens andrioli is het niet erg maar beter voor zekere dan onzekere

Commented [T(8]: Ik zou hier een bron neergooien, alleen maar om een bron te hebben

Commented [A(9]: Mischien een verwijzing naar een bron?

Commented [T(10]: Bron mist

Commented [T(11]: Is dit problem duidelijk toegelicht en onderbouwd?

In preliminary research ///, CBS compared different accelerometers to find the best individual sensor for activity classification and intensity measurement. The accelerometers taken into consideration were UKK, ActivPal, Hexoskin and IMU. The study tested individual accelerometers and a combination of these on a subject group the size of 40 people. The experiments were executed in a lab setting. Each person was instructed to perform certain activities while wearing the accelerometers and being monitored by Vyntus One ///. CBS concluded that the ActivPAL accelerometer is the best individual sensor for this goal ///. The conclusion also states: “ActivPal registers the approximate intensity of

Commented [W(12]: zal hier gelijk al de bron refereren

Commented [A(13]: Annemiek heeft een email gestuurd met een bestand over study design. Ik zou dit controleren of het ook daar staat

cycling well but is not accurate enough to distinguish between light and heavy cycling. ActivePal also overestimates the intensity of low intensity activities. " ///.

This paper is the result of a follow-up study on the ActivPal accelerometer. The data that resulted from the preliminary research by CBS is eligible for use in this study. Firstly, the theory chapter describes relevant terminology, formulas and literature this research is based on. Secondly, the methods are described. Specifically, it describes the data more thoroughly and the approach taken to create the Machine Learning models. Thirdly, the results state the objective results from our research. Finally, the conclusion aims to describe how Data Science models and the activPAL accelerometer can be used to answer, with reasonable accuracy, whether people adhere to the international recommendation of 150 minutes of moderate physical activity.

Introduction V3 - Matthew

It is general knowledge that performing moderately intense physical activities are good for your health. However, it is difficult for the general person to know if they have performed enough moderately intense physical activities to get the benefits. This could be because people are busy, they do not track it well or do not maintain a fixed schedule. According to the international physical activity recommendation, adults must spend a minimum of 150 minutes of moderately intense physical activity per week ///. CBS (Statistics Netherlands) has an interest in knowing whether people adhere to this norm. In the past, CBS has simply asked if people adhere to these norms in the form of SQUASH (Short Questionnaire to Assess Health-enhancing physical activity) ///. The problem is that people are not good at keeping track of how many minutes they have spent performing physical activities. A precise number of minutes people spend doing activities and the intensity of those activities are missing in the data. This inspired CBS to research how accelerometers could be useful to fill in this gap.

In preliminary research ///, CBS compared different accelerometers to find the best individual sensor for activity classification and intensity measurement. The accelerometers taken into consideration were UKK, activPAL, Hexoskin and IMU. The study tested individual accelerometers and a combination of these on a subject group the size of 40 people. The experiments were executed in a lab setting. Each person was instructed to perform certain activities while wearing the accelerometers and being monitored by Vyntus One ///. CBS concluded that the ActivPAL accelerometer is the best individual sensor for this goal ///. The conclusion also states: "ActivPal registers the approximate intensity of cycling well but is not accurate enough to distinguish between light and heavy cycling. activPAL also overestimates the intensity of low intensity activities. " ///.

The purpose of this study was threefold. The first target was to study whether machine learning could be used to predict the intensity of activities performed in a lab situation by a person, who is being monitored with Vyntus One and wearing ActivPal accelerometer. The second target was if the same

Commented [T(14]: In this study PA was measured during daily life in a population of healthy adults using a single accelerometer. Simultaneously, TEE was assessed using 63 Improving assessment of energy expenditure the gold standard technique of doubly-labeled water. The aim was to investigate whether the identification of activity type combined with a simple methodology to define activity type intensity could improve the estimation of TEE, AEE, and PAL as compared to daily activity counts.

Commented [T(15R14]: Bovenstaand is een voorbeeld hoe je een introductie kan afsluiten. Door een soort samenvatting te geven welke research questions zijn onderzocht.

Commented [T(16]: Persoonlijk vind ik dit een beetje vreemde manier. (zie dit ook niet op deze manier terugkomen in de papers die ik heb gelezen). Misschien kunnen we de research questions meer verwerken in de introductie?

