Project no: 689043
Acronym : SELFBACK
Title: A decision support system for self-management of low back pain
Activity: PHC-28-2015 Predictive modelling RIA

Work Package 2:	Predictive Monitoring Analytics
Deliverable D2.5:	Results from WP2 pilot study (PILOT) and empirical evaluation of algorithms
Organisation name of deliverable lead:	RGU
Author(s):	Sadiq Sani, Nirmalie Wiratunga, Stewart Massie, Kay Cooper
Reviewer(s):	Kerstin Bach
Participants:	NTNU
Type:	Other
Dissemination Level:	PU
Version:	1.0
Total no of pages:	37
Project Start date:	1. January 2016
Contractual delivery date:	30. September 2018
Actual delivery date:	30. September 2018
Keywords:	Feasibility Study, Empirical Evaluation
Status:	Submitted

Abstract
The aim of Deliverable 2.5 is to report on empirical evaluation of activity recognition algorithms and feasibility study of the physical activity and messaging components of the SELFBACK application which was conducted in Aberdeen. The knowledge generated in this study will inform amendments and further development prior to other consortium partners undertaking a pilot study of the near-final version of the app, which will in turn inform the randomized controlled trial due to commence in 2019.

This project has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement No 689043.

Document History

Version	Date	Author(s)	Description
0.1	25/09/18	Kay Cooper, Sadiq Sani, Nirmalie Wiratunga, Stewart Massie	Initial version of the document
0.2	29/09/18	Sadiq Sani	Revised version of the document
1.0	29/09/18	Kerstin Bach, Paul Jarle Mork	Submitted version to the EC

Table of Contents

1 Introduction

This report describes the empirical evaluation of physical activity recognition algorithms and feasibility study of physical activity monitoring for WP2, task 2.5. The original intention was to explore the feasibility, acceptability and usefulness of the initial version of the SELFBACK intervention (physical activity + exercise + education components). However, due to the unforeseen and unavoidable delay in the initial version of SELFBACK being ready for testing, the consortium decided that a feasibility study during months 23-26 to trial the baseline questionnaires and initial physical activity component of SELFBACK would provide important knowledge for further development of the SELFBACK intervention, and the design of the pilot and RCT studies.

The UK Medical Research Council guidance on the development of complex interventions [1] recommends that establishing feasibility and piloting are important elements of developing complex interventions, such as self-management interventions for low back pain (LBP). Whitehead et al [2] and Eldridge et al [3] further described feasibility studies as separate to pilot studies, defining them as preliminary studies that essentially ask whether "something can be done, should we proceed with it and if so, how" [3]. Arguably of equal importance is acceptability, defined as the extent to which people consider an intervention to be appropriate [4]. We therefore explored the feasibility and acceptability of the initial SELFBACK physical activity component in this study; hereafter referred to as the feasibility study.

The following report details the feasibility study aims and objectives, methodology and methods, results, and how the results informed further development of the SELFBACK intervention and the design of the pilot and RCT studies.

The Robert Gordon University (RGU) School of Health Sciences Ethics Review Panel granted ethical approval for the study (Ref: SHS/17/14). The ethcs approval is documented in WP8 (D8.1, D8.2, D8.7, D8.8).

2 Feasibility study: Aims & Objectives

The **aim** of the study was to explore the feasibility and acceptability of the available components of the SELFBACK intervention: baseline questionnaires, physical activity monitoring & feedback strategies.

The specific **objectives** were as follows:

2.1 Feasibility objectives

1. To measure completion rates (time & completeness) for baseline questionnaires in a sample of people with LBP using the SELFBACK app
2. To explore interaction with the SELFBACK app (user activity)
3. To explore any user-identified difficulties in engaging with the SELFBACK app

2.2 Acceptability objectives:

To explore the opinions of people with LBP on:

1. the content of the SELFBACK physical activity intervention component
2. the effort required in interacting with the selfBACK physical activity intervention component
3. the mode of delivery of the SELFBACK intervention physical activity component
4. barriers and facilitators to using the SELFBACK physical activity intervention component
5. the perceived usefulness and effectiveness of the SELFBACK physical activity intervention component
6. confidence that they could participate in the SELFBACK physical activity intervention longer-term

2.3 And:

To determine which amendments should be made to the SELFBACK physical activity component prior to further pilot testing

3 Feasibility Study: Methodology & Methods

3.1 Study design

We conducted a sequential explanatory mixed-methods study [5]. Quantitative data in the form of completion rates for questionnaires and user activity were collected, and an electronic survey was used to gather opinions from a sample of people with LBP on feasibility and acceptability of the initial SELFBACK physical activity component. Thereafter, a subgroup of participants took part in qualitative semi-structured interviews to further explain and interpret the quantitative findings.

3.2 Population and sample

The population of interest for the SELFBACK study is adults (aged 18+) with LBP (any duration) being recommended for self-management by their healthcare provider. As this feasibility study was not evaluating effectiveness of the SELFBACK physical activity component, but its acceptability and feasibility, it was not necessary to recruit from the target population. However, it was necessary to recruit participants with LBP in order for them to provide informed opinions on the intervention. We therefore recruited adults (aged 18+) with LBP (any duration) from (i) RGU: Wellness physiotherapy clinic; and (ii) RGU staff and students.

3.3 Recruitment

3.3.1 (i) RGU: Wellness

Participants with LBP who were being recommended for self-management (+/- further physiotherapy) and deemed suitable for the study by their physiotherapist were provided a study information pack (participant information sheet, reply slip, freepost envelope). Once they had read the information sheet, participants could either (i) indicate to their physiotherapist that they consented to their name & contact number being provided to the researcher who contacted them to discuss further, or (ii) return the reply slip/e-mail/telephone the researcher to discuss further.

3.3.2 (ii) RGU staff and students:

An advert was placed in the RGU e-bulletin, RGU website, and on the plasma screens across campus, and e-mails were sent to staff and students in the Schools of Health Sciences and Computing Science & Digital Media asking for volunteers to take part in the study. Interested participants contacted the researcher to discuss further.

On contacting the researcher inclusion/exclusion criteria were checked (Table 1), participants were provided with a participant information sheet if they had not already received one, and they were given the opportunity to ask questions and discuss the study with anyone they wished to prior to providing written, informed consent.

Participants did not receive payment for their participation in the study, but they were entered into a prize draw to win a Kindle Fire Tablet.

Table 1: Inclusion & exclusion criteria for feasibility study

Inclusion Criteria	Exclusion Criteria
Aged 18+	Serious pathology
Have LBP of any duration	Terminal illness
Own an Android smart phone with access to data	Serious depression
Willing to test SELFBACK app & monitoring strategies	Unable to read/speak/understand English
Independently mobile	Non-ambulatory or requires walking aids
	Pregnancy
	Does not own Android smart phone/no access to data
	Fibromyalgia
	Previous spinal surgery

3.4 Protocol

Figure 1 provides an overview of the study processes:

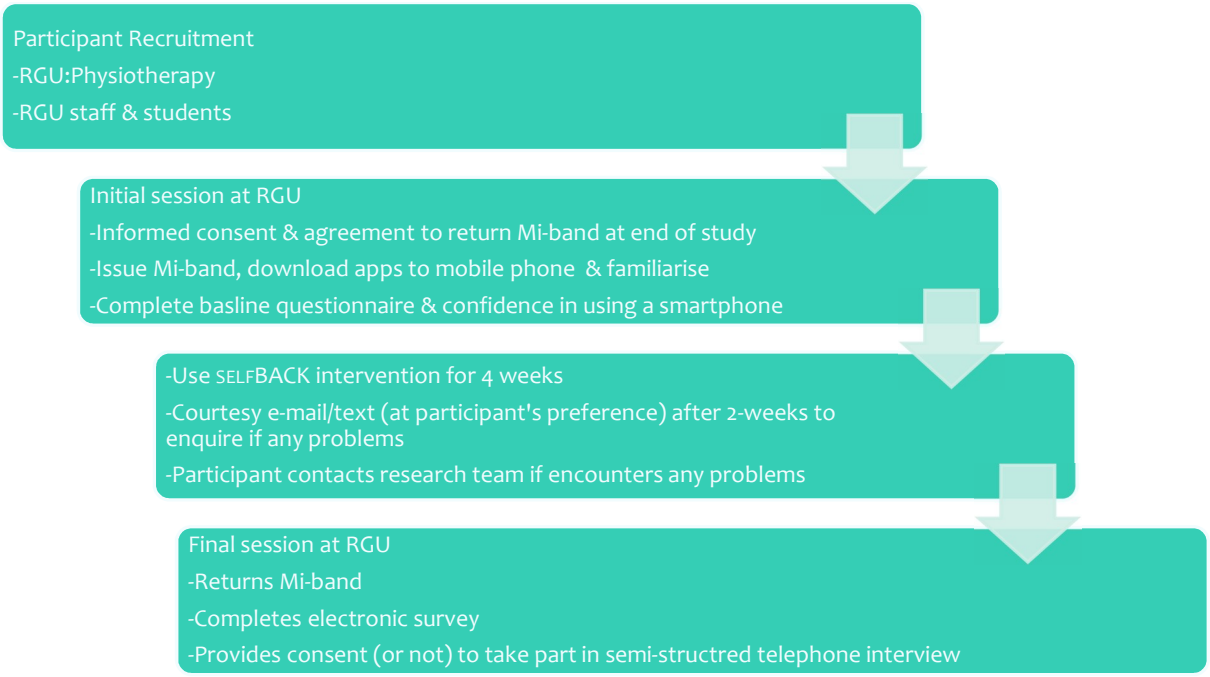

Figure 1: Overview of Feasibility Study Processes

3.5 Outcome measures

i. Demographics were collected from the baseline questionnaires in order to describe the sample and their similarity to the population of interest in the SELFBACK study.
ii. The baseline questionnaire has been reported elsewhere (D1.2). Participants completed this on a computer at the initial session, or if they preferred they could complete the questionnaires at home.
iii. Confidence in using a smartphone [6] was collected in order to further describe the population.
iv. The electronic survey consisted of the 10-item System Usability Scale [7] and a design questionnaire adapted from that used by Reynoldson et al [6] in their 2014 study of pain self-management apps.
v. The semi-structured telephone interviews explored participants' perceptions and experiences of the app, including:
    - Usefulness & appeal
    - Barriers & facilitators to using the app
    - Technical difficulties
    - Features they liked/disliked
    - Usability & interaction required from user
    - Feasibility of using the Mi-band
    - Quality/appropriateness/usefulness of physical activity component
    - Feedback feature (physical activity notifications)
    - General ease of use/navigation
    - Suggestions for improvement/amendments/changes

3.6 Data Analysis

Descriptive statistics were reported for baseline demographics, confidence in using a smartphone, the System Usability Scale, design questionnaire, completion rates for baseline questionnaires, and interaction with the SELFBACK app. Difficulties encountered whilst engaging with the SELFBACK app were reported. Due to the small sample size and feasibility/acceptability objectives we did not undertake any subgroup analyses. As the interview topic guide was relatively structured, transcription was conducted into proformas with appropriate headings and subheadings. Qualitative data was analysed by two members of the study team using framework analysis (descriptive analysis steps) [8].

