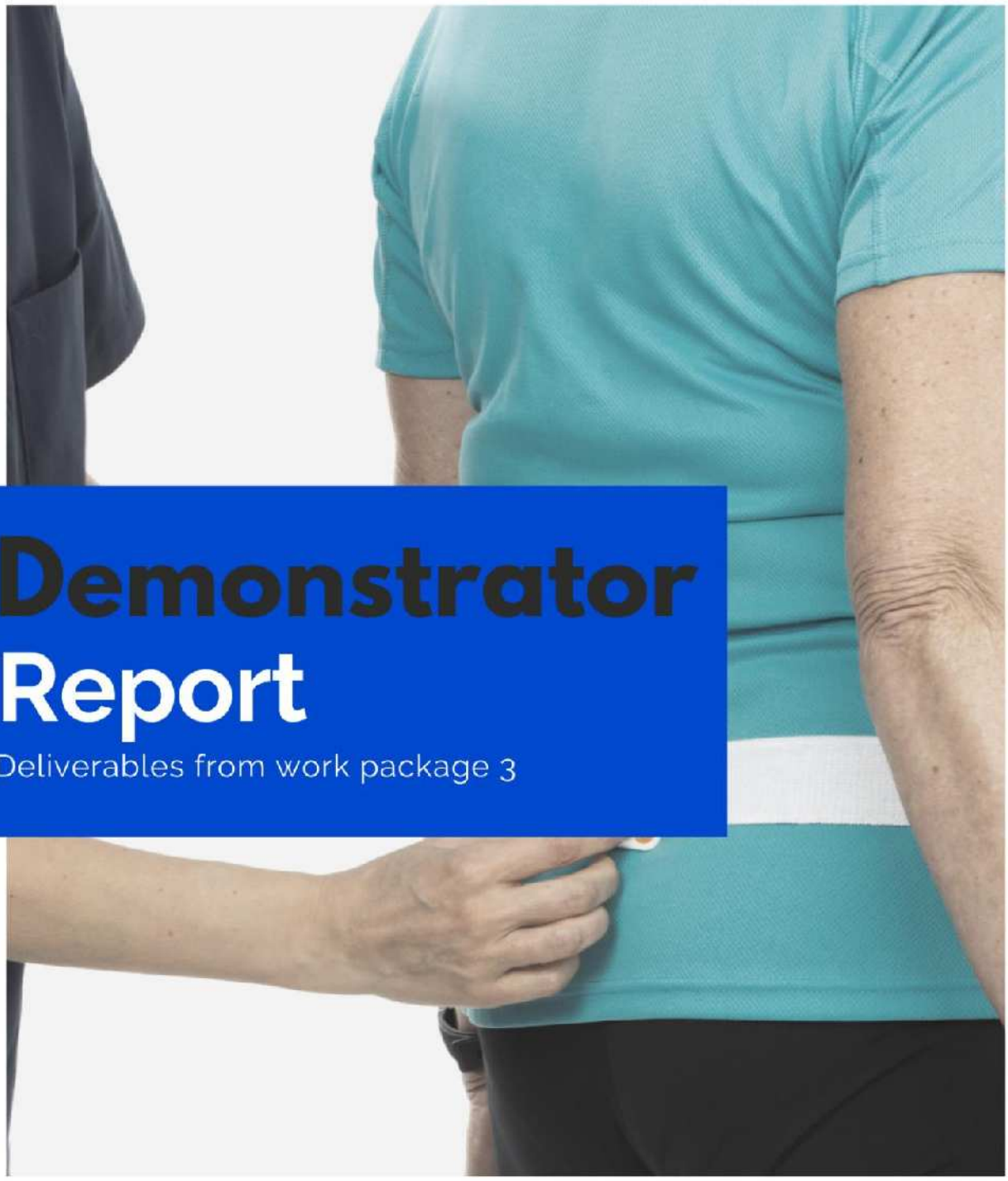


Demonstrator Report

Deliverables from work package 3

WEARABLE SENSOR TECHNOLOGIES
SENDoc

Northern Periphery and Arctic Programme
2014–2020

EUROPEAN UNION
Investing in your future
European Regional Development Fund

SUMMARY

SENDoc project started in 2017 with the aim to introduce the use of wearable sensor systems in ageing communities in northern remote areas. The concept of connected health wearable sensor systems in healthcare holds the potential of transforming the health care service but there is a need to test different solutions in real-life conditions and to bring together relevant stakeholders in larger scale projects to investigate implementation strategies. The project has tested technologies for connected health in demonstrator projects to prove feasibility and develop methodologies surrounding the technology. One specific topic of interest in SENDoc has been frailty that is related to the ageing process. Older people living with frailty are at risk of adverse outcomes such as dramatic changes in their physical and mental wellbeing after an apparently minor event that could challenge their health, such as a fall or new medication. To identify elderly at risk of frailty would be important as the knowledge that they have frailty and might be at increased risk for events such as injurious falls, can help health and social care professionals to take action to prevent the poor outcome for a particular intervention and to start a pathway of care to address the issues contributing to frailty. The demonstrators in SENDoc have focused on different aspects and indicators of frailty and has included development of a sensor-based frailty test, functional tests for frailty measurements and gait analysis for exercise plan.

TABLE OF CONTENT

INTRODUCTION

The silver tsunami is hitting the world and Europe in particular; Europe's population is ageing. Currently, an estimated 962 million people (13% of the population) worldwide are aged ≥60 years, and this group is growing at a 3% annual rate. The current economic situation has made these issues more acute. Across Europe, healthcare expenditure is expected to reach almost 16% of GDP by 2020. Coupled with a shortage of qualified personnel, European nations are facing increasing challenges in their ability to provide better integrated and sustainable health- and social services.

As a result of the age shift, in 2025, one-third of Europe's population will be 60 years or older. This "silver tsunami" with an increasing number of elderlies is a fantastic result of improved living conditions and medical treatment, but it also represents a great challenge for society. With increased age and increased incidence of non-communicable diseases, decreased ability for independent living, and potential impaired mental abilities there is a decline in functional and mental capability. As such, this demographic trend will affect the countries' economic and social development, welfare, health care, and the individual human being. There are two additional challenges in our region that attenuates the challenges associated with the silver tsunami: accessibility of services in remote areas and large regional differences in terms of access to and quality of health services. There is thus a need to develop a people-centered integrated care, as well as a need for strengthening health promotion efforts, to reduce the impact of age-related diseases. It is becoming increasingly important that the entire chain around health and care is optimally organized. The focus shifts from treatment in a care center to prevention and health promotion outside the care institute.

Though aging is not a disease per se, for most people it is a development of vulnerability, or frailty. Frailty as such is not attributable to a single factor and the pathophysiological pathways that lead to frailty are not well defined (please see Figure 1). Besides a common understanding of frailty as a concept, there is an evident need to target this vulnerable group in society and better define comprehensive assessments and subsequently multimodal rehabilitation. In fact, the European Commission (EC) has prioritized frailty within the health policy agenda of the majority of the European Union (EU) member states through its "Joint Action on Frailty Prevention"[1]

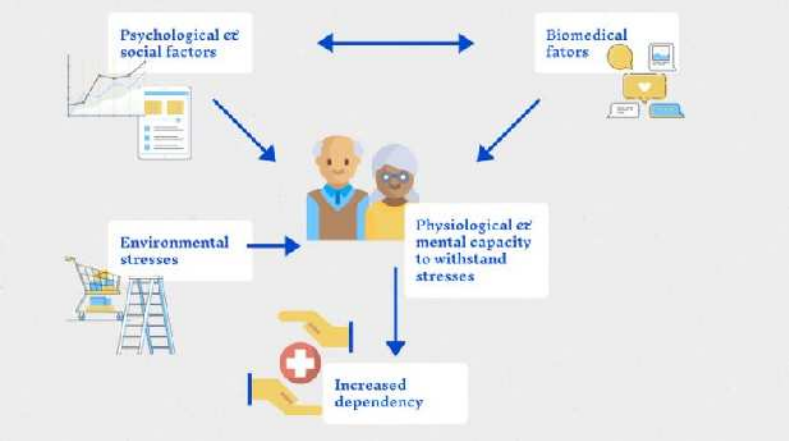


Figure 1 Factors related to frailty

Faced with our aging population, we must look for new ways to ensure that older adults can contribute to- and be cared for by- their communities. The future growth of the older population will necessitate a more integrated aging infrastructure, with increased social service, and health care options that meet the needs of both active- and frail- older adults. Thus, developing elder-friendly communities to meet the needs of todays´ older adults and prepare for elderlies' tomorrow is an issue of growing importance. Unfortunately, only a limited body of knowledge exists to assist policymakers and practitioners in creating elder-friendly social structures. However, current healthcare systems and processes are more focused on providing acute, reactive care rather than on treating and modifying the progression of these chronic conditions or their prevention. To increase the citizens access to- and value of- care, we will need to combine efforts that promote health, reduce process inefficiencies, and captures value from enabling technologies.