Commented [A(17]: ActivPal

Commented [A(18]: Ik weet niet of in de paper een soort samenvatting van de paper in zetten

Commented [T(19]: Moet de rode draait worden meegenomen in de introductie (op het laatste)? In de papers die ik las, zat dit niet.

Commented [T(20]: Wat is moderately?

Commented [T(21R20]: En waarom kies je voor moderately

Commented [T(22]: Ik zou hier een bron neergooien, alleen maar om een bron te hebben

Commented [A(23]: Annemiek heeft een email gestuurd met een bestand over study design. Ik zou dit controleren of het ook daar staat

Commented [W(24]: Als dit de eerste keer is dat Vyntus One wordt genoemd, moet het wat duidelijker worden wat het inhoud

Commented [T(25]: Persoonlijk vind ik dit een beetje vreemde manier. (zie dit ook niet op deze manier terugkomen in de papers die ik heb gelezen). Misschien kunnen we de research questions meer verwerken in de introductie?

results could be made without the use of the Vyntus One. This means that the measurement of PAL is only done by using a tri-axial accelometer, ActivPal. The last target of this study was to determine if Machine Learning be used to determine whether people did their 150 minutes of moderate activity by only monitoring with the ActivPal accelometer.

Introduction V4 – Ali

According to the WHO (World Health Organization), physical inactivity is the fourth leading risk factor for global mortality and physical inactivity levels are rising ///. It is already well known that physical inactivity is bad for one's health, so why is it such a big factor? It might be because it is difficult for people to keep track of how much physical activity they performed and how intense it was. The WHO is recommending internationally for 18–64-year-olds to spend at least 150 minutes of moderate intensity physical activity per week ///. CBS (Statistics Netherlands) has an interest in knowing whether people adhere to this norm. In the past, CBS has simply asked if people adhere to these norms in the form of a SQUASH (Short Questionnaire to Assess Health-enhancing physical activity) ///. The problem with this being that people do not have a good idea of how many minutes they have spent performing physical activities. A precise number of minutes people spend performing activities and the intensity of those activities is missing in the data. This inspired CBS to research how accelerometers could be useful to fill in this gap.

In preliminary research ///, CBS compared different accelerometers to find the best individual sensor for activity classification and intensity measurement. The accelerometers taken into consideration were UKK, activPAL, Hexoskin and IMU. The study tested individual accelerometers and a combination of these on a subject group the size of 40 people. The experiments were executed in a lab, each person was instructed to perform certain activities while wearing multiple accelerometers and being monitored by a Vyntus® CPX system ///. In short, the Vyntus® CPX system is a device that analyses a person's breathing. After the lab session, people were instructed to return to their normal lives while wearing the accelerometers for one week. CBS concluded that the activPAL accelerometer is the best individual sensor ///. The conclusion also states: "activPal registers the approximate intensity of cycling well but is not accurate enough to distinguish between light and heavy cycling. activPAL also overestimates the intensity of low intensity activities. " ///.

The purpose of this study was threefold. The first target was to study whether machine learning could be used to predict the intensity of activities performed in the lab, based on the Vyntus® CPX data. The second target was to study whether the intensity of activities performed throughout the week could be predicted with just the activPAL accelerometer data, based on the Vyntus® CPX data. The last target of this study was to determine if Machine Learning be used to determine whether people, who are wearing the activPAL accelerometer, adhere to the international recommendation of 150 minutes of moderate intense physical activity for a week.

Theory – V1 (Mark)

Commented [T(26]: Wat wil je bij subject field precies terugzien.

- Opdracht uitleggen
- Terminologie beschrijven
- Literatuur noemen
    - Andere onderzoeken die vergelijkbare dingen doen
    - Uitleggen waarom niet methode overnemen of dat je het wel doet
- Hier ben je een hele pagina voor kwijt
- Vragen aan groepsgenoten waar ze informatie vandaan hebben gehaald

Colin:

- Na introductie een hoofdstuk over de opdracht > wat is MET waarde, wat zijn verschillende activiteiten, relevante literatuur, opdrachtomschrijving, begrippenlijst

How can machine learning be used to recognize activities in accelerometer data. And how can machine learning be used to estimate the MET values (Metabolic Equivalent of Task) in accelerometer data. Underneath we will explain our findings with references to the literature we used in our research of this subject field.