App interaction statistics were automatically collected on the app by logging, at the end of each day, information such as the number of times a participant launches the app, the time of day the app is first launched, the count of steps taken, the steps goal for the day, the list of messages sent, whether the user opened each message or not and whether the user liked or disliked each message.

4 Feasibility Study: Results

4.1 Demographics

Sixteen participants were recruited and completed the baseline questionnaires. Their demographics and selected baseline measures are reported in Table 2. It can be seen that the mean age was 51.5 (SD 13.9), 63% were male, 75% had chronic LBP, pain and disability levels were mild to moderate, and most participants had high levels of confidence in using a smartphone.

The baseline questionnaires took on average 18-minutes to complete (range 10-38 minutes). Whilst the majority of participants reported no issues with the baseline questionnaires, two participants commented on its length (too long), two felt there were some ambiguous questions, one felt it was confusing if too much thought was given to it (better if completed quickly), and one commented that there was some repetition of questions.

Table 2: Demographics & Baseline Measures (n=16)

Gender	Male 63% (n=10)
Age	23-71 (mean 51.5 SD 13.9)
Employed: Part-time Full-time Retired Other	 25% (n=4) 44% (n=7) 25% (n=4) 6% (n=1)
Body Mass Index	18.8-32.8 (mean 26.2 SD 4.2)
LBP Duration (current episode) 1 week 4 weeks 12 weeks >12 weeks	 6% (n=1) 19% (n=3) 31% (n=5) 44% (n=7)
Pain worst	1-10 (mean 5.3 SD 2.8)
Pain average	1-7 (mean 3.8 SD 2.0)
RMDQ	1-17 median 5.0)
FABQ	7-24 (median 11.5)
FABQ LBP Cause	0-6 (median 3)
Pain self-efficacy	14-58 (mean 51.5)
EQ-5D	40-95 (mean 75.1 SD 14.1)
Perceived Stress Scale	3-25 (median 12)
Perceived Health Questionnaire	0-13 (median 3~)
Confidence using smartphone	Median 4/5

LBP=Low Back Pain; RMDQ = Roland Morris Low Back Pain Disability Questionnaire; FABQ = Fear Avoidance Beliefs Questionnaire; EQ-5D = European Quality of Life Questionnaire

The design questionnaire also asked participants about the baseline questionnaires. Most participants found them easy to complete (94%), found the time taken to complete

acceptable (88%), and reported they would be happy to provide the information in a commercial version of the app (88%). 56% found the questions relevant, whilst 44% were unsure, indicating that a clear explanation of the purpose of the questions and the need for some repetition (due to using a variety of validated tools with some overlap of constructs) is required for the pilot and RCT.

It took an average of 37-minutes to download the required apps (Mi-Fit, Google Fit & SELFBACK; range 20-60 minutes). The long download time might be due to using University Wi-fi; a user guide has been compiled for the pilot and RCT with clear, detailed instructions for downloading and synchronizing the required apps. Two participants were reluctant to download all three apps and questioned why they were all required. In hindsight we had not provided clear information on this in the participant information sheet; this has likewise been incorporated into the information sheets for the pilot and RCT.

4.2 User-App Interaction

Participants' interactions with the app were automatically logged for the entire duration of the trial. The information logged for each participant were weekly step-count goals, daily number of steps taken, number of times they opened the app per day, and number of messages (notifications) received per day. The average step-count goal set was 7003.7 and participants took an average of 5233.1 steps per day. The app was opened an average of 6.2 times per day and an average of 1.8 messages were received per day. Additional details are provided in Table 3.

Table 3: Summary of participants' interactions recoded by app

	Step-count Goal	Steps taken per day	Number of times app opened per day	Messages per day
Min – Max	3,000-12,500	2-20,791	0-95	0-10
Mean (SD)	7003.7 (2931.5)	5233.1 (4397.3)	6.2 (11.8)	1.8 (2.4)
Mode	10,000			

4.2.1 Messages Sent

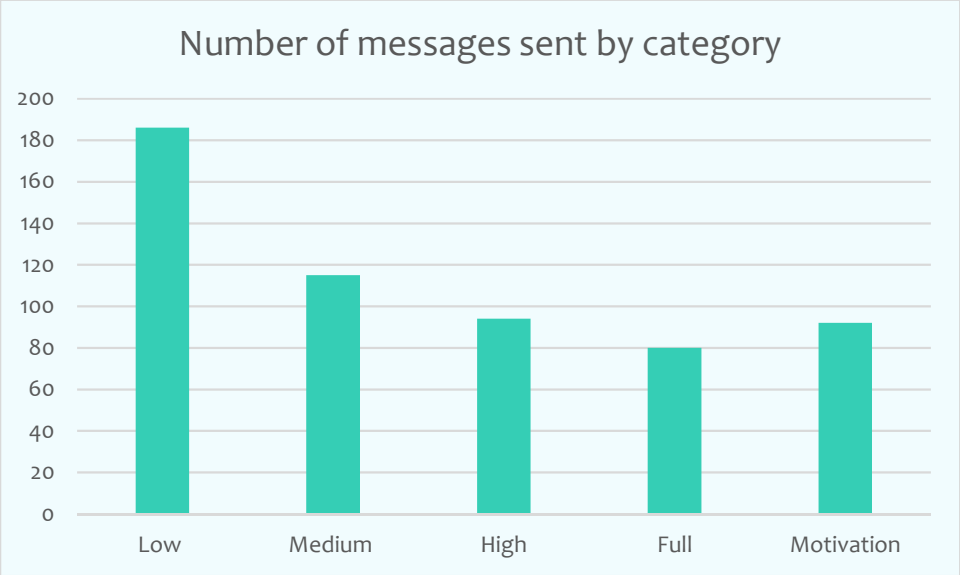

Figure 2: Distribution of total number of messages by category

A total of 569 messages were sent to the 16 participants over 321 recorded trial days. Out of the message categories, the most amount of messages sent were low physical achievement messages (32.7%) while the second highest are medium achievement messages (20.2%). Low achievement messages indicate a step count of less than 50% while medium achievement messages indicate a step count of less than 75%. The fact that most messages belong to these two categories is reasonable considering that an average user is likely to have achieved no more than 75% or their goal most of the day. A full achievement message is sent once a user achieves their goal. A total of 80 full achievement messages were sent which indicates that participants achieved their goal on 25% of the total trial days.

4.2.2 Messages Opened

A log is kept of whether or not a participant opened a message during the trial. A message is opened if the user taps on the message in the Android notification tray or if the user opens the Messages tab in the app and taps on a message in the list. Doing this opens a pop up window containing the message. Note that not opening a message does not indicate that the user did not read the contents of message as the user is able to do so from the Android notification tray without opening the message. Figure 3 provides the distribution of opened and unopened messages. In total, 42.2% of messages were opened during trial. Figure 4 presents the distribution of message by message category.

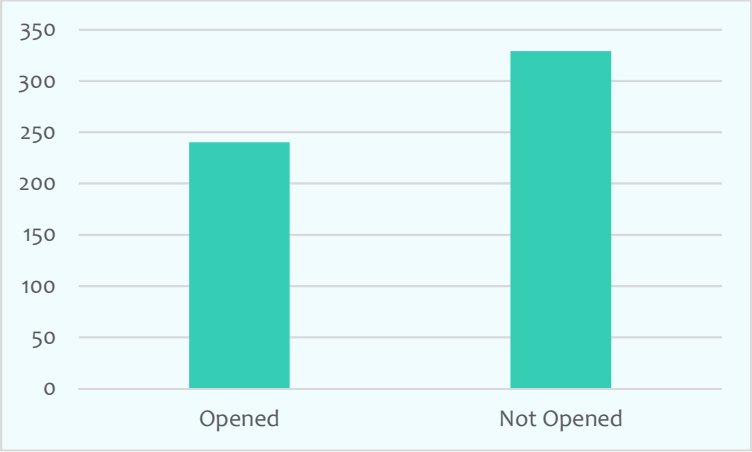

Figure 3 Distribution of opened and unopened messages

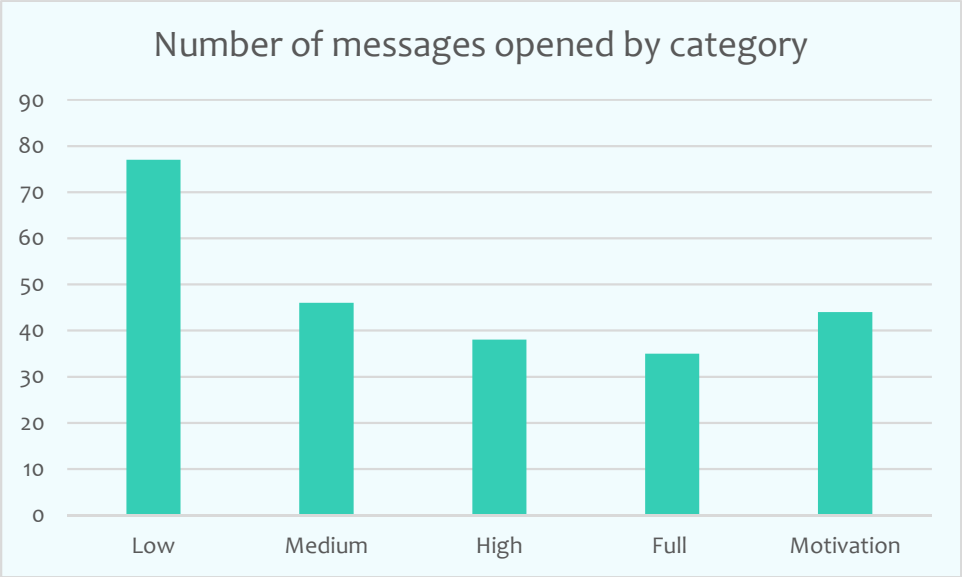

Figure 4 Distribution of opened messages by category

From Figure 4, we can see that Low and Medium achievement categories have the highest numbers of opened messages. However these two categories also have the highest numbers of messages. Thus, it would be beneficial to consider the ratio of opened to total messages received by category. This is presented in Figure 5.