The focus needs to be on healthcare information and patient/citizen data, facilitating its secure movement out of system silos and transforming it into intelligence for improved patient administration and enablement of patient-centered, coordinated care. Progress must continue, but in equal measure new ways of grasping the potential of rapidly emerging technologies must be sought, such as more integrated use of electronic health records (EHRs), greater sharing of digital communication between patients and clinicians, and more prevalent use of technologies that drive remote diagnosis, treatment, care and patient education, with the objectives of achieving three critical aims:

1. Increasing citizen access to- and value of- care
2. Collaboration to improve quality, outcomes, and personalized care
3. Building of sustainable, cost-effective healthcare systems

The SENDoc project has been working with one part of this equation, namely evaluating the use of wearable sensor systems in ageing communities in northern remote areas. One of the key innovations of SENDoc has been to use sensors to measure mobility, strength, and balance, in order to support independent living in rural communities. SENDoc has focused on changing existing rehabilitation programmes, transferring research and development in wearable systems, applying connected healthcare concepts and creating platforms where data and experiences can be shared and supported, capturing data through innovative sensor wearable software systems, measuring impact on health, testing technical and social acceptability, and testing the functionality in cold climate conditions.

Frailty needs to be investigated from many different angles due to that frailty is not attributable to a single factor and the patho-physiological pathways that lead to frailty are not well defined (please see Figure 2).

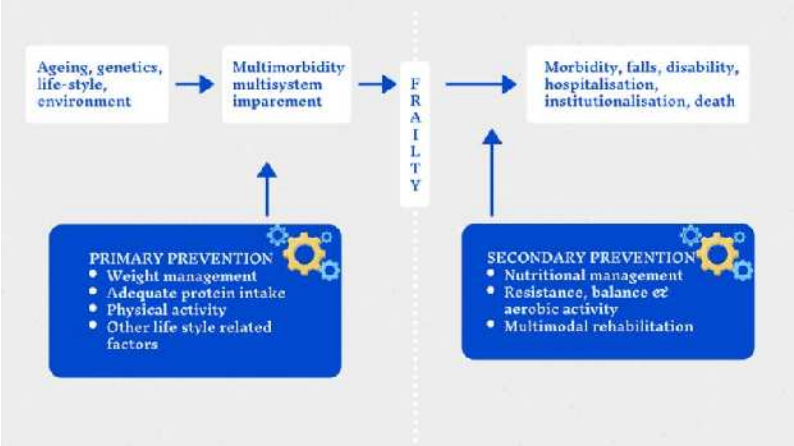


Figure 2 Assessment and multimodal rehabilitation

Within SENDoc we have performed several demonstrators to investigate how we can potentially use wearable sensors when working with elderly and the frailty concept. We have investigated stability, usability, predictability and diagnosability of different systems.

THE SENDOC DEMONSTRATORS

For the using of wearables to become common practice in health care there is a need for the systems to be durable and able to handle multiple testing in different settings. The systems also need to be easy to use and to interpret. We have performed a wide array of demonstrators. The SENDoc project has used an informed, structured process for the selection of health wearables in the context of wearables demonstrators. The process started with the identification of user needs, followed by their translation to technical specifications defined in terms of quantifiable features. These features were

defined using ideal and marginally acceptable values based on national regulations and constraints imposed by the specific context. The devices available in the market were screened and rated against these technical specifications. Then the selected devices were submitted to preliminary verification and usability tests. Finally, the devices were validated for its use under real-life conditions.

Description of test centers in SENDoc

Finland

The SENDoc demonstrators in Finland have been done in collaboration with our associate partner Siun sote - North Karelia's Joint Municipal Authority, Social and Healthcare Rehabilitation Unit. Usability evaluation procedure in real rehabilitation working life processes and rehabilitees have been invaluable. Key actors have also been Karelia University of Applied Science's service units and learning environments FysioTikka and Voimala. Karelia's Movement & functional capacity lab facilities have been basic structure of accuracy evaluations of different sensor systems. Rehabilitation with remote connections and sensor demonstrators have been conducted together with private rehabilitation companies and with Finnish NPA area companies, that manufacture and develop sensor systems and remote rehabilitation services.

Ireland

The SENDoc demonstrators in Cork, Ireland have been performed by two groups in University College Cork: Tyndall National Institute, ICT and Deep Tech experts, as well as the Centre for Gerontology and Rehabilitation in association with St Finbarr's Hospital. We have also collaborated with our SENDoc colleagues in Northern Ireland to evaluate technology from Ulster University

Northern Ireland

The SENDoc demonstrators in Northern Ireland have been performed in collaboration with the Western Health and Social Care Trust (WHSCT) Altnagelvin hospital and the Clinical Transnational Research and Innovation Centre (C-TRIC) in Derry/Londonderry. The C-TRIC is a clinical state-of-the art research facility not-for-profit personalized medicine, which connects patients directly with research opportunities, and it has partnerships between Derry City and Strabane District Council, Ulster University and the WHSCT. Also, demonstrations have taken place at the Intelligent Systems Research Centre at Ulster University, which is a state-of-the-art research environment comprised by laboratories for bio-inspired and micro engineered-, brain-computer interfacing-, computational neuroscience, cognitive robotics, and instrumentation, also including facilities for ambient intelligence and wireless sensor networks.

Sweden

The SENDoc demonstrators in Sweden have been performed in collaboration with Life Medicine, a testbed at the Northern University Hospital, Region Västerbotten and Umeå University. Life Medicine, a testbed working /aiding R&D and innovation in personalized medicine, carries out testing and development of new and innovative patient-healthcare solutions within Personalised Medicine by taking advantage of the region's expertise in innovation power, Big Data, research and susceptible regional healthcare that acts in favor of improved healthcare and care. The testbed shortens the distance from idea to innovation and is a platform for creating new commercial services and products for export. The concept consists of four parts: showroom, tech-lab, designed/simulated user environment and real user environment, digital portal, and open innovation.

Findings from demonstrators

1. Development of sensory based frailty test

1.1 Balance and the risk of falling[2]

Studies have shown that over 90% of all hip fractures are caused by falls, illuminating the need for ways to accurately predict and preferably prevent falls among the increasing number of older adults. Fall risk may be influenced by postural stability, also known as the individual´s ability to maintain the

body's center of gravity within the base of support (BOS) during dynamic or static activities3. We have previously published results revealing that impaired postural stability can increase the risk of incident falls by 75–90% among older adults in the highest quintile of postural sway, measured objectively as center of pressure (COP) oscillations when the sensorimotor system compensates for perceived shifts in the body's center of mass4,5. (However, less is clear regarding the potential interaction between COP sway and limits of stability (LOS) in relation to fall risk. LOS can be defined as to how far an individual is willing to extend their COP relative to their BOS without tripping, slipping or falling. Researchers have shown that the BOS and LOS decrease with age and in individuals suffering from neurological deficits such as Parkinson's disease and multiple sclerosis.

Instruments and measurements

The investigated sample was drawn from the Healthy Ageing Initiative (HAI) cohort study, ongoing since June 2012 in Umeå, Sweden, which invites all 70-year-old individuals residing in the Umeå municipality to participate. Eligible participants, drawn from population registers, received written information about the study and were subsequently contacted via telephone. If agreeing to participate, they arrived at a prescheduled visit where they provided written consent and commenced testing. The sample used for this study was based on the first 2396 men and women with complete data on postural stability and prospective fall assessments from the HAI cohort. Data were collected between June 2012 and December 2016. This study was permitted by the Umeå University Research Ethics Committee and followed the principles outlined in the World Medical Association's Declaration of Helsinki.

Assessment of postural sway and limits of stability

COP and LOS data were collected using a Wii Force Plate (WFP; Nintendo, Kyoto, Japan), connected via Bluetooth to a stationary computer. Data were acquired by the custom written Visual Studio software Balans-Test using WiimoteLib v1.7 (https://wiimotelib.codeplex.com/). The WFP contains four vertical force sensors, which delivers data at ~60 Hz and our software sampled it at 100 Hz after interpolation. Force signals were then exported to MATLAB R2014b (Mathworks, Inc, Natick, MA, USA), filtered with a 3rd degree Butterworth filter (10 Hz), and finally, down-sampled to 20 Hz. LOS assessment was initiated with each participant standing stationary on the WFP while maintaining eyes open (EO). They were subsequently introduced to a leaning protocol by a research nurse, who instructed the participants to lean as far as they could from the initial static position in a circular motion during 60 s. The motion direction was chosen by the participant although the research nurse instructed a clockwise direction if the participant seemed unsure on how to begin. From a monitor the participants received visual feedback on their COP excursion from the center while forming the LOS, which also provided the final representation when the trial ended. Participants were instructed to maintain a handhold on a provided bar in front of them and apply horizontal forces throughout the whole LOS-estimation procedure. As surprisingly shown in pilot tests, these forces do not influence the vertical force and the gravitational force on the force plate during LOS assessments. The general rationale behind this approach was to enable a slow-velocity leaning protocol that has been suggested for accurate, elliptical LOS estimations [6], while also reducing the participant's potential anxiety of falling when attempting to extend the LOS closer to the possible maximum.