Due the strong correlation between accelerometry and human energy expenditure, accelerometer data can be used to estimate the intensity of the activities performed, using only an accelerometer and machine learning (Eidi Sasaki et al., 2016). The MET values (our measurement for human energy expenditure) can be estimated using the linear relationship between the sum of magnitude of acceleration and the metabolic equivalents (METs). The sum of magnitude of acceleration is the sum of the rate of change of velocity of the tree axis of the ActivPal accelerometer. This is calculated using the following formula: VM = +. Later, the MET values can be classified into different activity intensity categories. These different activity intensity categories are developed using cut-points.

Commented [A(27]: What is cut-points? Van waar komen ze?

Commented [A(28]: Dit is nu beetje random. Waarvoor wordt het geclassified in intsensity categorie?

Recognizing the type of physical activity can be done using the tri axial ActivPal accelerometer and the use of classification trees, neural networks, or hidden Markov models (Bonomi, 2010). For our task of classifying the types of activities performed we used a random forest model as this model is less likely to overfit compared to a classification tree model and because this model gave good results.

Prediction models that account for the type of activity performed result in more accurate estimates of human energy expenditure (Bonomi, 2010). …….

For next iteration:

- At the end explain with references from literature how you could interpret the results to see if people did enough exercising.
- Fill in the paragraph with the dots to explain with references from literature how you can estimate human energy expenditure.
- Add in-text citations while you're writing and later on add them to the bibliography at the end of the paper.

Theory – V2 (Mark)

2 Theory

In this chapter we will talk about the underlying theory of this research and the problems we are trying to solve to make it clearer what the different important elements of our research are.

The goal of this research is to use machine learning and the accelerometer data to calculate if the respondents moved 150 min, moderately intense in the time period of a week. This can be done by firstly classifying the activities performed with machine learning in the accelerometer data, and after this is done using regression analysis to estimate the MET (Metabolic Equivalent of Task) value of the activities performed. Then, with this MET value it is possible to know what the intensity of the activities performed is.

2.1 The activPAL accelerometer

An accelerometer is a sensor that is used to measure acceleration due to gravity, this way together with the build in gyroscope an accelerometer can figure out the angle at which it's tilted at with respect to earth. The accelerometer has three axes so that the vector in the x, y and z direction can be measured. With these measurements a person's movement can be measured. The data of the accelerometer is saved as a CSV-file.

Commented [W(29]: THEORY V2: Moet meervoud zijn

Commented [A(30]: Is dit nodig? lijkt me beetje overbodig

Commented [B(31R30]: Volgens mij is dit juist wel een mooie inleiding en is het beter om er niet gelijk in te springen

Commented [W(32]: THEORY V2: dit deel van de zin is wat vaag, ik kan er niet uithalen waar het over gaat

Commented [B(33R32]: nu duidelijker?

Commented [W(34]: THEORY V2: Volgens mij beter verwoorden als; the goal of this research is.. ipv ons goal is. Correct me if im wrong

Commented [S(35]: dat is niet het goal

Commented [S(36]: "to calculate if the respondents performed 150 minutes of modetely intense physical activity per week." ik zou het zo opschrijven

Commented [W(37]: THEORY V2: Is dit echt het doel? Volgens mij kan dit veel concreter beschreven worden. Op blackboard staat: Are we at the end able to calculate if people have moved 150 minutes, moderately intense? Dit is volgens mij belangrijk om te vermelden.

Commented [A(38]: Dieper ingaan. Wat wordt MET voor gebruikt?

Commented [B(39R38]: Je moet in dit hoofdstuk niet teveel in detail treden en ik ga er al enigzins op in in de volgende zin

Commented [S(40]: beetje een rare zin. Volgens mij probeer je hier te zeggen dat je de MET waarden kan categoriseren?

Commented [B(41R40]: ja, maar hoezo is dit raar dan?

Commented [S(42R40]: Wat je zegt is dat je met de MET waarde kan inschatten hoe intensief een activiteit was..... het is raar omdat MET een meeteeinheid van "intensiviteit" is en die wordt juist geschat door de modellen zelf

Commented [W(43]: THEORY V2: Hier beter de activPAL accelerometer benoemen ipv accelerometer in het algemeen

Commented [S(44]: de hoeken wordt gemeten door een gyroscope sensor in de activPAL en heeft niet met de zwaartekracht sensor te maken, zo staat het wel in de zin

2.2 MET

The MET value is the objective measure of the ratio of the rate at which a person expends energy, relative to the mass of that person, while performing some specific physical activity compared to a reference, set by convention at 3.5 mL of oxygen per kilogram per minute, which is roughly equivalent to the energy expended when sitting quietly.