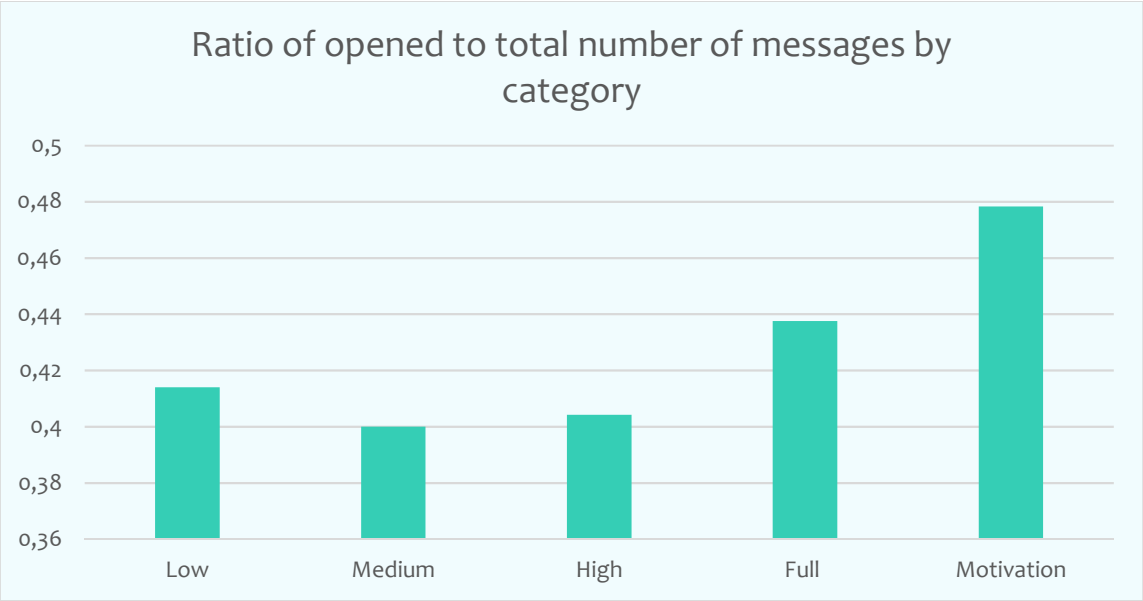

Figure 5 Ratio of opened to total number of messages by category

From Figure 5, the category with the highest ratio of opened to total messages is Motivation, followed by Full. Motivation messages are sent in the morning at the start of the user's day. This indicates that the messages most opened by participants are the ones received at the start of the day.

4.2.3 Messages Unopened

The distribution of unopened messages by category is presented in Figure 6. Here, in absolute numbers, Low message category has the highest number of unopened messages. However, this is also the category with the highest number of messages in total. Accordingly, we present the ration of unopened to total number of messages in Figure 7.

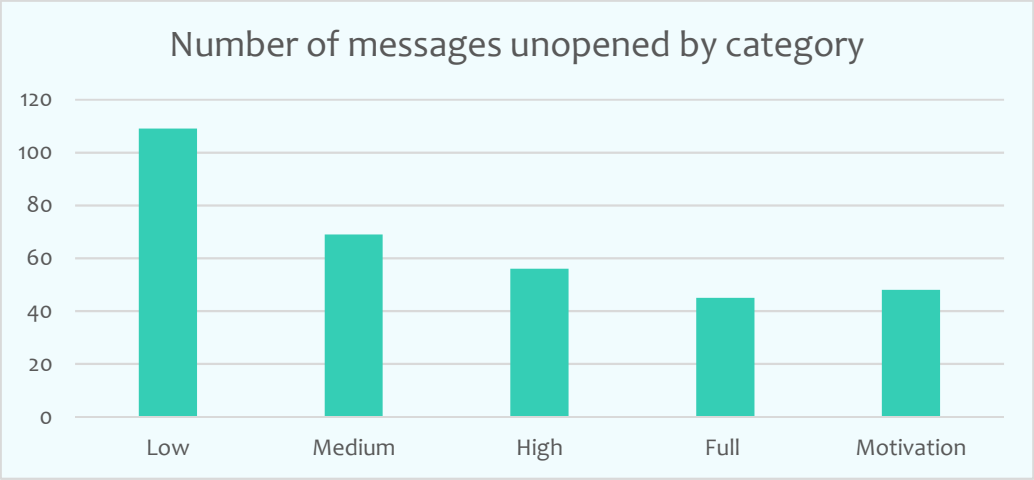

Figure 6 Distribution of unopened messages by category

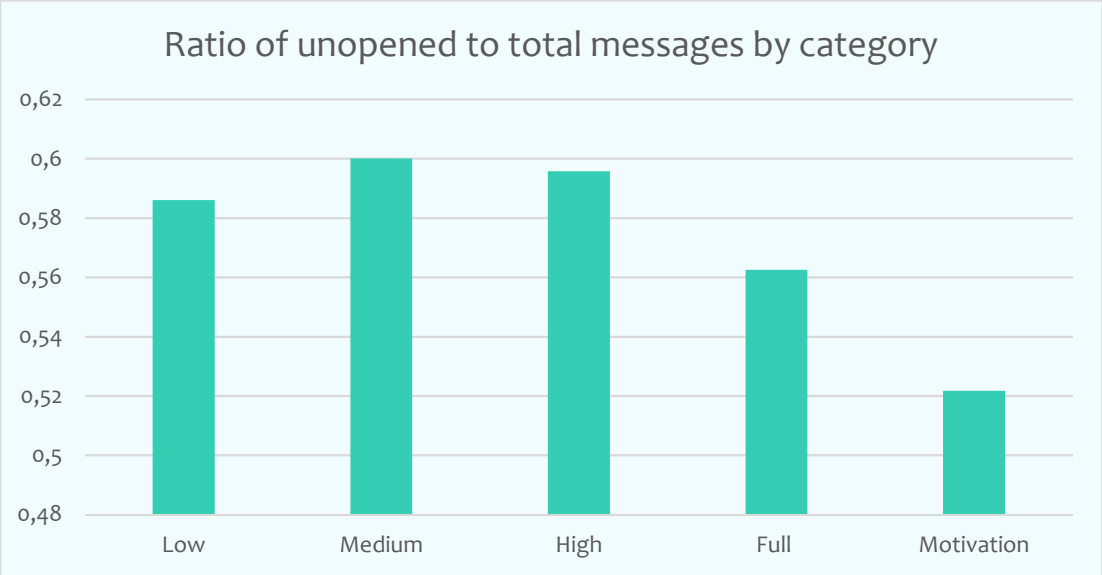

Figure 7 Ratio of unopened to total messages by category

From Figure 7, Medium category has the most number of unopened messages as a ratio of total messages. Motivation messages have the least number of unopened messages, which means this is the most opened category of messages.

4.2.4 Messages Liked

Here, we present sentiment analysis of messages. Sentiment is captured by providing a "Like" and "Dislike" button on the pop-up window of opened messages. Accordingly, sentiment is only available for opened messages. Opened messages where the user does not express an explicit like or dislike are considered "Neutral". Figure 8 presents the distribution of sentiment for all opened messages.

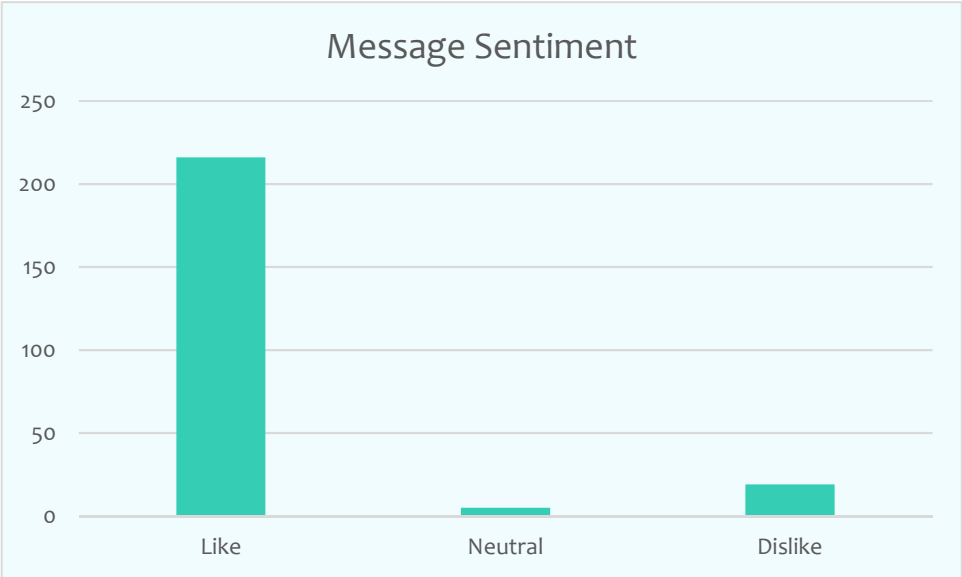

Figure 8 Distribution of message sentiment

From Figure 8, most messages that were opened were liked (216 out of 240) and participants only disliked 19 messages. This indicates that participants overwhelmingly liked the messages they received. Sentiment was not expressed on only 5 opened messages, which further indicates the willingness of participants to indicate their opinion on messages during the trial.

Figure 9 presents the distribution of sentiment by message category.