Lastly, participants were also instructed to only use ankle joint movement when leaning. A secondary LOS measurement was performed if the participant used hip joint movement for leaning, took a step in any direction or would generally not follow the instructions provided by research nurses. With LOS assessment complete, participants removed the hands from the bar in front of them and conducted one EO trial and one eyes closed (EC) trial during 60 s each, measuring COP sway length in a bipedal quiet stance. They were instructed to avoid body movements, stand relaxed with their arms resting at their sides and maintain an upright position. Measurements were repeated in the event instructions were not followed. Given the positions of the force sensors, we calculated the COP from the force signals. We constructed convex polygons enclosing the trajectories of the COP. The polygons were

characterized by their area (centimeters squared). As indicators of the participants' behavior, we investigated Sway-Area-to-LOS (SA:LOS, 0–100), i.e., the ratio between the sway area and the LOS, as well as Minimal distance between sway area and LOS (MinD:LOS; centimeters). The direction of the MinD:LOS parameter could occur in a 360° angle since it was located wherever the LOS border and the sway area were in closest proximity to each other.

Assessment of physical activity and capacity

Participants wore triaxial accelerometers (GT3X+; Actigraph, Pensacola, FL, USA) during 7 consecutive days following the baseline measurements for physical activity (PA) assessments. Instructions provided to participants and accelerometer settings used have been previously described [7]. Total Moderate-to-Vigorous PA per day was classified as the sum of activity minutes above moderate and vigorous intensity thresholds. Wear time validation protocols defined non-wear time as ≥60 min of consecutive zero activity with a 2-minute spike tolerance. Measurements were excluded from analysis if accelerometer wear time was below 4 days with 10 h per day in total. Isometric muscle strength was examined using a hydraulic hand dynamometer (Jamar; Patterson Medical, Warrenville, IL, USA) to measure each participant's maximum grip strength in the non-dominant hand. Participants were instructed to maintain the elbow in proximity to the waist and keep the arm at a 90° angle, while compressing the handgrip dynamometer with maximal effort. The maximum value obtained in two consecutive attempts was used in further analyses. Participants also completed the Timed- Up-and-Go (TUG) test to assess gait ability, lower leg muscle strength and overall functional mobility [14].

Fall data collection

On the day of the examination participants reported retrospective falls that occurred up to 12 months prior to the examination through a questionnaire. Prospective falls were collected by contacting participants at follow-up 6 and 12 months later where a research nurse asked the question: 'During the past 6 months, have you experienced a fall at the same level?' Qualifying retrospective and prospective falls were defined as low-energy incidents where the participant came to rest on the ground unexpectedly by themselves.

Results

During follow-up, 337 out of 2,396 participants (14%) had experienced a fall. Unadjusted regression models from the EO trial revealed increased fall risk by 6% (OR 1.06, 95% CI 1.02–1.11) per each centimeter squared increase in sway area and by 16% (OR 1.16, 95% CI 1.07–1.25) per 1-unit increase in Sway-Area-to-LOS ratio. Odds ratios were generally lower when analyzing EC trials and only slightly attenuated in fully adjusted models.

Significance

Integrating postural sway and LOS parameters provides valid fall risk prediction and a holistic analysis of postural stability. Future work should establish normative values and evaluate clinical utility of these measures.

1.2 Machine learning modelling for mortality prediction in a population of older adults

Introduction

Nowadays, the investigation about mortality prediction has become crucial because, with an increasing utilization of hospital healthcare by elderly population, an immediate assessment on patient's condition has become the first step to guarantee an improvement in the long-term survival, while reducing the total healthcare demand.

Nevertheless, current clinical indexes show a number of limitations, making the implementation of new diagnostic modelling for the development of risk prediction algorithms of primary importance.
This work aimed to investigate and then demonstrate the possibility to find new solutions based on machine learning, using clinical data collected over a 5 years' span. In particular, the hypothesis of this project was that all-cause mortality could be predicted starting from data collected via activity trackers and questionnaires, simplifying the standard approaches that are mainly based on lifestyle choices, such as smoking and alcohol consumption, and health factors, such as cancer or heart disease history. Because of their low-cost and ease of use, this new study might drastically reduce the healthcare expense burden, guaranteeing at the same time accurate and objective predictions.

Dataset

The data adopted in this study were taken from the Healthy Ageing Initiative study in Umea, Sweden, concerning medical history, lab tests, physical activity, and behaviour, including alcohol- and tobacco use and mental wellbeing. It is composed of 156 parameters extracted from a 70-years old population, after a 3-hour health examination carried out on each participant, totalling 2291 elderly entries. All the measurements were repeated only once, in a period from January 2013 to December 2017, and the subject's conditions were monitored to know which patients passed away in the time between their data collection and the end of study date (31st December 2019). The study aimed to generate a dataset including various and heterogeneous features to evaluate all the possible aspects that influenced older people's daily life. Nevertheless, the obtained dataset presented one of the most common scenarios findable in clinical studies, represented by a low amount of data belonging to positive class, since only 92 (4%) of the total patients passed away during the study period. For this reason, the most challenging aspect of the work was related to properly manage the severely imbalanced dataset.

Methods

To overcome the problem, several machine learning techniques have been considered, involving data augmentation, feature engineering, over/under-sampling, probability calibration, and ensembles methodologies by means of several base classifiers.
Furthermore, since current literature showed the efficiency of ensemble model while managing highly imbalanced dataset, the final solution was found by implementing all the previous techniques in conjunction with an ensembling classifiers structure. The resulting model involved a first feature selection procedure, by means of the FSCA algorithm, followed by the removal of outliers by using Isolation Forest. Finally, ten Adaboost models were used in parallel, trained by ten different subsets of training data randomly oversampled through the Random Balance algorithm (Figure 3 and Figure 4). Each classifier was given a weight based on its accuracy and the result was computed with a soft-voting approach.

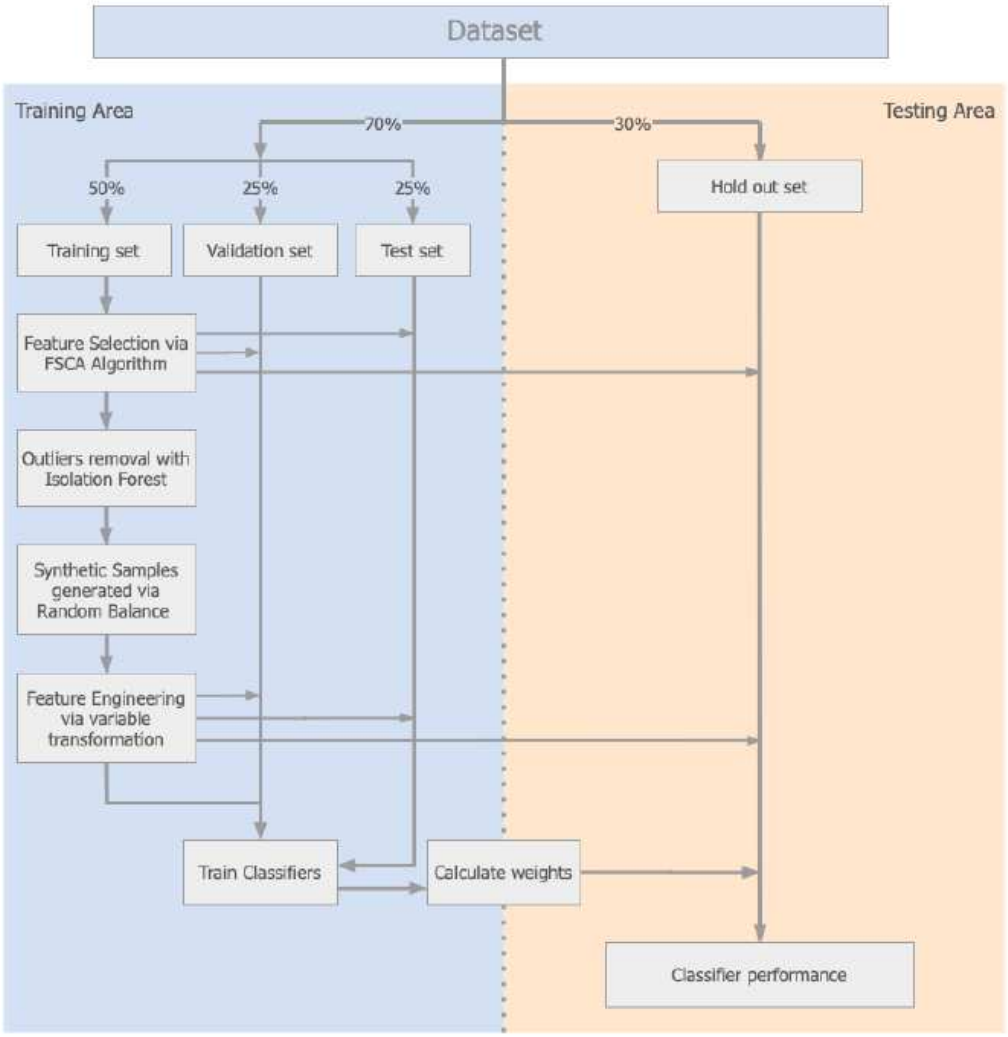

Dataset
Training Area
Testing Area
70%
30%
50%
25%
25%
Hold out set
Training set
Validation set
Test set
Feature Selection via FSCA Algorithm
Outliers removal with Isolation Forest
Synthetic Samples generated via Random Balance
Feature Engineering via variable transformation
Train Classifiers
Calculate weights
Classifier performance