Physical activities may be rated using METs to indicate their intensity per time unit of one minute and are calculated using the following formula:

$$\text{MET} = \frac{Energy\ expenditure}{Resting\ energy} = \frac{VO2;max}{3;5\ kg\ min}$$

While minute-estimates of MET are of interest in some studies, researchers are typically more interested in obtaining estimates of time spend in different activity intensity categories during the day or the week. In this context, the literature has established MET cut-points for activity intensity categories, which are: under 3 MET: light-intensity, 3 to 6 MET: moderate-intensity and over 6 MET: vigorous-intensity (Eidi Sasaki et al., 2016).

2.3 The Vyntus® CPX system

The Vyntus® CPX system is an oxygen gas meter that does pulmonary function testing and cardiopulmonary exercise testing. This way it can analyze the breathing of a person and therefore together with the accelerometer data can give a clearer picture of a person's energy expenditure.

Commented [A(45]: Dieper ingaan. Wat wordt MET voor gebruikt?

Commented [A(46]: Wetenschappelijk bron vinden

Commented [A(47]: Bron?

Commented [S(48]: referentie ontbreekt

Commented [S(49]: denk dat ook vermeld moet worden dat MET een meeteinheid per minuut is

Commented [S(50]: ik zou dit eerst onder 2.2 zetten en dan de formule

Commented [B(51]: cut-off points beschrijven

The device consists of a mask that can be placed over the mouth and the nose of a respondent, the amount of oxygen a respondent breathes can then be measured with this device. To calculate the MET, we use the VO_2 max value (also known as maximal oxygen uptake) that this device produces. This is the maximum rate of oxygen consumption measured during incremental exercise; that is, exercise of increasing intensity. The name is derived from three abbreviations: "V" for volume, "O_2" for oxygen, and "max" for maximum.

Commented [S(52]: is het maximal of maximum?

Commented [B(53R52]: maximal

Commented [A(54]: Beschrijv welke data we ervan gebruiken

Commented [S(55]: misschien te gedetailleerd en kunnen we dit weglaten? weet het niet zeker

Commented [B(56R55]: lijkt mij juist wel informatief en het gaat volgens mij ook niet heel erg in op de details

Methods

Describe the context and setting of the study

- Identify the main study variables
- Specify the study design
- Describe the 'population' (patients, doctors, hospitals, etc.)
- Describe the sampling strategy
- Describe the intervention (if applicable)
- Describe data collection instruments and procedures
- Outline analysis methods

Subjects

The total data population was composed of 34 adults (23 male and 18 female). In the data cleaning process, we had to leave out 16 subjects because of the following reasons …. . This left us with population of 25 adults (13 male and 12 female).

Commented [T(57]: Criteria van de participanten missen

Commented [T(58R57]: Wie zijn uitgesloten of wie zijn included?

Commented [T(59]: Verwijzen naar je tabel anders staat die er voor niks. Wat lees ik in de tabel.

Table 1. Subject characteristics (N = 25)

Parameter	Mean +- SD	Range
N (M/F)	25 (13/12)	
Age (categorical)		20 - 74
Body Mass(kg)	75.08 ± 12.60	50.4 – 100.3
Height (cm)	178. 76 ± 10.52	157 – 194.5
Body Mass Index	23.55 ± 3.91	18.18 – 32.72

Commented [T(60]: Tabel volgens paper regels doen.

Table 1. Female subjects characteristics.

	Baseline n = 24	After weight loss n = 12	95% CI
Age, y	47 ± 10	46 ± 11	
Height, m	1.70 ± 0.06	1.71 ± 0.07	
Body weight, kg	119.5 ± 26.8	99.8 ± 18.7**	(- 9.9; - 19.3)
BMI, kg/m^2	41.2 ± 9.1	33.6 ± 4.9**	(- 3.4; - 6.5)
Waist circumference, cm	114.6 ± 16.8	96.4 ± 8.6**	(- 8.8; - 14.3)
Fasting plasma glucose, mmol/L	5.8 ± 2.4	5.4 ± 1.8	
Fasting plasma insulin, mmol/L	11.9 ± 7.8	7.8 ± 3.4*	(- 0.1; - 5.6)
HOMA	3.7 ± 4.8	1.9 ± 1.1*	(- 0.04; - 1.45)
Blood Pressure			
Systolic, mmHg	138 ± 15	130 ± 13*	(- 2; - 21)
Diastolic, mmHg	83 ± 8	79 ± 12*	(- 2; - 11)
Heart rate, bpm	78 ± 10	74 ± 9	

*, P < 0.05;
**, P < 0.001;
95% CI, confidence interval of differences (After weight loss - Baseline).