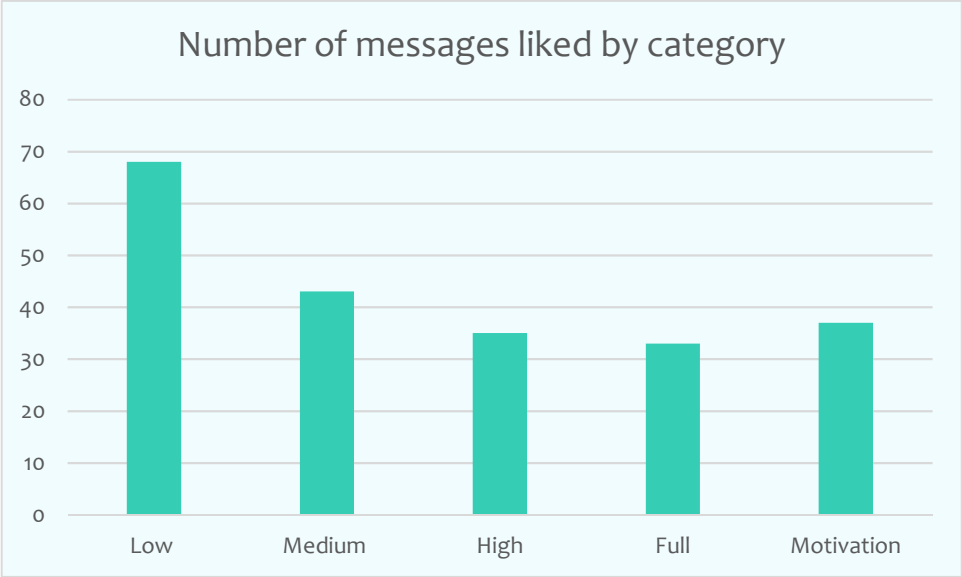

Figure 9 Distribution of message sentiment by category

From Figure 9, we can see that Low achievement messages have the most number of likes and Full achievement messages have the least number of likes. However, Low and Full achievement messages have the highest and lowest number total number of messages respectively. Hence, we further analysed the sentiment of messages by category, as a ratio of the number of opened messages in that category as presented in Figure 10.

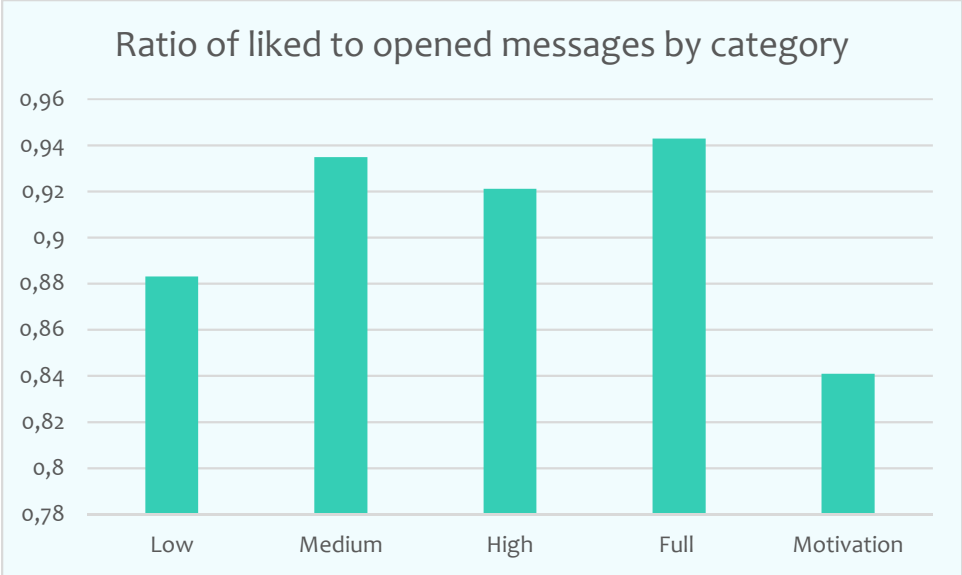

Figure 10 Ratio of liked to number of opened messages by category

From Figure 10, we can see that Full achievement messages are the most liked. This is of little surprise as this message category is the most positive. However, we can also observe that participants liked Medium and High as well. The message category with least likes is Motivation.

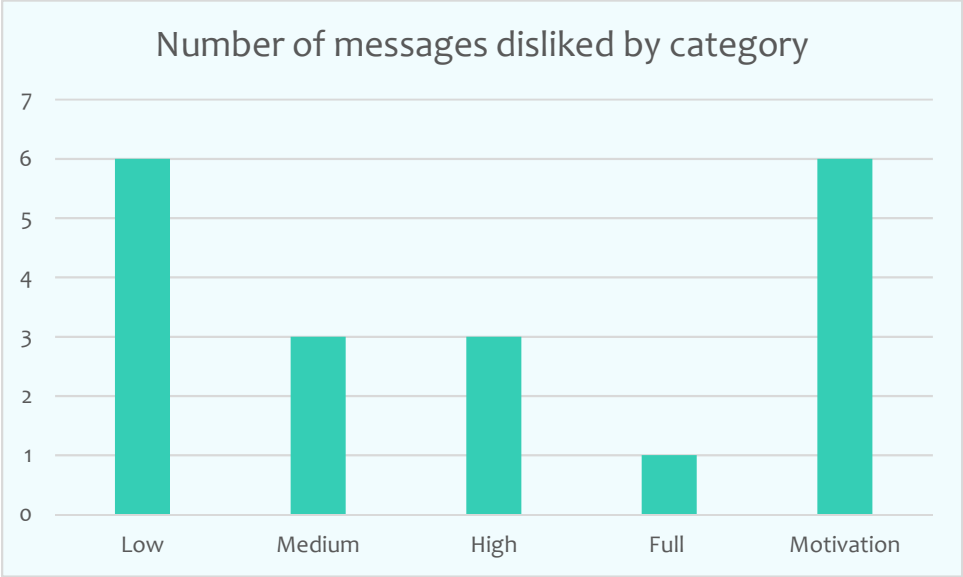

Figure 11 Distribution of messages disliked by category

There are only 24 messages that were disliked and the distribution of these is shown in Figure 11. The most disliked categories are Low and Motivation messages (6 messages were disliked each).

4.3 Electronic survey (n=16)

4.3.1 System Usability Scale

Results of the System Usability Scale are shown in Table 4. Participants found SELFBACK easy to use, felt confident using it, and thought most people would learn to use it quickly. It was not felt to be unnecessarily complex, and support of a technical person was in general not required. However, it was felt that there was inconsistency and only 31% of participants found the functions to be well integrated. Participants were neutral on whether it was cumbersome to use and whether they needed to learn a lot of things before using SELFBACK.

4.3.2 Design questionnaire

Results of the design questionnaire are presented for each of the four sections of the questionnaire: (1) set-up & general use; (2) app content; (3) suggestions; and (4) general questions on app use.

4.3.2.1 (1) Set-up & General Use

37.5% of participants reported it was very easy or easy to set-up and use the app, 31.3% reported that it was difficult or very difficult, and 31.2% were undecided. This confirms the need for clear and detailed instructions for participants in the pilot and RCT. Some of the difficulties in downloading the app were due to the age/version of some participants smartphones; minimum operating system requirements have been documented as an inclusion criterion for the pilot and RCT. Battery life was not unduly affected by the SELFBACK

app (Figure 12). Most participants (81.25%) wore the Mi-Band all day (Figure 13) whilst 50% wore it during the night also. However, two participants were unable to sync the Mi-Fit band with their smartphone and consequently did not use it at all. Several comments were made by participants in relation to set-up and general use; they can be seen in Table 5.

Table 4: System Usability Scale results (n=16)

	Agree (%)	Disagree (%)	Neutral (%)
1. I think that I would like to use the SELFBACK system frequently	43.8	24.9	31.3
2. I found the SELFBACK system unnecessarily complex	12.5	62.5	25.0
3. I thought the SELFBACK system was easy to use	75.0	12.5	12.5
4. I think that I would need the support of a technical person to be able to use the SELFBACK system	26.7	60.0	13.3
5. I found the various functions in the SELFBACK system were well integrated	31.2	43.8	25.0
6. I thought there was too much inconsistency in the SELFBACK system	37.5	56.3	6.2
7. I would imagine that most people would learn to use the SELFBACK system very quickly	81.3	6.2	12.5
8. I found the SELFBACK system very cumbersome to use	12.5	18.8	68.7
9. I felt very confident using the SELFBACK system	75.0	12.5	12.5
10. I needed to learn a lot of things before I could get going with the SELFBACK system	12.5	18.8	68.7