Figure 3

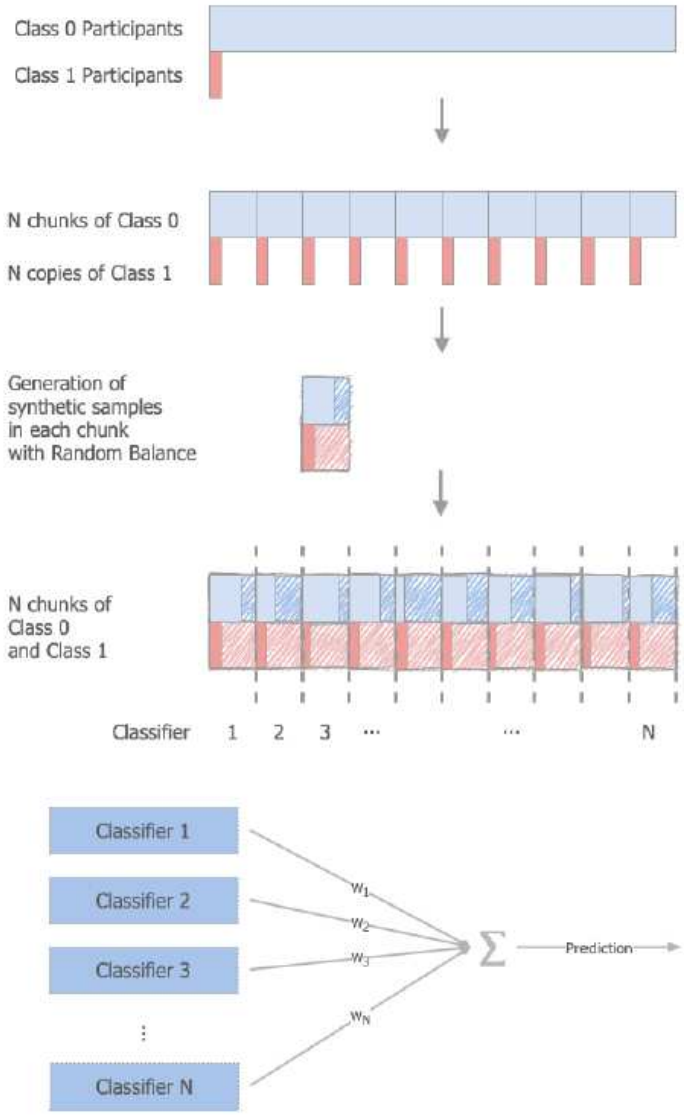


Figure 4

Results and Discussion

The achieved results identified the AdaBoost as most performant base classifier, followed by Decision Tree, Random Forest, Logistic Regression and Support Vector Machine as last.
Nevertheless, when a cost-sensitive technique was applied, the overall performance of each model was getting worse with respect to all the other scenarios, implying its inefficiency in performing the required task.
At the same time, by looking at results reached with the implementation of Monte Carlo data augmentation, the overall performance became even worse, highlighting the inadequacy of the solution for this kind of dataset. For this reason, only over sampling and probability calibration methods have been adopted in the final model, choosing as base classifier the AdaBoost model.

Finally, the achieved result with the ensemble model showed a mean AUC-ROC value of 0.880 (Table 1 and Figure 5 below), a very promising result, aligned to the state-of- the-art results in literature on the topic.

Model		Metrics						
		AUC-ROC	AUC-PR	Brier score	F1 score	Accuracy	Recall	Precision
Ensemble model	Test	0.880 (0.873 - 0.887)	0.572 (0.563 - 0.578)	0.217 (0.198 - 0.232)	0.272 (0.257 - 0.287)	0.783 (0.768 - 0.802)	0.986 (0.971 - 0.997)	0.158 (0.149 - 0.169)
	Train	0.884 (0.881 - 0.887)	0.576 (0.574 - 0.578)	0.223 (0.216 - 0.230)	0.265 (0.259 - 0.270)	0.777 (0.771 - 0.784)	1.000 (1.000 - 1.000)	0.153 (0.149 - 0.156)

Final ensemble model metrics mean and 95% confidence interval values, for both test and training sets.

Table 1

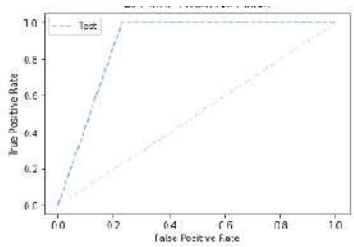


Figure 5

Moreover, a further investigation was performed by considering only subsets of features in different cases:

- Case 1: Demographic/Anthropometrics and Questionnaires.
- Case 2: Demographic/Anthropometrics, Questionnaires and Wearables data.
- Case 3: Demographic/Anthropometrics, Questionnaires, Wearables data and Lab tests.
- Case 4: Demographic/Anthropometrics, Questionnaires, Wearables data and Others.

This analysis was repeated in two cases: for all the original patients, and for only the patients passed away for cancer illness as samples of class 1. Also in this case, the results obtained were extremely interesting, reaching an AUC-ROC value of 0.882. Moreover, the scenario in which only features related to demographic, questionnaires and wearable data were used, showed a minimum loss on the AUC metrics (0.882 vs 0.857), enhancing the reliability of data from questionnaires and wearable devices to predict accurately cancer-related mortality in older adults (Tables 2 and 3 below).

Model		Metrics						
		AUC-ROC	AUC-PR	Brier score	F1 score	Accuracy	Recall	Precision
Case 0	Test	0.880 (0.873 - 0.887)	0.572 (0.563 - 0.578)	0.217 (0.198 - 0.232)	0.272 (0.257 - 0.287)	0.783 (0.768 - 0.802)	0.986 (0.971 - 0.997)	0.158 (0.149 - 0.169)
	Train	0.884 (0.881 - 0.887)	0.576 (0.574 - 0.578)	0.223 (0.216 - 0.230)	0.265 (0.259 - 0.270)	0.777 (0.771 - 0.784)	1.000 (1.000 - 1.000)	0.153 (0.149 - 0.156)
Case 1	Test	0.540 (0.504 - 0.564)	0.188 (0.128 - 0.244)	0.308 (0.255 - 0.347)	0.072 (0.052 - 0.089)	0.692 (0.650 - 0.731)	0.307 (0.207 - 0.408)	0.041 (0.029 - 0.051)
	Train	0.605 (0.587 - 0.626)	0.346 (0.307 - 0.368)	0.405 (0.344 - 0.463)	0.110 (0.103 - 0.118)	0.595 (0.537 - 0.678)	0.616 (0.534 - 0.662)	0.060 (0.056 - 0.066)
Case 2	Test	0.763 (0.647 - 0.872)	0.459 (0.346 - 0.570)	0.264 (0.219 - 0.330)	0.204 (0.146 - 0.262)	0.736 (0.671 - 0.789)	0.793 (0.582 - 0.993)	0.118 (0.086 - 0.151)
	Train	0.811 (0.741 - 0.873)	0.510 (0.451 - 0.568)	0.259 (0.215 - 0.327)	0.222 (0.184 - 0.256)	0.741 (0.675 - 0.782)	0.887 (0.773 - 0.994)	0.127 (0.106 - 0.145)
Case 3	Test	0.621 (0.517 0.725)	0.354 (0.283 0.468)	0.380 (0.318 0.438)	0.126 (0.091 0.186)	0.620 (0.567 0.688)	0.621 (0.496 0.825)	0.070 (0.049 0.107)
	Train	0.699 (0.642 - 0.774)	0.433 (0.385 - 0.502)	0.368 (0.300 - 0.422)	0.152 (0.122 - 0.205)	0.632 (0.571 - 0.704)	0.772 (0.681 - 0.887)	0.085 (0.067 - 0.117)
Case 4	Test	0.879 (0.874 - 0.884)	0.571 (0.562 - 0.578)	0.218 (0.200 - 0.231)	0.270 (0.259 - 0.283)	0.781 (0.768 - 0.801)	0.986 (0.957 - 1.000)	0.156 (0.149 - 0.165)
	Train	0.882 (0.881 - 0.884)	0.573 (0.571 - 0.576)	0.221 (0.216 - 0.227)	0.265 (0.261 - 0.269)	0.779 (0.773 - 0.784)	0.994 (0.987 - 1.000)	0.152 (0.150 - 0.155)

Ensemble model metric results and correspondent 95% confidence intervals for both test and training sets, while using different subsets of features. Case 0: All the available features; Case 1: Demographics / Anthropometrics + Questionnaires; Case 2: Demographics / Anthropometrics | Questionnaires | Wearables data; Case 3: Demographics / Anthropometrics | Questionnaires | Wearables data + Lab tests; Case 4: Demographics / Anthropometrics + Questionnaires + Wearables data + Others.