Commented [T(61]: Dit design aanhouden

Study design – V1 (Adnan)

Data was collected from subjects in a lab setting. Each subject was invited for a lab research. In the lab the subject was given an ActivPal accelerometer which were then attached to their upper thigh. The subject was then attached to Vyntus device. Vyntus device were used to collect data about their oxygen intake (VO_2) while they were performing activities. Each subject executed following activities, cycling light, cycling heavy, jumping, climbing stairs, sitting, standing for 5 minutes except climbing stairs and jumping. After the lab session each subject carried ActivPal device for one week. At the same time, they kept a diary in which they wrote their activities.

Commented [T(62]: Onnodige regel

Commented [T(63]: Toevoegen dat traplopen en springen uitgevoerd zijn voor 1 minuut

Study design – V2 (Mark)

The data was collected from respondents in a lab setting. This was done instead of the method CBS used to do where they would let respondents fill in a SQUASH-questionnaire (Short Questionnaire to Assess Health-enhancing physical activity). This is because the questionnaire was not accurate enough in getting to know how active a person is. Some activities were overestimated while other activities like housekeeping were forgotten by the respondents.

Instead, each respondent was invited for lab research. In the lab the respondent was given an ActivPal accelerometer which were then attached to their upper thigh. The respondent was then attached to the Vyntus device. The Vyntus device was used to collect data about the respondent's oxygen intake (VO2) while the respondents were performing activities. Each respondent executed the following activities: cycling light, cycling heavy, jumping, climbing stairs, sitting, standing for 5 minutes except climbing stairs and jumping. After the lab session each respondent carried an ActivPal device for one week. At the same time the respondents kept a diary in which they wrote their activities.

Study design – V3 (Matthew)

Each respondent was invited for lab research. In the lab the respondent was given an ActivPal accelerometer which were then attached to their upper thigh. The activPAL accelerometer measures the physical activity level (PAL) of the respondent. Oxygen intake (VO2) was measured in the midday after the regular exercises. This measurement was done by using the VYNTUS™ ONE pulmonary function system. Each respondent executed the following activities: cycling light, cycling heavy, jumping, climbing stairs, sitting, standing for 5 minutes except climbing stairs and jumping while being inside the lab. After the lab session each respondent carried an ActivPal device for one week without supervision. In this time the respondents were asked to keep a diary in which they wrote what physical activities they been doing while wearing the activPAL.

Commented [S(64]: wordt in hoofdstuk Theory al uitgelegd dus kan je gwn VO2 gebruiken vgm

Commented [T(65]: Controleren Anniemeke

Commented [S(66]: veranderen naar andere naam van Vyntus

Commented [S(67]: "performed" vind ik beter passen in deze zin

Commented [S(68]: comma weg moet "and" zijn

Commented [S(69]: "they have been doing"

Dataset split in Train, Validation and Test – V1 (Adnan)

The splitting of the train, validation and test sets was executed in two steps. First 3 of 25 subjects was extracted from the subjects to be used as test dataset. These 3 were most representative against the training dataset which was determined with the algorithm described in previous Alinea. Then the data of remaining 22 subjects was split in to training (80%) and validation (20%) dataset which was done with training_test_split function from Sklearn.

Dataset split in train, validation and test – V2 (Mark)

To assure representativeness of the test set against the training set, a random forest model is used. To this random forest model, a value of 0 is assigned to the random state parameter. The value of 0 is used because this makes sure the results are reproducible. Also, a random number of 500 trees was assigned with the use of the n_estimators parameter. If this model cannot classify between what the train dataset, and what is the test dataset is, by giving a classification score of 0.5. The selected test dataset is representative of the train dataset.

The splitting of the train and test dataset is executed in two steps. First, 3 of the 25 subjects were extracted from the subjects to be used as test dataset. These 3 are most representative of the training dataset, which is determined with the algorithm described in the previous alinea. Then the data of the remaining 22 subjects is split into a training (80%) and validation (20%) dataset. This is done making random partitions for the two subsets.