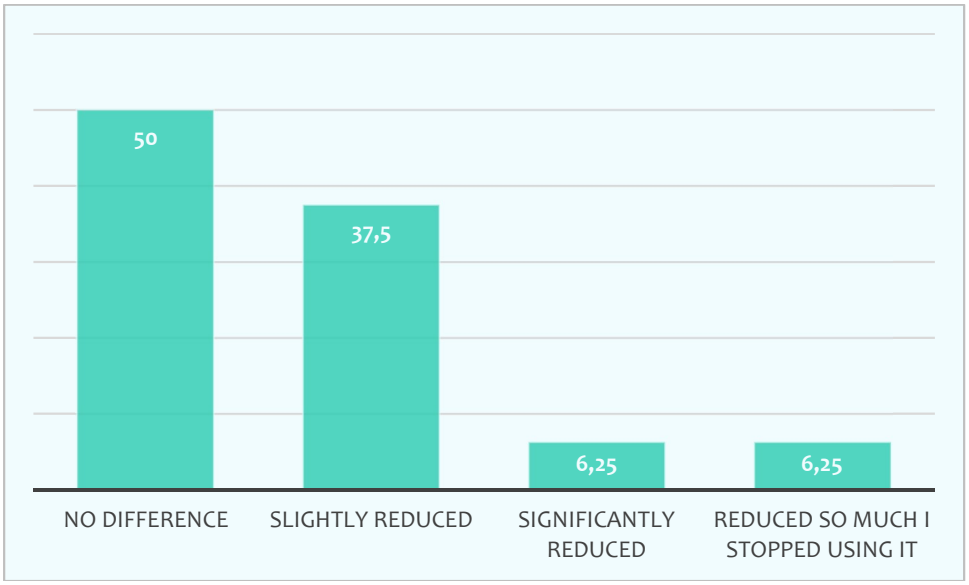

Figure 12: Effect of SELFBACK on smartphone battery life

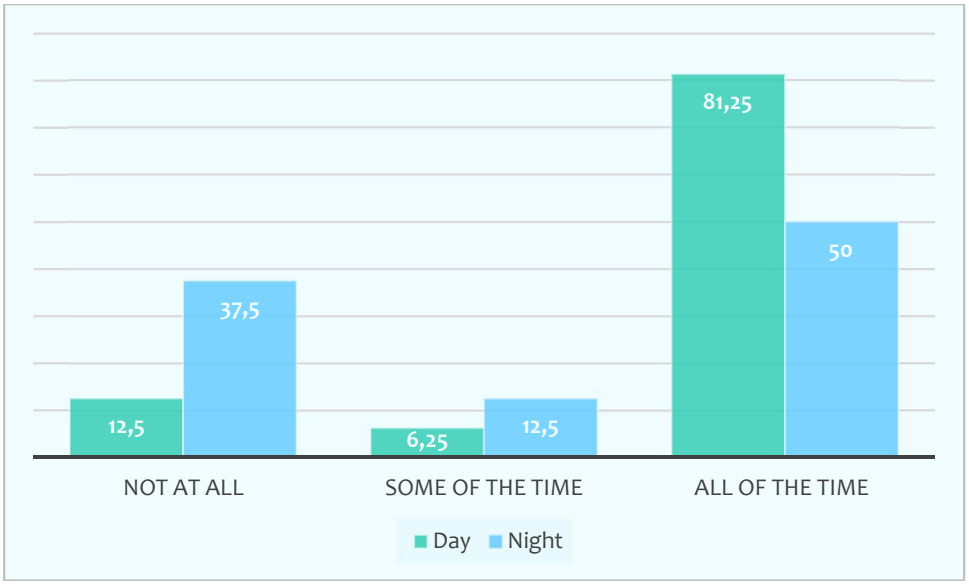

Figure 13: Participants use of the Mi-Band

Table 5: Comments on set-up & general use

Negative	Positive
Syncing device & apps was annoying App did not sync properly despite receiving technical help Having to install 3 apps felt a bit much & **SELF**BACK didn't seem to take any data from M-iFit Set up was very protracted and complicated with the downloading of the different apps to allow **SELF**BACK to work. I can see the positives in SELFBACK and would love to use the app, I liked the wrist band and the concept but found during my trial it didn't work at all Unfortunately, I had some problems setting up my band and app so that's why I have scored lower but when I finally connected, I had no problems whatsoever! I could never use the wrist band because of incompatibility of my phone operating system	Set up and sync with Mi and Fit was straight forward. General use is easy as there is not much functionality

4.3.2.2 *(2) App content*

Step-count information was reported to be useful by over 60% of participants (Figure 14). However, only 50% reported it to be accurate, with 44% reporting it as inaccurate and 6% undecided. There was less agreement on the usefulness of the goal achievement and motivational messages (Figure 15). In terms of the number of messages, 43% felt they were just right, 43% not enough, and 14% too many. Timing of the messages was considered appropriate by 60% of participants, and 80% reported that the messages were personalized. Seventy-five percent of participants did not want to be rewarded for goal achievement. It is difficult to draw conclusions from these findings; they perhaps reinforce the varied ways people like to interact with apps and the challenges involved in designing an app suitable for a range of preferences.

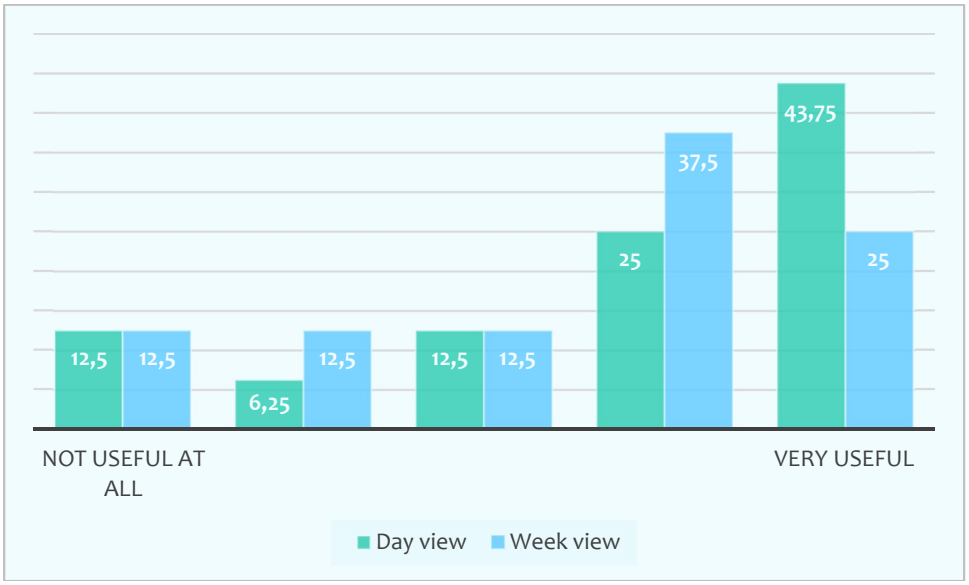

Figure 14: Usefulness of step count information (%)

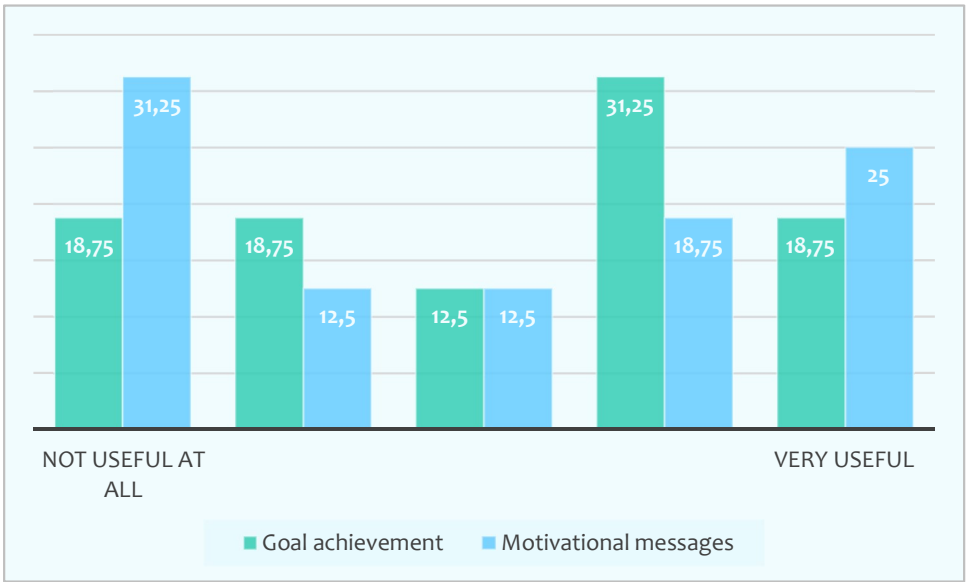

Figure 15: Usefulness of goal achievement & motivational messages (%)

Only 25% (n=4) of participants were asked by the app about barriers to physical activity; this was triggered by failing to meet step-count goals. Participants were somewhat undecided about their appropriateness and the usefulness of the information on overcoming them (figures 16-17). This perhaps reflects the immature nature of the app at the stage of testing; the education and information components have since been significantly developed. Fifty percent of participants reported that the app helped them to increase their physical activity over the 4-week period of use. A number of comments on app content were made by participants and can be seen in Table 6. Accuracy of step-counting and appropriateness of messages were the main concerns raised.

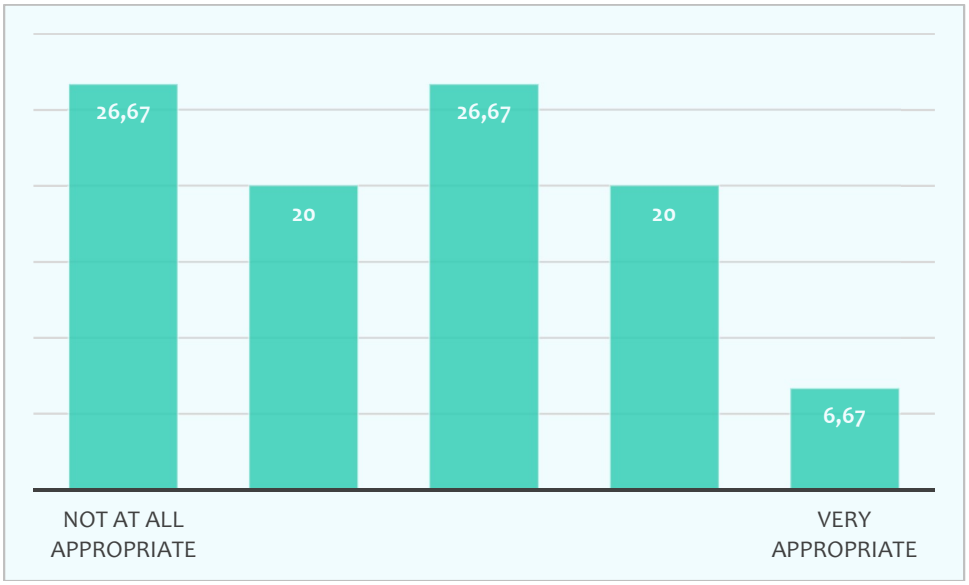

Figure 16: Appropriateness of question on barriers (%)

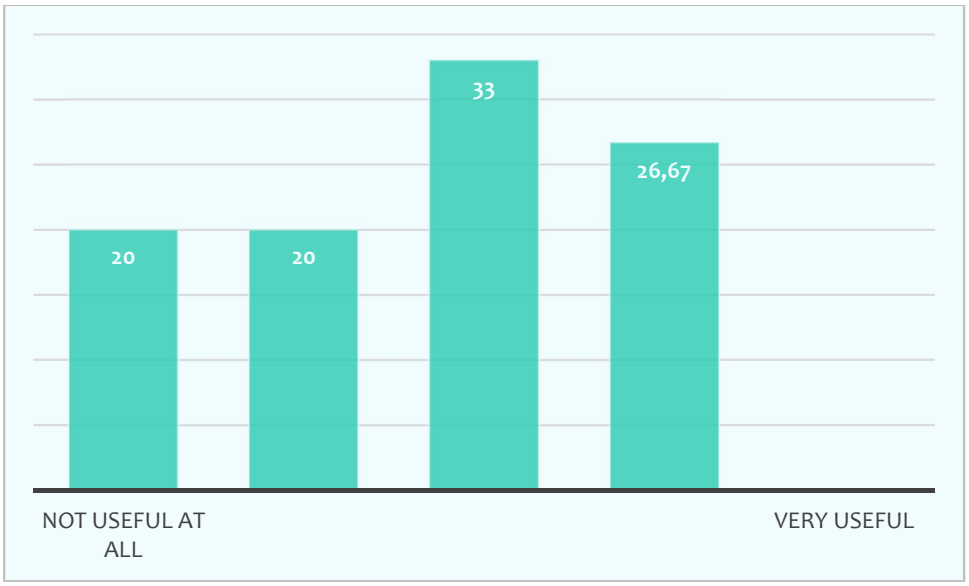

Figure 17: Usefulness of information about overcoming barriers (%)