Table 2

Model		Metrics						
		AUC-ROC	AUC-PR	Brier score	F1 score	Accuracy	Recall	Precision
Case 0	Test	0.882 (0.870 - 0.896)	0.540 (0.523 - 0.550)	0.218 (0.199 - 0.234)	0.169 (0.159 - 0.186)	0.783 (0.767 - 0.803)	0.987 (0.960 - 1.000)	0.093 (0.085 - 0.102)
	Train	0.886 (0.882 - 0.889)	0.544 (0.543 - 0.546)	0.223 (0.216 - 0.231)	0.163 (0.157 - 0.166)	0.777 (0.768 - 0.783)	1.000 (1.000 - 1.000)	0.089 (0.086 - 0.091)
Case 1	Test	0.533 (0.505 - 0.567)	0.259 (0.192 - 0.312)	0.415 (0.296 - 0.510)	0.048 (0.043 - 0.058)	0.585 (0.495 - 0.677)	0.480 (0.326 - 0.600)	0.026 (0.023 - 0.030)
	Train	0.602 (0.597 - 0.606)	0.307 (0.260 - 0.351)	0.406 (0.291 - 0.497)	0.063 (0.057 - 0.071)	0.594 (0.500 - 0.681)	0.512 (0.388 - 0.606)	0.034 (0.030 - 0.038)
Case 2	Test	0.857 (0.814 - 0.887)	0.516 (0.476 - 0.544)	0.229 (0.206 - 0.252)	0.156 (0.147 - 0.172)	0.771 (0.748 - 0.794)	0.947 (0.893 - 1.000)	0.085 (0.081 - 0.092)
	Train	0.878 (0.872 - 0.883)	0.537 (0.530 - 0.543)	0.228 (0.216 - 0.243)	0.158 (0.151 - 0.166)	0.772 (0.757 - 0.784)	0.988 (0.976 - 1.000)	0.086 (0.082 - 0.090)
Case 3	Test	0.670 (0.632 - 0.713)	0.204 (0.129 - 0.273)	0.229 (0.194 - 0.265)	0.062 (0.038 - 0.076)	0.771 (0.737 - 0.802)	0.360 (0.206 - 0.474)	0.031 (0.021 - 0.041)
	Train	0.751 (0.728 - 0.763)	0.355 (0.346 - 0.366)	0.227 (0.188 - 0.267)	0.140 (0.123 - 0.158)	0.773 (0.737 - 0.810)	0.730 (0.706 - 0.753)	0.077 (0.068 - 0.089)
Case 4	Test	0.875 (0.864 0.887)	0.532 (0.519 0.545)	0.220 (0.202 0.236)	0.165 (0.156 0.178)	0.780 (0.762 0.800)	0.973 (0.946 1.000)	0.091 (0.085 0.097)
	Train	0.886 (0.883 - 0.888)	0.544 (0.543 - 0.545)	0.223 (0.217 - 0.229)	0.163 (0.159 - 0.166)	0.777 (0.772 - 0.783)	1.000 (1.000 - 1.000)	0.088 (0.087 - 0.090)

Only patients passed away because of cancer illness have been considered as class 1. Case 1: Demographics / Anthropometrics + Questionnaires; Case 2: Demographics / Anthropometrics + Questionnaires + Wearables data; Case 3: Demographics / Anthropometrics + Questionnaires + Wearables data + Lab tests; Case 4: Demographics / Anthropometrics + Questionnaires + Wearables data + Others. For both test and train sets, all the metrics mean and their 95% confidence interval are reported.

Table 3

The results showed the performance of the model and, as a consequence, the possibility to adopt the developed model on equivalent health-related scenarios (i.e. sepsis prediction, arrhythmia detection), which are plagued by the common problems associated with severely imbalanced datasets.

1.3 Measuring sedentary behaviour as a potential risk factor for depression[8]

Sedentary behaviour has previously been associated with the risk of depression. Older adults have proven to be more sedentary and more depressed than other age groups. However, there is a lack of data using objective measures of sedentary behaviour and taking physical activity into account. Thus, the purpose of this population-based demonstrator was to examine how total sedentary time and length of sedentary bouts were associated with the risk of depression among 70-year-olds using wearables.

Participants met individually with a qualified research assistant at the department in Umeå. They were interviewed about their medical history, health status and lifestyle and then performed several tests. The tests were associated with general health, including a screening instrument to identify depression. The whole procedure was around three hours. To measure sedentary behavior and physical activity, participants were sent home with an accelerometer for one week. The first measurement occasion was followed by a revisit after one week to discuss the participants individual results.

Instruments and measurements

Data from a total number of 3,633 individuals were collected and analysed. Sedentary behaviour was measured objectively with the ActiGraph GT3X+ and depression was measured with the Geriatric Depression Scale (GDS-15), several covariates such as physical activity were examined.

Geriatric Depression Scale (GDS-15)

GDS-15 is a 15-item self-assessment questionnaire developed to identify major depressive disorder in older adults[9]. In this study a Swedish GDS-15 version, which has been used in previously[10], was used. Respondents are given a score of 0-15 and a score ≥5 is a validated cut-off value of clinical depressive

symptoms[11]. The GDS-15 has shown to be valid to detect major depressive disorder and has good internal consistency with a value of Cronbach's α at 0.80[12].

Accelerometer

To measure SB and PA the accelerometer ActiGraph GT3X+ (ActiGraph Corporation, Pensacola, FL, USA; www.actigraphcorp.com) was used. The accelerometer provides objective measurements of activity, summarized into "counts" where counts per minute (CPM) are used when analysing data. The participants wore the accelerometer for four to seven continuous days. Data was stored every 60 seconds and can be categorized based on duration and intensity of the activity. Cut off points based on CPM were classified into categories according to[13]. For the analyses in this study two measures of SB were calculated from the accelerometer. One variable with percent of the day spent sedentary and one with an average length (mean) of sedentary bouts defined as periods of time in uninterrupted activity.

Covariates

Data from the accelerometer was also used as a measure of PA which was included as a covariate. Participants were divided into groups of fulfilled or not fulfilled WHO recommendations for PA (at least 150 min moderate-intensity PA or 75 min vigorous-intensity PA throughout the week; World Health Organisation, 2018). Data from Timed Up and Go (TUG), which is a measure of basic mobility skills [14], and The JAMAR Hydraulic Hand Dynamometer (Patterson Medical, Warrenville, IL, USA; www.pattersonmedical.com), which gives a measure of grip strength were used as measures of functional performance in accordance with Holmqvist et al. [15]. The occurrence of self-reported diabetes type 2 and stroke were included as covariates in the analyses, as was sex.

Findings

The purpose of this study was to examine if sedentary behavior was associated with depression among elderly using wearables. To investigate this, two measures of sedentary behavior were used from accelerometer hardware and the results indicated that both higher percent of the day spent sedentary and longer average length of sedentary bouts increased the risk of being depressed, also when adjusting for several covariates such as physical activity. Sedentary behavior is thus a potential risk factor for depression among older adults. These findings were strengthened by the adjustment for physical activity and the use of an objective measure in the form of data collected with accelerometer on a large sample. Using wearable accelerometer technology, new information about the risks associated with increased length of sedentary bouts was provided. To our knowledge, this is the first time on this subject that use an objective measure of sedentary behavior and physical activity on a population-based sample of this size.

Data from a total number of 3,633 individuals were collected and analysed from the Healthy Ageing Initiative, an ongoing cross-sectional research project in Umeå. Sedentary behaviour was measured objectively with the ActiGraph GT3X+ and depression was measured with the Geriatric Depression Scale (GDS-15), several covariates such as physical activity were examined.