MET regression models (V1 By Colin)

To prepare the features for the MET prediction models we combined data from multiple sources. For the creation and calculation of the features we used the activPAL accelerometer data, Vyntus One data and the 'respondents' file data that was supplied by CBS. This 'respondents' file contains different characteristics from all the respondents that participated in the lab research. The characteristics we used from the 'respondents' file for our features are length, weight, gender, age category, if the respondent meets the balance guidelines, if the respondent meets the bone and muscle guidelines and if the respondent is sporting.

Since most of these characteristics were non-numerical or string values, we converted all features to numerical values. These numerical values differed between simple True/False converted to 1 or 0 and numerical values that represent a category, in our case the age category. The age category '15-19' got a numerical value of 0, '20-24' got a numerical value of 1, etcetera. By converting our features to numerical values our Machine Learning models were able to be trained and evaluated (Brownlee, 2020).

Commented [A(70]: Papers die ik heb gelezen bevat geen ik vorm. Mag je ik -vorm gebruiken in paper?

Commented [S(71]: de lezer wordt zo in het water gegooid. misschien de zin anders verwoorden leest wat makkelijker. bijv " The regression models were developed with feature selection in mind. The features were retrieved from multiple sources ....."

Commented [A(72]: ActivPal

Commented [A(73]: Je hoeft geen bestand naam te geven. Je kan zeggen "and file containing charastrics of subject"

Commented [A(74]: Dit kan dan "This file contains" zijn

Commented [A(75]: Ik denk dat het beter klinkt als we het zeggen "file as our features"

A few other features have more complex computations. Since MET is measured in minutes (more in this in chapter... Van Mark), the following features were also resampled to 1 minute. The 'sum of magnitude of acceleration', which means the total acceleration within a certain timespan, is resampled to 1 minute after applying the following formula (Measurement of Physical Activity Using Accelerometers, 2016) on the *X*, *Y* and *Z* data from the activPAL accelerometer:

$$\sqrt{x^2 + y^2 + z^2}$$ // to do equation numbers

The last feature is the speed, which is also resampled to 1 minute for every activity. The following formula (Calculate speed from accelerometer, 2014) was used to calculate the velocity from the acceleration of the X, Y and Z axis. The velocity is needed to calculate the speed.

$$v(t) = v(0) + \sum a \times \delta t$$

The *t* is the *time interval* of the x, y or z velocity. In this case 0.05 seconds (Why activPAL?, z.d.). The *a* is the *acceleration* of the *X, Y* or *Z* axis. After calculating the velocity for the X, Y and Z axis it is possible to calculate the speed. The following (Calculate speed from accelerometer, 2014) formula has been used. The *X, Y and Z* inputs are taken from the velocity formula named above.

$$|v| = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

$$\sqrt{v_x^2 + v_y^2 + v_z^2}$$

The MET value, sum of magnitude of acceleration and speed were the only features that were not static and would change its value based on the resampled minute.

Commented [A(76]: Je kan dit verwijzen als equation 1. Dit zie ik in andere paper ook gebeuren.

Commented [A(77]: Dit wordt dan equation 1

Commented [A(78]: zou je het niet moeten assignen naar een getal zoals eerder formule?

Commented [A(79]: Overbodig?

Brownlee, J. (2020, 26 augustus). 3 Ways to Encode Categorical Variables for Deep Learning. Machine Learning Mastery. https://machinelearningmastery.com/how-to-prepare-categorical-data-for-deep-learning-in-python/

Measurement of Physical Activity Using Accelerometers. (2016, 1 januari). ScienceDirect. https://www.sciencedirect.com/science/article/pii/B9780128020753000024

Why activPAL? (z.d.). PALT. http://www.palt.com/why-activpal/

Calculate speed from accelerometer. (2014, 14 december). Physics Stack Exchange. https://physics.stackexchange.com/questions/153159/calculate-speed-from-accelerometer

Commented [B(80]: verplaatsen naar referentie pagina

MET regression models (V2 By Colin)

To predict the MET values of the activities, different features are needed for our prediction models. Multiple data sources were combined to prepare the different features for the prediction models. For the creation and calculation of the features we used the activPAL accelerometer data, Vyntus One data and a file containing the respondent characteristics that was supplied by CBS. This file contains different characteristics from all the respondents that participated in the lab research. The following characteristics were used for the creation of our features: length, weight, gender, age category, if the respondent meets the balance guidelines, if the respondent meets the bone and muscle guidelines and if the respondent is sporting.