Table 6: Overall comments on SELFBACK content

Negative	Positive
Obviously because the app didn't work properly its affected my answers as I didn't receive helpful messages or information App needs much more content especially about back pain. It needs to be a bit more specific to the type of back pain the user has and not just counting steps The app only displayed step count; Step count was often less that that shown on the MiFit app; MiFit step count was some 40% less than my Fitbit The messages were not timely and did not grab my attention. Consequently their value was minimal. Many of the messages were so generalist and simple as to be of little help. It felt like the messages were trying to replace common sense I couldn't use the wrist step count as it was not compatible with my cell phone android version therefore the step counts were far from anything accurate Grammar in motivational messages a bit hit or miss The app is quite decent but I think there are better ones available. I think the idea of increasing physical activity to reduce back pain is valid but I'm not sure it merits a specific / separate step tracker app	Reviewing goal achievement and barriers and setting new goals at the end of each week was a good idea. Simplicity & ease of use Encouraged me to complete steps Fact the band buzzed to remind me to move so often was great! Motivational messages were tailored to my own step count so well done personalization! Great concept and would love it to be more functional for folks with back pain. I think it has something very different to give and encourage movement and confidence If the user has a sedentary life style and is married to a mobile phone then the app could be useful when mature I really enjoyed this app. Unfortunately I did not take part in achieving my step goals as much as I should have due to exam revision, uni schedule, cold weather and illness but it would be a great app for the warmer months! Please add some more features to the app and I would definitely buy it!

4.3.2.3 (3) Suggestions

Participants made a number of suggestions for making the app more useful. These included:

- Providing comparisons against normal population average
- Being able to add other (non-stepping) activities e.g. swimming & yoga
- More information
- Exercises for back pain
- Healthy eating tips
- Gentle reminders to do something
- Notifying user when being inactive

Some of these suggestions are within the scope of the SELFBACK study (e.g. more information; exercises for LBP) and some will be useful to the research team for future work.

Finally, 69% of participants said they would download the app and use it long-term, and 63% would recommend it to a friend.

4.3.2.4 (4) General Questions on app use

In order to inform the future commercialization of SELFBACK, participants were asked some questions on general health app use. Forty-four percent reported that they currently used a health app. Fitness apps were most common, followed by general apps such as Apple and Samsung, Relaxation/mindfulness, and "other" which included Strava, Garmin, MyFitnessPal, Nike Health, and Pokemon Go. Reasons for liking the apps that participants used included: quality and quantity of data logging, being able to see achievements and record fitness statistics, the ability to set targets, providing a pleasurable experience, and having extra features. Participants were also asked what phrase they would type into Google to find help for back pain. Responses were varied but all included the term back pain, with the exception of one which used sciatica instead.

Just over half (56%) the participants were concerned about privacy. Their concerns related to knowing who would have their personal information, the safety of their personal information, use of their personal information for marketing purposes, and confidentiality.

Participants were asked whom they thought were trusted sources of advice regarding health apps. General Practitioners, Physiotherapists, Other patients and Scientific articles were trusted by more than half the participants. Online reviews, app stores and Facebook were conversely not trusted by more than half the participants.

4.3.3 Semi-structured Interviews

Ten participants took part in semi-structured telephone interviews. Their mean age was 51 (range 30-74), and 60% were male, therefore they reflected the sample as a whole.

Table 7 provides a summary of the categories, classes and themes that emerged during data analysis. It can be seen that the interviews reinforced the electronic survey results. Data unique to the interviews concerned barriers and facilitators to using the SELFBACK intervention. We did not detect any typologies during analysis of the interviews, e.g. male/female or younger/older opinions towards using SELFBACK. Rather, we detected a range of opinions across the sample.

4.3.3.1 Barriers

Participants suggested that older people may find it difficult to use SELFBACK. Likewise, people with disabilities may be prevented from using the app, for example:

> *"…maybe old people if they don't use a SMART phone obviously they can't benefit from this service as much…Maybe if some people have some sort of disability that they cannot walk or run from, the step count is not helpful for them."* (Participant 01)

> *"I would imagine older people would share lack of confidence with technology or just don't have a smart phone at all. So I don't know you would maybe need to have a desk top version." (Participant 14)*

Participants identified that older smartphones also posed a barrier to using the app, for example:

> [Asked about barriers to using SELFBACK] *"And all those who have other phones like myself which is not compatible with the App"* (Participant 01)

Finally, having to carry the 'phone around was seen as a potential barrier, as the notifications appear on the smartphone and not the wristband, for example:

> *"The biggest barrier is if you carry your phone all the time. If you are married to your iPhone then that probably makes a lot, although its only android specific, but what I really mean to say is if you are married to your phone like I see a lot of people then fine, if you are not like I am not then that's the biggest barrier for me."* (Participant 02)

4.3.3.2 Facilitators

Three facilitators were highlighted during the interviews. The notifications were viewed as a potential facilitator, as illustrated by these participant quotes:

> [Asked about facilitators to using SELFBACK] *"Just the motivational messages, just keeping you going."* (Participant 07)

> *"I think the physical activity monitoring, the step counting especially, if you were having a quiet day and you see your steps are low and you are being reminded through the messaging that you are well off target for the day compared to yesterday you have done X amount of steps more that is particularly motivating and encouraging you to get off the couch and perhaps take the dog for a walk, so that is good and that can encourage you to use it" (Participant 11)*

> *"Em just I think just seeing the percentage go up, it was quite nice seeing it did have the percentage on the messages, the notifications some of the times so that was quite nice to see as a percentage how much you had done and how far you had to go and that encouraged you to do a little more in the evening. To think oh well I am almost at my target I might go for a short walk to the shop or just walk around the house for a wee bit and get it up to the total." (Participant 15)*

Receiving reports on daily activity was likewise viewed as a potential facilitator:

> *"As for the App and your day to day steps and making you aware of what you have achieved each day, how much you have moved, I thought it was brilliant and I didn't think I would find that very helpful."* (Participant 12)

Finally, it was suggested that people with LBP would be more likely to use SELFBACK on recommendation from a health professional such as a General Practitioner:

> *"… it seems to me if a person went to their GP with a new case of lower back pain and the GP could refer them to this App then they might stand a better chance of using the App and the App being of some value to them." (Participant 04)*

Table 7: Summary of Framework Analysis

Themes	Classes	Categories
Technical issues	App-related technical issues Wristband-related technical issues Need for availability of technical support	Difficulty with wristband sync Restarting app frequently Over-riding other apps Downloading 3 apps Need technical support
Notifications	Notifications – Positive	Notifications motivational Notifications personalised
	Notifications – Negative	Repeat themselves after a while Could be demotivating if don't meet goals and keep getting told that Grammar "dodgy" in some notifications
	Notifications – Personalisation	Notifications tailored to activities done Don't want "silly" notifications Notifications not related to back pain
	Notifications – Frequency	Notifications – prefer 3 x day Notifications – too frequent Notifications not at right time (see them end of day) Notifications – want more frequent Notifications just right
Step-counting	Step-counting inaccuracy	Step-count inaccurate compared to Mi-fit & Fitbit
	Step – counting – positive	Making aware of activity/inactivity really useful
Goal-setting	Goal-setting – positive	Flexibility good (adjusting goal up/down) Good to get recommendation based on previous week
	Goal-setting – negative	Lack of flexibility (n=1)
Graphs/charts	Graphs/charts – negative	Not as informative as other fitness apps
	Graphs/charts – positive	Easy to understand Motivational
Suggestions for improvement	Low back-related features/exercises Notifications on wristband Ability to log other activities Inactivity reminders Interaction with other apps (e.g. Garmin, Strava, MyFitnessPal) Further personalisation	Low-back related features Notifications on wristband would help Prompt to look at phone Swimming/cycling/weight-training not recorded Need something to remind you to be active
Positive aspects of app	Improved back pain & overall health Simplicity No adverts Motivational	Improved back pain & overall health Simplicity No adverts Motivational
Barriers	Older people People with older phones People with disabilities	Older people People with older phones People with disabilities Having to carry phone around (to see messages)

	Having to carry phone around (to see messages)	
Facilitators	Notifications Report on daily activity – motivational GP recommendation	Notifications Report on daily activity – motivational GP recommendation
Wristband use	Wristband – positive	Comfortable Holds charge long time
	Wristband – negative	Clasp difficult to fasten Difficult to read in daylight Can't use with gloves

5 Informing the Pilot & RCT studies

The objectives of this feasibility study were achieved: feasibility and acceptability of the SELFBACK intervention has been explored, and required amendments identified. It is possible to recruit people with LBP, for them to complete baseline measures, and use the Mi-Fit band and SELFBACK app for a one-month period. People with LBP are also willing to complete follow-up measures and participate in semi-structured interviews. In relation to the specific objectives:

5.1 Feasibility objectives

1. To measure completion rates (time & completeness) for baseline questionnaires in a sample of people with LBP using the SELFBACK app
    a. Baseline measures can be completed in an acceptable time.
2. To explore interaction with the SELFBACK app (user activity)
    a. Participants were reasonably willing to interact with messages delivered by the app, with just under 1 in every 2 messages received being opened.
    b. Participants were overwhelmingly willing to give their opinion on messages they received with 90% of opened messages being liked
3. To explore any user-identified difficulties in engaging with the SELFBACK app
    a. Age/version of smartphone resulted in difficulties for some participants. Failure of Mi-Band to synchronize with smartphone can be a further difficulty.