Results

Results from two hierarchical logistic regressions showed that both higher percent of the day spent sedentary and longer average length of sedentary bouts increased the risk of being depressed.

Conclusions

The present study verifies previous findings about the relationship between sedentary behaviour and depression and provides new information about the risks associated with increased length of sedentary bouts. These findings may be important to consider in future recommendations for preventing depression among older adults.

1.4 Recoverability of gait analysis and effects of day rehabilitation to fall risk of older persons

Introduction

Research was conducted in seven Siun sote Older person's day rehabilitation groups between October 2019 and January 2020. In practice measures was done in the beginning and the end of period of the rehabilitation with GWalk sensor system. The starting and ending values of gait parameters of higher fall risk and other tests of functional capacity (SPPB, FRAT, MMSE) and connections to life quality (WHOQOL-BREF, RAND-36) were compared. Statistical analysis was done with SPSS- program. Researcher interviewed the day rehabilitation group instructors, about their experiences of recoverability of gait analysis in rehabilitation process.

Results

Participants walked 20 m distance with their normal walking speed. RAND -36 Health survey (Table 4) was also conducted at the beginning and the end. Collected was additional information from Siun sote: SPPB (short physical performance battery) from 30 participants and FRAT (fall risk assessment tool) from 12 participants. 59 older people participated, and they were divided into four different groups according to activity:

1.) 1 x week, no strength training in gym, duration 8-9 weeks, no physiotherapist (N=12)
2.) 1 x week, strength training in gym, duration 7 weeks, physiotherapist (N=9)
3.) 2 x month, no strength training, duration 8-10 weeks, physiotherapist (N=16)
4.) 2 x week, strength training in gym, 6-7 weeks, physiotherapist (N=14)

Mean age 82 years (69-94 years old), 44 (74,6%) were women and 15 were men (25,4%)

	1	2	1	2	1	2	1	2
Walking speed (m/s)	0.72	0.73	0.66	0.81	1.03	1.05	0.72	0.88
Cadence (steps/ min)	93.8	97.2	95.1	101.0	107.7	107.5	92.4	102.6
Symmetry (%)	89.3	89.8	86.3	87.3	93.6	95.0	94.2	92.7
Propulsion (right)	4.1	4.4	3.6	4.6	6.4	7.0	4.2	5.1
Propulsion (left)	4.1	4.2	3.7	4.9	7.0	6.8	4.5	5.3
Physical Functioning	45.25	45.75	41.67	38.33	70.63	66.56	38.57	50.71
Energy/ Fatigue	57.75	58.25	48.89	57.22	71.56	67.81	54.29	66.43
General health perceptions	47.75	47.25	41.67	43.89	54.38	55.00	43.93	48.93

Emotional Well-Being	77.20	77.00	79.11	86.22	82.25	79.00	76.86	80.57
Social functioning	73.78	78.13	75.00	81.95	86.72	87.20	74.11	93.75
Bodily pain	57.25	56.25	53.33	61.11	72.50	70.78	68.93	71.96

Table 4

Statistically significant (Wilcoxon) results in Group 4 when comparing initial and final measurements:
Group 4: 2x week, strength training in gym, 6-7 weeks, physiotherapist (N=14)
RAND -36 Health Survey:

- Physical functioning, Energy/Fatigue, Social functioning

G-WALK :

- Walking speed (m/s)
- Cadence (steps/min)
- Propulsion right and left

QUALITATIVE DATA: How Day Rehabilitation instructors and physiotherapists experienced the use of the G-Walk sensor system during Day rehabilitation periods?

- Four physiotherapists and one Day Rehabilitation instructor followed the measurement situations (initial and end measurements) with G-Walk sensor system with the participants
- After this they were interviewed about how they felt and how they experienced using this device in Day Rehabilitation periods
- The interviews were face-to-face unstructured interviews. Interviews were strongly participant-led and interviewer asked supplementary questions based on what the interviewee said (Braun & Clarke [16], 78).
- Interviews will be transcribed and analyzed in summer 2020. Inductive thematic analysis is used in the content analysis. The aim is to identify themes and patterns of meaning across the dataset in relation to the research question. In inductive thematic analysis the idea is to generate an analysis from the "bottom up" from the data, with no pre-existing theory, however, considering the researcher´s standpoint, disciplinary knowledge, and epistemology. (Braun & Clarke 2013, 175.)

1.5 Predicting fall risks in elderly from accelerometer data using neural networks

Introduction

Fractures, especially hip fractures in the elderly population, is a big public health issue, which is predicted to worsen in the next couple of years in Sweden and many other nations with an ageing population. Fractures result in increased morbidity, mortality, and health expenditures[17].
As 90% of all fractures are a direct result of falls[18], we would like to have an easy and practical way to identify an elderly person at risk of falling before a sad outcome.

Dataset

The investigated sample was drawn from the Healthy Ageing Initiative (HAI) cohort study, ongoing since June 2012 in Umeå, Sweden, which invites all 70-year-old individuals residing in the Umeå

municipality to participate. Eligible participants, drawn from population registers, received written information about the study and were subsequently contacted via telephone. If agreeing to participate, they arrived at a prescheduled visit where they provided written consent and commenced testing. The sample used for this study was based on the 1705 men and women with complete accelerometer, lab-based gait analysis data and prospective fall assessments from the HAI cohort. Data were collected between June 2012 and December 2016. This study was permitted by the Umeå University Research Ethics Committee and followed the principles outlined in the World Medical Association's Declaration of Helsinki.

Limitations of Existing Fall Prediction Work

Preliminary work carried an analysis of the literature and identified three key issues that future research should focus on to advance the research area of Sensor based Falls Risk Testing (SFRT). These issues relate to:

1) fall classification criteria
2) data acquisition methodology
3) validation protocols used

Each issue was raised in at least two recent review articles.

1) Classification Criteria:

The classification criterion is the baseline measure used to compare a proposed fall risk measurement technique against. It is the basis of all performance metrics used to evaluate the fall risk assessment tool and thus should be carefully considered when designing a fall risk assessment study.

The method used to classify fall risk differs among the literature. Participants are commonly classified as fallers or non-fallers based on one of several different methods including Clinical Assessment, Prospective Falls and Retrospective Fall.

- Clinical Assessment: A person is classified as a faller or non-faller based on their performance of an assessment, or set of assessments, in clinical settings (e.g., Timed Up and Go test).
- Retrospective Falls: A person is classified as a faller or non-faller based on self-reported fall history denoting the presence or absence of fall occurrences in the past.
- Prospective Falls: A person is classified as a faller or non-faller based on self-reported fall occurrence within a follow-up period from the assessment (commonly one year).

It is accepted in the literature that clinical assessment and retrospective falls only act as proxy measure for prospective falls, which is defined as the gold standard measure. However, in the literature review, we identified a total of only 15 papers which use prospective falls as the criterion measure [17 18 19 20 21 22 23 24 25 26 27 28 29 30 31].

2) Free Living based Data Acquisition:

Performance of standardised gait/balance tests in a lab or clinic setting is the most used approach in the literature for SFRT. However, this approach has several disadvantages which we discuss below.

One issue relates to the participants' awareness of being observed (Hawthorne Effect) during performance of gait and balance tests in controlled settings. Research has shown that gait performance can differ when participants are being observed. Thus, gait-based measures in lab/clinic settings may not reflect naturalistic behaviour. Another issue with controlled lab-based measures is the cost and time associated with administering the test. Lab based assessments require at least one expertly trained health professional to set up the equipment and administer the test. Additionally, participants are required to physically attend the lab. As previously discussed, for fall risk assessments

to be adapted in clinical practice it is vital that they are easy and cost effective to administer. Current lab-based approaches are therefore rarely adapted for regular fall risk screening.

As an alternative to acquiring data under controlled lab-based conditions, gait-based measures can be extracted from free living behaviour where sensors are worn by participants in their natural daily living environments. There is clearly a significant benefit to this method of data collection as participants are not required to attend specialist clinics or lab settings and participants do not need to be monitored or supervised by clinicians or other health care professionals. Furthermore, behaviour measured in free living conditions should be considered more representative of natural behaviour compared to measures taken in lab-based settings.

In the 15 previously identified works on prospective falls, 2 were based on prospective falls.

3 Problematic Machine Learning Methodology

Recent research has highlighted some troubling trends in relation to the presentation of over-optimistic SRFT results. Concerns have been raised in relation to sample size, questionable modeling and problematic validation methodologies.

One if the biggest challenges in SFRT is acquiring a large enough sample size to ensure sufficient study power. This is particularly challenging for prospective based falls studies where there is a requirement for a 6-12 month follow up with each participant. However, from the set of 15 reviewed prospective falls papers, we found that the largest number of participants used was only 319 and the average number of participants used was 127 (+/- 86). Most studies have therefore been too small to gain any real statistical insight into the effectiveness of techniques if applied to a larger population.