Since most of these characteristics were non-numerical or string values, all of these features were converted to numerical values. These numerical values differed between simple True/False converted to 1 or 0 and numerical values that represent a category, in our case the age category. The age category '15-19' got a numerical value of 0, '20-24' got a numerical value of 1, etcetera. By converting our features to numerical values our Machine Learning models were able to be trained and evaluated (Brownlee, 2020).

A few other features have more complex computations. Since MET is measured in minutes (more in this in chapter... Van Mark), the following features were also resampled to 1 minute. The 'sum of magnitude of acceleration', which means the total acceleration within a certain timespan, is resampled to 1 minute. The formula (Measurement of Physical Activity Using Accelerometers, 2016) can be found in **equation #add number**. The *X*, *Y* and *Z* data from the activPAL accelerometer is used to calculate the acceleration.

$$\sqrt{x^2 + y^2 + z^2}$$ // **Equation #**

The last feature is the speed, which is also resampled to 1 minute for every activity. **Equation #add number** (Calculate speed from accelerometer, 2014) was used to calculate the velocity from the acceleration of the X, Y and Z axis. The velocity is needed to calculate the speed.

$$v(t) = v(0) + \sum a \times \delta t$$ // **Equation #**

The *t* is the *time interval* of the x, y or z velocity. In this case 0.05 seconds (Why activPAL?, z.d.). The *a* is the *acceleration* of the *X, Y* or *Z* axis. After calculating the velocity for the X, Y and Z axis it is possible to calculate the speed. **Equation #add number** (Calculate speed from accelerometer, 2014) has been used

to calculate $|v|$ which is the total speed. The *X, Y and Z* inputs are taken from **Equation (velocity number)**.

$$|v| = \sqrt{v_x^2 + v_y^2 + v_z^2}$$ // **Equation #**

Once the features were created, 2 different ensembled decision tree regression models were configured. The Random Forest and XGBoost model were chosen and configured as identical as possible to pick the best performing model for each activity. To pick the optimal combination of features, the implementation of Recursive Feature Selection (RFE) was applied on both models. The chosen features were extracted from the prepared data frame with the method described in chapter from Mark(about train/test/split method). Finding the optimal amount of decision trees was decided experimentally. A function was written to find the most optimal amount of trees between a certain range of the related model. Hyperparameter tuning was eventually applied on both models to make sure the configuration was implemented in the best possible way.

Activity classification

International recommendation of PA - V1 Ali

Once the classification and regression models have been completed, software was created to analyze a subject's weeks' worth of activPAL data. The method of data preprocessing applied to this data is the same as previously described in activity classification and MET regression chapters. The idea is to let the activity classification model label each data row with an activity, so the data could be grouped by activity and passed onto the corresponding MET regression model. The MET regression models add the predicted MET-value to the data it has gotten and resample the data to minute by minute. In the end the MET regression models combined would have predicted MET-values for all the activities that were recognized in the data. In order to retrieve the time spent performing moderately intense PA in the predicted and classified data, "moderately intense PA" must be defined. According to /// moderately intense PA is PA with a MET-value ranging from 3 to 6 MET. To answer whether people performed at least 150 minutes of moderately intense PA per week, the classified and predicted data were filtered by rows containing a MET-value higher than 3. The number of rows in this filtered data represent the number of minutes spent performing moderate and intense PA. If the number of rows is higher than or equal to 150, it can be concluded that the subject adhered to the recommendation of at least 150 minutes of moderately intense PA for that week.

Commented [A(81]: in chapter methods

Commented [A(82]: Dit deel verwart me erg

Commented [A(83]: "could be" klinkt alsof het gedaan kan worden. Ik denk dat je beter kan zeggen dat het gedaan is

International recommendation of PA - V2 Ali

In order to answer if a subject adheres to the international recommendation of PA, software was written that analyzes a subject's week data. The software implements the same way of data preprocessing, classification and regression models as previously described in methods chapter. After preprocessing, the software uses the classification model to label the week data with recognized activities. Thereafter, the activities are grouped by activity and passed onto the corresponding regression models who then add the MET predictions. At this point the week data is classified and contains MET predictions. To retrieve moderately intense and vigorous PA in the data, moderately intense and vigorous PA must first be defined. According to /// moderately intense PA is PA with MET ranging from 3 to 6 MET and vigorous PA is PA with MET above 6. So, by filtering out data with MET lower than 3, moderately intense and vigorous PA data remains. The number of rows in this filtered data represents the number of minutes spent performing at least moderately intense PA, since each row in the data represents a minute. If this number is equal to or greater than 150, it can be assumed the subject adheres to the international recommendation of PA for that week. The output of the software is this number and a linear plot, colored by activity, showing MET on Y axis and time on X axis.