5.2 Acceptability objectives:

To explore the opinions of people with LBP on:

1. the content of the SELFBACK physical activity intervention component
    a. *Content generally viewed positively. Notifications require some careful revision. Additional content planned for full version of* SELFBACK *app should be well received.*
2. the effort required in interacting with the SELFBACK physical activity intervention component
    a. *Generally viewed as acceptable*.
3. the mode of delivery of the SELFBACK intervention physical activity component
    a. *Generally viewed as acceptable.*
4. barriers and facilitators to using the SELFBACK physical activity intervention component
    a. *Barriers and facilitators were identified and will be considered (where possible) in the conduct of the pilot and RCT.*
5. the perceived usefulness and effectiveness of the SELFBACK physical activity intervention component
    a. *Generally viewed as useful and effective, with some caveats that in the main relate to adding further content (as planned).*

6. confidence that they could participate in the SELFBACK physical activity intervention longer-term
    a. *Most participants would participate in the long-term.*

5.3 And:

To determine which amendments should be made to the SELFBACK physical activity component prior to further pilot testing

The following amendments should be made to the SELFBACK app:

1. *Provide a clear explanation of the purpose of the baseline questionnaires and why there might appear to be some repetition*
2. *Determine minimum software requirements for smartphone to be eligible for study participation*
3. *Provide clear information to participants on the need for 3 apps to be downloaded to smartphone*
4. *Provide clear and detailed instructions to participants and research assistants for downloading apps and synchronizing Mi-Band to smartphone*

Review notifications – amount/type/grammar/tone – ensuring a range are available as there is no "one size fits all"

6 Empirical Evaluation of Activity Recognition Algorithms

In this section of the report, we present an overview of the human activity recognition (HAR) algorithms we have developed and the results of empirical evaluation of these algorithms. In Deliverable D2.1, we have described and presented empirical evaluation of our initial activity recognition and step counting algorithm. Since then, we have explored other important issues regarding activity recognition such as the use of deep learning approaches, personalizing activity recognition algorithms to individual users, as well as improving single sensor HAR performance using privilege learning.

6.1 Dataset

In Deliverable D2.1, we provided a detailed explanation of the SELFBACK HAR dataset. This dataset is used in all empirical evaluations.

6.2 Experimental Design

In all experiments, except where explicitly mentioned, a leave-one-person-out validation approach is used where, the data of one person is held out for testing and the remaining is data is used for training. This is repeated until every person in the dataset has been used exactly once for testing. Performance is reported using macro-averaged F1-Score.

6.3 Comparison of Shallow and Deep Features

Deep learning has recently gained a lot of attention and has demonstrated state-of-the-art performance in a variety of tasks from image recognition to language translations. Deep learning algorithms perform automatic feature extraction by identifying important patterns in the data, saving the need for manually defined features. Convolutional Neural Networks (CNNs) are particularly relevant for HAR due to their ability to model local dependencies in accelerometer data, thereby leading to better feature extraction. Accordingly, we evaluate the performance of Deep CNNs in comparison to the shallow feature approaches we presented in D2.1. The algorithms compared are:

- Time: time domain features with support vector machine (SVM) classifier
- Freq: frequency domain features with SVM classifier
- FFT: fast fourier transform features with SVM classifier
- DCT: discrete cosine transform features with SVM classifier
- CNN: Convolutional neural networks with soft-max classifier
- CNN-SVM: convolutional neural network features with SVM classifier
- CNN-kNN: convolutional neural network features with k-nearest neighbor classifier

Results are presented in Figure 18.

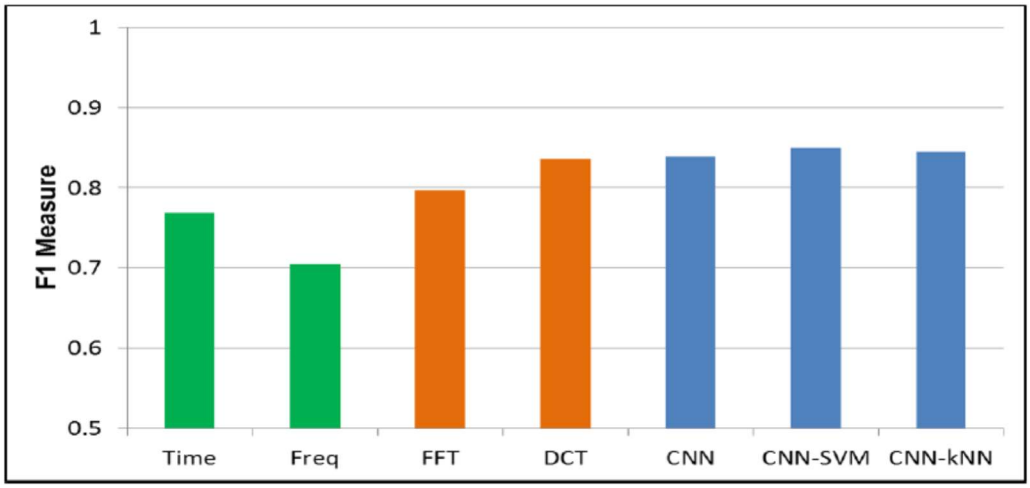

Figure 18: Comparison of Shallow and Deep Algorithms for HAR

As can be observed from Figure 18, deep learning approaches generally outperform shallow feature approaches, with the exception of DCT which is the best performing approach presented in D2.1. Our results highlight the effectiveness of applying deep CNNs for HAR, however, it also shows that DCTs remain competitive. This work has been published and further details can be found in [9].

6.4 Personalised Human Activity Recognition

Individual differences such as height, gait, posture etc., make standard HAR models trained on a general population to be less accurate compared to personalized models that are developed specifically for target individuals. However, training a personalized model for each individual user is impractical for real-world applications due to the burden this places on the user to provide sufficient data for model training. This problem motivates the need for hybrid approaches that can bootstrap a general model with minimal user data to achieve improved performance.

The issue of personalizing HAR models is particularly relevant for SELFBACK where there is a need to recognize activity intensity e.g. walking slow vs. walking fast. This level of granularity in HAR is particularly difficult to achieve using a general model as we have explained in D2.1. In the following sub-sections we present two approaches we have developed for personalized HAR.

6.4.1 CBR Approach

Although each user is unique, we can assume that similar users have similar patterns of movement. This assumption of similarity is a core principle of Case-based reasoning. Accordingly, after collecting data from a general population, we can use example data from the end user to select (sample) the most similar instances from the general population data and use these to build a personalized model for the target user. Similarity is established using a k-nearest neighbor (kNN) algorithm. Accordingly, we call our approach KnnSamp. Also, the user provided examples can be included in the model training which we call KnnSamp+. Accordingly, we compare the following algorithms:

- All Data: a baseline general model trained using all data
- KnnSamp: our proposed approach where model is trained using only instances similar to user examples
- knnSamp+: model trained using similar instances and user provided examples
- Random: random sampling of training instances

Random is included in our comparison to ensure that any performance gain by KnnSamp and knnSamp+ is not due to smaller sample of training instance. Results are presented in Figure 19.

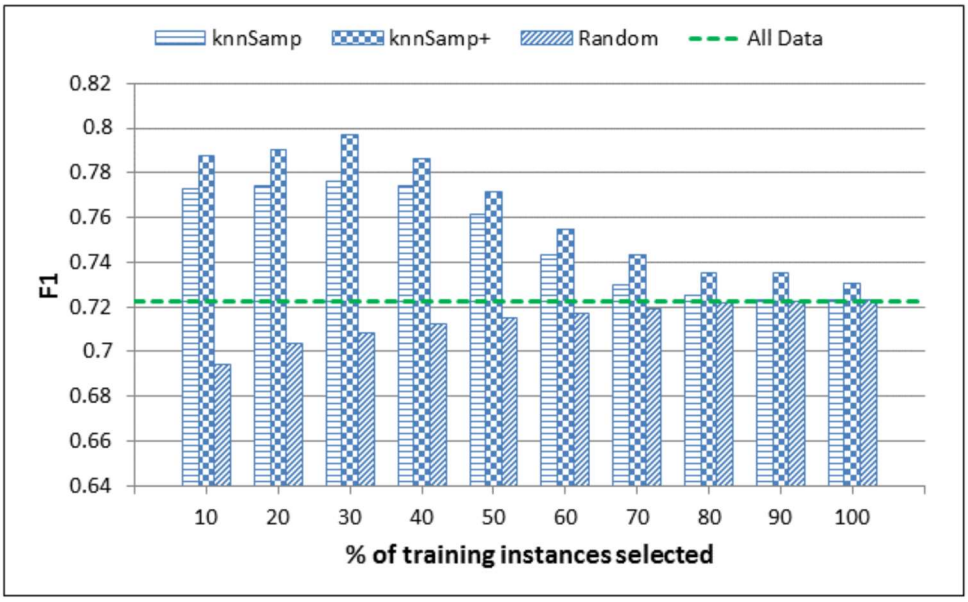

Figure 19: Results of CBR approach to personalized HAR

The vertical axis of Figure 19 is the F1 score and the horizontal axis is the percentage of training data that is sampled. The baseline (All Data) is presented as the green horizontal line. Note from the graph that our proposed approaches (KnnSamp and KnnSamp+) both outperform the baseline with as little as 10% sampled training data. The best performance is achieved by sampling 30% of training data. Random, as expected, does not outperform the baseline at any percentage of sampled data. knnSamp+ slightly out performs knnSamp which indicates that additional performance gain can be achieved by including user provided example in model training.

Overall, results show that our proposed hybrid approach does outperform a standard general model in HAR. This work has been published and further details are available in [10].

6.4.2 Deep Learning Approach

One major drawback of the CBR approach to personalized HAR we presented in the previous section is the fact that the model needs to re-trained every time the user provides new example data. This limitation can be quite serious in real-world HAR applications where models are typically deployed on small portable devices with limited processing power.

Accordingly, we developed an approach to personalized HAR that does not require model-re-training using a deep learning architecture called matching network. We refer to our approach as MNet, and further details of this work can be found in our publication [11].

We present an empirical evaluation our MNet approach by comparing the following algorithms:

- kNN: k-nearest neighbor classifier with DCT features
- SVM: support vector machines classifier with DCT features
- MNet: Matching networks with DCT features and MLP embedding
- MLP: DCT features with MLP embedding and soft-max classifier

Results are presented in Figure 20.