Another issue relates to misuse of model validation methodologies in SFRT studies. In particular, a common problem found was that testing data is used in some of the model training pipeline steps, such as feature selection, model selection, parameter tuning. These practices lead to models that are biased towards the available data and thus can produce models that are over-trained and produce inflated accuracy scores that are unlikely to maintain their reported performance during real world use. To build a model that is not biased towards the available data, it is vital that testing data should never be used to inform the feature selection, model selection or parameter tuning. For example, learning which features to select should be conducted on training data only. Learned features can then be selected from training and test set. This holds true for K-fold cross validation (CV) also, where K train-test sets are used. In this case, feature selection, model selection and parameter tuning should be performed independently on the training data for each the K splits.

Research also highlights issues with the use of CV as the primary mechanism to evaluate prediction techniques. Specifically, there is no way to know how reliable a CV based performance measures is. Furthermore, CV has been shown to have a large uncertainty for small sample sizes common in health and medical research where sample sizes are in the hundreds. It is therefore advised, particularly for small datasets with N < 1000, that performance be reported using a single holdout test set.

Of the previously identified 15 works, only 1 of these use an external holdout test set.

Aim

The aim of this study is to address these 3 key issues and develop machine learning models to predict prospective falls using free living-based accelerometer data.

Data Acquisition

1,705 participants, all 70 years old, took part in the study (817 female and 888 male). All participants attended an examination session conducted by a research nurse. During the examination session,

participants were asked to perform a set of standardized functional tests; Timed Up and Go (TUG), Gait Velocity during a 6 Meter Walk Test and Non-Dominant Hand Grip Strength. Participants also performed a gait assessment on a pressure sensitive walkway (GAITRite, CIR Systems Inc, USA). Participants were asked to walk the length of the walkway (6 Meters) and 65 gait based measures were calculated for each participant using proprietary software (GAITRite). Gait measures include Step Time, Cycle Time, Step Length, Heel to Heel Base Support Distance, Single Leg Support Time, Double Leg Support Time, Swing Time, Stance Time, Step Extremity Ratio and Toe In/out angle.

History of falls in the 12 months prior to the study commencing was also recorded based on participant recall. A fall was defined as an event which results in a person coming to rest inadvertently on the ground or floor or other lower level.

After the examination session, participants were provided with a hip mounted tri-axial accelerometer (Actigraph GT9X, Actigraph LLC, USA) which they were asked to wear for 7 consecutive days. Acceleration for x, y and z axis were recorded for the duration of the 7 days at 30Hz. At the end of the 7 days, the device was returned by the participant and acceleration data was retrieved for each participant (average of 911MB per participant).

Six and twelve months after the examination session, follow-up telephone interviews were conducted to ask whether participants have experienced a fall since their examination session. Prospective follow up phone calls identified 255 participants reporting at least one fall within 12 months after the examination session with 16.4% (n=134) of females and 13.6% (n=121) of males reporting a fall. 525 participants reported having a fall prior to the study commencing. Of the 525 participants reporting a fall in the 12 months prior to the study, only 19% (n=98) of those reported a fall in the 12 month follow up period. The remaining (n=151) prospective falls were from participants who did not report a fall in the 12 months prior to the study commencing.

Methods

This work aims to calculate gait quality-based features from accelerometer data retrieved from the tri-axial accelerometer worn by participants to predict fall risk. Prior to extracting gait-based measures from free living-based data, automatic detection of periods of ambulatory activity is first performed.

Singal Processing

With the accelerometer worn around the waist, the aim was that participants would wear the device such that the accelerometer axis aligned with anatomical axis such that accelerometer x, y and z axis aligned with sagittal, longitudinal, and frontal axis respectively. It was observed that participants consistently aligned the accelerometer y axis with the anatomical longitudinal axis due to the constraints imposed by the sensor belt mounting mechanism. However, due to the potential to mount the sensor at any position on the waist between the left hip and right hip, alignment of the sensor x and z axis with the anatomical frontal and sagittal axis was performed inconsistently by participants. The x and z axis were therefore combined into a single horizontal acceleration magnitude A_{horiz} to ensure measurements were consistent across all participants. Overall acceleration magnitude is also calculated, defined as A_{mag}.

Ambulatory Activity Detection

Prior to extracting gait features, periods of ambulatory activity were automatically identified from a filtered accelerometer signal. Candidate steps were first identified by performing peak detection on a vertical acceleration signal $A_{y}^{'}$ filtered using a 4th order butterworth bandpass filter (0.25 to 2.5 Hz). All identified peaks were evaluated and defined as a candidate step only if the signal had a zero-crossing and crossed both a positive and negative threshold of ± 0.8 on either side of the zero-crossing. All candidate steps were then grouped into clusters based on temporal proximity to one another.

Clusters of candidate steps that met a set of pre-defined criteria were classified as periods of ambulatory activity. While it is important that many periods of ambulatory activity are correctly identified (i.e., true-positives), it is more important that other periods of activity, that are not ambulatory in nature, are not included for further analysis (i.e., false-positives). The criterion was therefore designed to detect steady state ambulatory activity to minimize the number of false positive periods of ambulatory activity detected. The criteria implemented is defined as follows:

- Time between each step peak should be between 0.2 and 3 seconds
- Standard Deviation of time between all step peaks should be within +/- 0.8 seconds
- There should be a minimum of 25 steps within the cluster.

Feature Extraction

For each period of detected ambulatory activity, raw accelerometer data for that period was processed and features were extracted. A 4th order butterworth bandpass filter (0.05 to 3.0 Hz) was applied to the three raw accelerometer signals, A_y, A_{horiz} and A_{mag}. A single feature vector F was computed for each participant from the median of all feature vectors f_i where f_i is the feature vector computed from the i^{th} bout of detected ambulatory activity. Each feature vector consists of a step based feature vector s_i and a frequency based feature vector p_i such that $f_i = \{s_i, p_i\}$.

The step-based feature vector s_i is comprised of a set of statistical based features computed for each filtered signal (A_y, A_{horiz}, A_{mag}). Step features include maximum, minimum, index of maximum, index of minimum, variance, signal mass center, number of peaks, absolute energy, auto-correlation mean and sample entropy. Step features are computed for each step, within the i^{th} bout of ambulatory activity, between the local minima preceding and succeeding the step peak. This was performed for all peaks within the i^{th} bout of ambulatory activity and the overall step feature s_i was calculated as the median of all step features within the i^{th} bout of ambulatory activity.

The frequency-based feature vector p_i is comprised of FFT and Wavelet coefficients which are calculated from a fixed size 10 second window within the bout of ambulatory activity. For bouts of ambulatory activity that last longer than 10 seconds, a sliding window approach was used to find a 10 second period that minimizes the standard deviation of time between step peaks. Thus, the goal was to identify a 10 second period of ambulatory activity with the most consistent step cadence to calculate FFT and Wavelet based features from. Figure 6 provides a visualisation of candidate steps and how periods of ambulatory activity are identified. Figure 6 also shows an example of a 10 second window being select for FFT and Wavelet coefficients to be extracted from.