Commented [A(84]: Extra

Commented [A(85]: Waarom zeg je niet dat je filter op data met waarde gelijk of hoger dan 3

Results

- Report on data collection and recruitment (response rates, etc.)
- Describe participants (demographic, clinical condition, etc.)
- Present key findings with respect to the central research question
- Present secondary findings (secondary outcomes, subgroup analyses, etc.)
- Describe your results and include appropriate plots. Do not include a discussion of your results.

MET regression

Activity classification

International recommendation of PA - V1 Ali

The output of the software is a plot showing an overview of activities recognized in the activPAL data and the corresponding MET-values. This result has been visually compared to the result of activPAL's own software named palAnalysis. The reliability of the software solely depends on the models integrated in the software. The results of these models have been described in chapters …….

The first image underneath is the result of our software. It shows the activities of a subject performed on a single day, starting from 13:00, and the predicted MET-value for those activities. The recognized activities were sitting, standing, walking, running and cycling light & heavy. The MET-values range from 1 to 6.5 for this day. The second image is that of activPAL's software; palANALYSIS. The palANALYSIS software does not differentiate between walking & running and cycling light & heavy. Also, the color for cycling is purple in palANALYSIS and light blue in our software.

/// bronvermelding

Commented [A(86]: Hoort dit niet in methods?

Commented [S(87R86]: yup kan vgm wel daarnaar verplaatst worden

Commented [A(88]: Hoe werkt het wel?

Commented [S(89R88]: het maakt geen onderscheid tussen lopen en rennen bv is het dan niet duidelijk dat ze samengevoegd zijn?

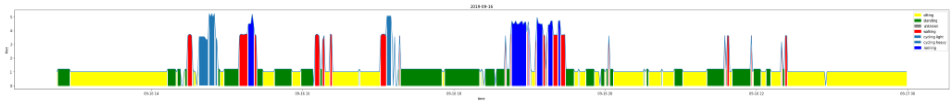
/// bronvermelding

International recommendation of PA - V2 Ali

// Software result

// color code table

Figure (add figure number) is one of the plots from the developed software output. It shows activity and MET data of a single subject on a single day, starting from 13:00. The recognized activities are sitting, standing, walking, running and cycling light & heavy. The MET-values range from 1 to 6.5. The number of minutes spent performing at least moderately intense PA is 645 for that week. The reliability of the software solely depends on the models integrated in the software. The results of these models have been described in chapters …….



/// bronvermelding

Discussion (Matthew)

Punten:

- Speed calculation > kalibratie toepassen voor accurate berekeningen. Huidige speed berekening zorgde voor betere model resultaten, alleen weten we niet of de speed correct is toegepast omdat er niet gekalibreert is.
- Grote van dataset voor de MET prediction modellen, te weinig data waardoor overfitting van de modellen plaatsvond. Meer data kan door de activiteiten langer uit te voeren of meer respondenten aan het onderzoek mee te laten doen.

- De MET-regression models werken goed wanneer ze eigenschappen van subjecten zien. Zodra er onbekende subject in de model komt dan werken ze heel slecht. Dit kan meerdere oorzaken hebben maar wat wij vermoeden te weining data
- De ActivPal accelerometer moet gecallibebreerd worden. We zien in data dat acceleratie bijna nooit op 0 waarde krijgt. Dit heeft invloed op de MET-regression models omdat we magnitude of acceleration en snelheid feature gebruiken. De snelheid stijgt constant over de tijd.
- De data die we mee hebben gewerkt gebruikt categorieen voor leeftijd. Het zou beter zijn als we getallen hebben.

· State the main findings of the study

· Discuss the main results with reference to previous research

· Discuss policy and practice implications of the results

· Analyse the strengths and limitations of the study

· Offer perspectives for future work

· Discuss your results and compare your results to similar studies.

Conclusion (adnan + colin)

Reference

Voor introduction:

Global recommendations on physical activity for health (who.int)

physical-activity-recommendations-18-64years.pdf (who.int)

- Sasaki, E., Samara da Silva, K., Gonçalves Galdino da Costa, B., & John, D. (2016). Measurement of Physical Activity Using Accelerometers.