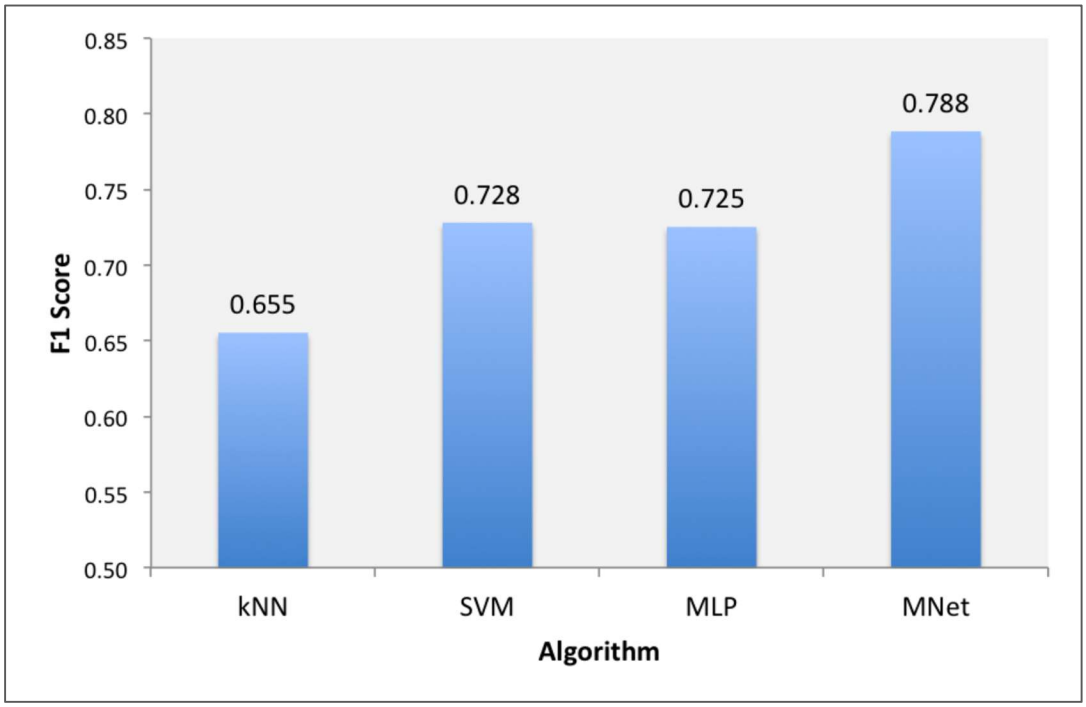

Figure 20: Result of Deep Learning approach to personalized HAR

It can be observed from Figure 20 that MNet produces the best result; whilst SVM and MLP have comparative performance but kNN comes in last. The poor performance of kNN compared to SVM and MLP is consistent with our previous evaluations in D1.2. MNet out performs both SVM and MLP by more than 6%, which shows the effectiveness of our matching network approach at exploiting personal data for activity recognition.

6.5 Improving HAR using Multiple Sensors with Privilege Learning

The SELFBACK system uses a single wrist-mounted sensor for HAR to maximize convenience to the user. However, the use of multiple sensors has been shown to provide more accurate activity recognition compared to using a single sensor. Thus, the goal of an optimal HAR solution is to utilize the fewest number of sensors at deployment, while maintaining performance levels close to that achieved using multiple sensors. This approach is often

referred to as privilege learning. To this end, we introduce a novel approach that allows us to train a model using multiple sensors while allowing the user to wear a smaller number of sensors after deployment. Our approach uses a neural network translator which takes data from available sensors and artificially generates data for the missing sensors. Further details of this work can be found in our publication [12].

6.5.1 Dataset

On a subset of 34 users in the SELFBACK dataset, we collected data from the thigh in addition to the wrist. This dataset of wrist and thigh sensors is what is used for the evaluation in this section.

6.5.2 Evaluation

The aim of this evaluation is to demonstrate how we can train a HAR model using both wrist and thigh sensor data and allow the user to only wear a wrist sensor deployment, using our neural translator to generate the missing thigh sensor data. Accordingly, we compare the following algorithms:

- Wrist: model trained and tested using only wrist sensor data
- $Wrist_{NNT}Thigh$: model trained on both wrist and thigh data and test on only wrist data, using neural translator to generate missing thigh data
- Wrist+Thigh: model trained and tested on wrist and thigh data

Wrist is the default SELFBACK scenario where only a single sensor is used. *Wrist+Thigh* is a multi-sensor scenario which we would like to avoid because it requires the user to wear a sensor both on the wrist and thigh. Results are presented in Figure 21.

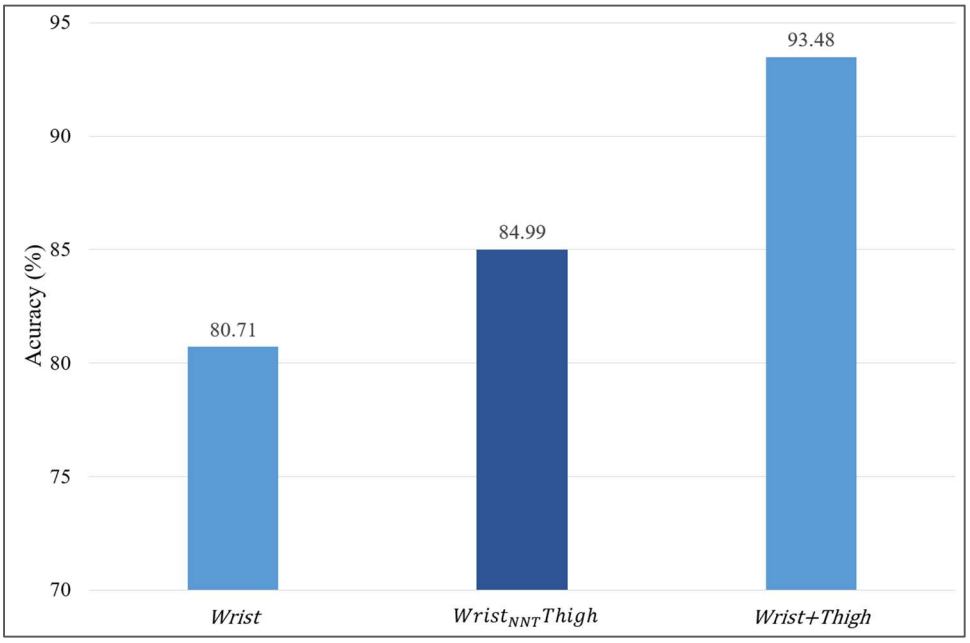

Figure 21: Results of HAR with privilege learning

As can be observed in Figure 21, our proposed approach $Wrist_{NNT}Thigh$ produces over 4% improvement in accuracy compared to *Wrist*, although it uses only wrist sensor data for testing. However, $Wrist_{NNT}Thigh$ is substantially outperformed by *Wrist+Thigh* which indicates that our neural translator is not perfect at simulating thigh data. Future improvements will investigate using deeper layers in the neural translator.

In conclusion, our results demonstrate that it is possible to improve on the performance of single sensor HAR without the inconvenience of requiring the user to wear additional sensors.

7 Conclusion

In this report, we have presented analysis of the Aberdeen feasibility study. Our findings show that it is possible to recruit people with LBP, for them to complete baseline measures, and use the Mi-Fit band and SELFBACK app for a one-month period. People with LBP are also willing to complete follow-up measures and participate in semi-structured interviews. Our findings also indicate that participants generally liked the physical activity monitoring function of the app and the notifications they received. The findings from the feasibility study are of crucial importance in preparation for the Trondheim feasibility and pilot studies of the full SELFBACK app in terms of recruitment and information given to participants. In addition, the feedback on notifications we receive in the Aberdeen feasibility study has helped to revise the design and delivery of notifications in the full SELFBACK app.

We also presented empirical evaluation of physical activity recognition algorithms that we have developed that address many important issues regarding physical activity monitoring that are relevant for SELFBACK. The issues we explored include use of deep learning approaches, personalizing activity recognition algorithms to individual users, as well as improving single sensor HAR performance using privilege learning. All the algorithms have been published and presented at conferences and list of references is provided below.

- Sani S, Massie S, Wiratunga N, & Cooper K (2017). Learning Deep and Shallow Features for Human Activity Recognition. In *International Conference on Knowledge Science, Engineering and Management* (pp. 469-482).
- Sani S, Wiratunga N, Massie S, & Cooper K (2017). kNN sampling for personalised human activity recognition. In *International conference on case-based reasoning* (pp. 330-344). Springer, Cham.
- Sani S, Wiratunga N, Massie S, & Cooper K (2018). Personalised human activity recognition using matching networks. In *International conference on case-based reasoning*. Springer, Cham.
- Wijekoon A, Wiratunga N, Sani S, Massie S, & Cooper K (2018). Improving kNN for Human Activity Recognition with Privileged Learning using Translation Models. In *International conference on case-based reasoning*. Springer, Cham.

8 References

1. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. *Developing and evaluating complex interventions: new guidance* [online]. Medical Research Council. 2008. Available from: www.mrc.ac.uk/complexinterventionsguidance [accessed 11/09/18].

2. Whitehead AL, Sully BG, Campbell MJ (2014). Pilot and feasibility studies: is there a difference from each other and from a randomised controlled trial? *Comtemporary Clinical Trials,* 38(1):130-3.

3. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, Bond CM (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLOS one,* 11(3) e0150205.

4. Sekhon M, Cartwright M, Francis JJ (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Services Research,* 17:88.

5. Creswell JW & Plano Clark VL. Designing and conducting mixed methods research. Thousand Oaks, California, US: Sage Publications; 2007.

6. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs J, Jones R et al (2014). Assessing the quality and usability of smartphone apps for pain self-management. *Pain Medicine,* 15(6), pp898-909.

7. Brooke (1996). SUS – A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland I. *Usability evaluation in industry.* Taylor & Francis, London, UK.

8. Ritchie J, Spencer L (1994). Qualitative data analysis for applied policy research. In: Bryman A, Burgess RG editors. *Analyzing Qualitative Data.* London, Routledge, 1994.

9. Sani S, Massie S, Wiratunga N, & Cooper K (2017). Learning Deep and Shallow Features for Human Activity Recognition. In *International Conference on Knowledge Science, Engineering and Management* (pp. 469-482).

10. Sani S, Wiratunga N, Massie S, & Cooper K (2017). kNN sampling for personalised human activity recognition. In *International conference on case-based reasoning* (pp. 330-344). Springer, Cham.

11. Sani S, Wiratunga N, Massie S, & Cooper K (2018). Personalised human activity recognition using matching networks. In *International conference on case-based reasoning*. Springer, Cham.

12. Wijekoon A, Wiratunga N, Sani S, Massie S, & Cooper K (2018). Improving kNN for Human Activity Recognition with Privileged Learning using Translation Models. In *International conference on case-based reasoning*. Springer, Cham.