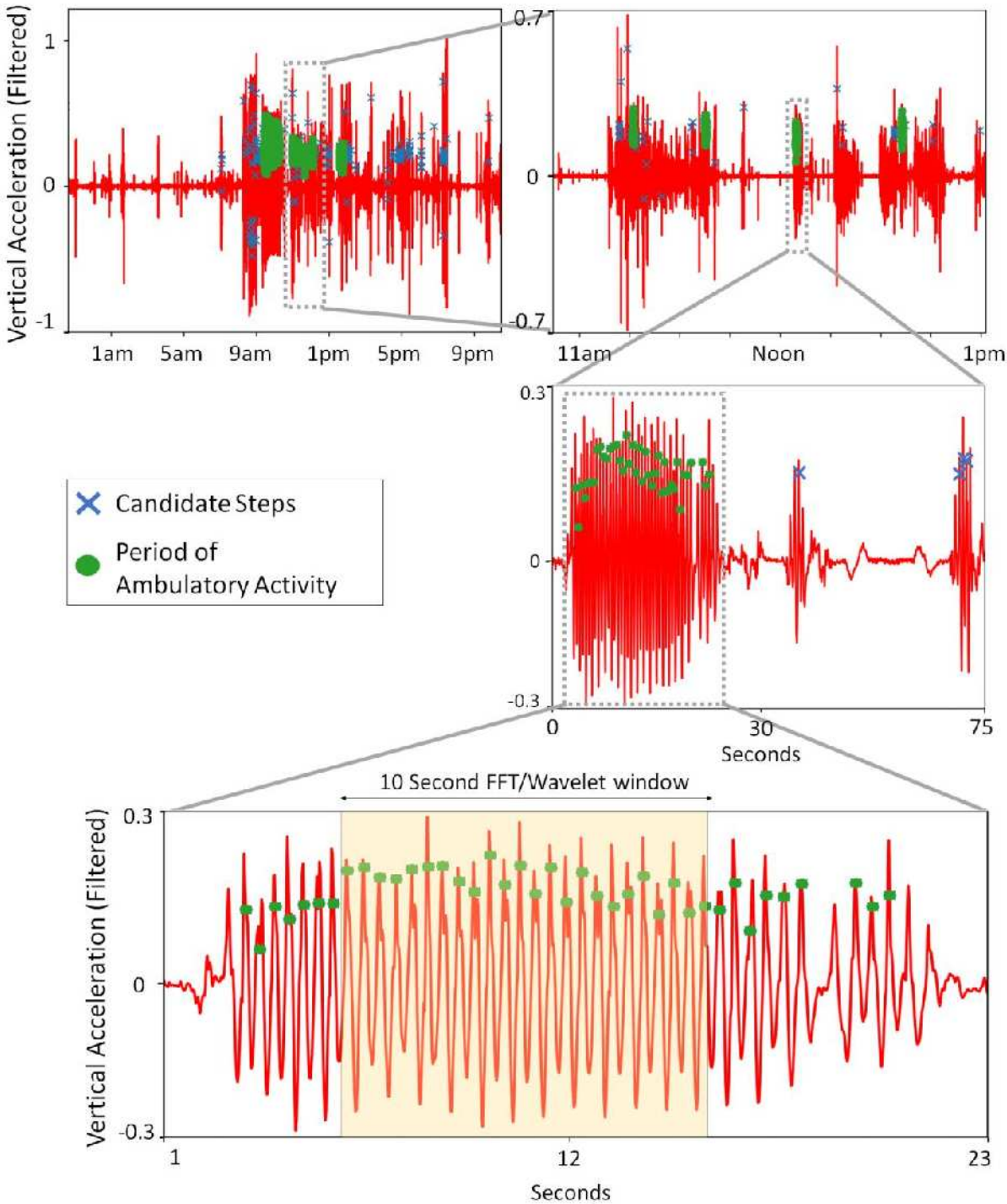


Figure 6: Filtered Accelerometer Signal for 1 Day (with multiple zoom levels) showing detection of candidate steps and periods of ambulatory activity

Model Training

As previously discussed, model training should be considered as a pipeline of steps comprising feature pre-processing, feature selection, training, parameter tuning and model selection. Model training, and any of the individual pipeline steps, will only be performed on the training set. This work will evaluate the use of 3 different data types to predict prospective falls:

- (FT) Functional test scores from Grip Strength, TUG tests, Gait Velocity
- (PSW) Gait measures from supervised lab-based pressure sensitive walkway
- (FLA) Proposed gait measures from free living accelerometer data.

Three categories of prediction models will therefore be trained and evaluated to directly compare the prediction performance of the 3 different data types. Fall History, defined as a fall occurring in the 12 months prior to the study, will also be assessed as predictor of future falls and as a feature to complement each of the three data types.

The three data types (FT, PSW, FLA) are available for all 1,705 participants. A holdout test set of participants, stratified for occurrences of prospective falls, was created using 25% of the data (N=428, 64 Fallers). A training set of participants was also created using remaining participants not in the holdout test set (N=1279, 191 Fallers). Data type specific test and training sets were then derived from the set of test and training participants.

The scope of this paper is not to develop or propose novel machine learning models or configurations. The aim is to evaluate the proposed FLA features and compare them to more standard FT and PSW features, in predicting prospective falls. To build predictive models, consideration needs to be given to the type of feature prepossessing and learning algorithms that will be used. In addition, appropriate hyper-parameters need to be set for the chosen algorithms. To remove any potential bias that could be introduced, by choosing prepossessing techniques, learning algorithms or hyperparameters that favours one of the 3 data types over another, a systematic and objective methodology to select algorithms and hyperparameters is implemented. This method uses an automated machine learning methodology based on Bayesian optimization methods to select from 15 classifiers and 14 feature preprocessing algorithms. The final prediction model, including algorithms, models, features and hyperparameters, is selected based on performance calculated from 10-fold cross validation on the training set. Overall model performance of the final prediction model is evaluated using the holdout test.

Results

In this work, 3 different data types were used to train and test machine learning models for fall risk prediction as previously described. For each data type, an automated Bayesian Optimization based machine learning methodology was utilized to configure a machine learning pipeline and train models. Models were configured and trained using training set data only. The training set was split into a cross-validation set (N=1087, 163 Fallers) and a validation set (N=192, 28 Fallers). The Bayesian Optimization system was implemented to maximize the average g-mean, over 10 folds, computed from the cross-validation set. Using early stopping, the validation set was utilized to reduce the models over-fitting on the cross validation set. After each epoch of the Bayesian Optimization process, g-mean was calculated for the cross validation set and the validation set. Early stopping of the Bayesian Optimization process was performed when validation set performance began to diverge from cross validation set performance.

After model configuration and training was completed for each of experiment data type condition, testing was performed on each model using the holdout test set (N=428, 64 Fallers). Experiments were performed to compare retrospective and prospective fall classification accuracy using different combinations of FLA, FT and PSW data types. Experiments included independent assessment of FLA, FT and PSW data types while also evaluating potential complementary information provided by combining data types. Feature combinations FLA+FT, FLA+PSW, FT+PSW and FLA+PT+PSW were therefore evaluated to assess if retrospective and/or prospective fall classification improved when using combined data types.

Tables 5 and 6 show fall classification performance scores for retrospective and prospective fall respectively for the different data types. Sensitivity and Specificity scores are shown, in both tables, for the cross validation set and the holdout test set when early stopping was and was not implemented.

	No Early Stopping				With Early Stopping			
	CV		Test Set		CV		Test Set	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
FLA	0.65	0.63	0.52	0.58	**0.61**	**0.67**	**0.61**	**0.68**
PSW	0.59	0.62	0.50	0.56	0.58	0.61	0.57	0.60
FT	0.60	0.58	0.58	0.6	0.59	0.60	0.60	0.61
FLA + PSW	0.61	0.67	0.55	0.63	0.64	0.6	0.64	0.58
FLA + FT	0.59	0.70	0.52	0.65	0.6	0.66	0.58	0.68
PSW + FT	0.60	0.58	0.58	0.63	0.60	0.60	0.60	0.61
FLA + FT + PSW	0.67	0.64	0.50	0.60	0.60	0.63	0.59	0.65

Table 5: Performance on Retrospective Falls

	No Early Stopping				With Early Stopping			
	CV		Test Set		CV		Test Set	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
FLA	0.67	0.68	0.5	0.64	**0.64**	**0.63**	**0.61**	**0.66**
PSW	0.55	0.57	0.51	0.55	0.54	0.61	0.55	0.61
FT	0.52	0.57	0.42	0.58	0.55	0.6	0.56	0.62
FLA + PSW	0.59	0.68	0.44	0.64	0.58	0.61	0.58	0.63
FLA + FT	0.63	0.72	0.47	0.69	0.62	0.63	0.63	0.64
PSW + FT	0.5	0.64	0.5	0.64	0.54	0.61	0.53	0.62
FLA + FT + PSW	0.66	0.66	0.45	0.61	0.6	0.61	0.59	0.62

Table 6: Performance on Prospective Falls

Discussion

The aim of SFRT systems is to improve the screening process for identifying people at high risk of falling in the future. Machine learning based experiments, using features constructed from different combinations of FLA, FT and PSW data types showed that FLA performed best at predicting prospective and retrospective falls. Like results of the statistical analysis, FT and PSW data did not perform well in predicting prospective falls. Both FT and PSW performed with moderate performance when classifying prospective fallers.

Results show that over fitting occurs on the cross validation set when early stopping is not implemented. For retrospective falls, when early stopping was not implemented, sensitivity and specificity for the holdout test set dropped by 8% and 3% respectively compared to the cross validations set. Similarly for prospective falls, when early stopping was not implemented, sensitivity and specificity for the holdout test set dropped by 12% and 3% respectively. No significant difference in performance is seen between cross validation and holdout test set when early stopping is implemented.

As previously discussed, it vital that performance of an SFRT system be evaluated using data that has not been used during training. Performance on the holdout test set is therefore of most interest as it is indicative of performance in real world settings. Results show that the FLA data type performs best for retrospective falls with a sensitivity and specificity of 0.61 and 0.68 respectively. Similarly, the FLA data type also performs best for prospective falls with a sensitivity and specificity of 0.61 and 0.66 respectively. Interestingly, while performance scores were marginally higher for retrospective falls, there was no significant difference between the 2 best performing models for retrospective and prospective falls.

The best performing model for retrospective falls using FLA data, as configured and selected by Bayesian Optimization algorithm, was based on a Naive Bayes classifier. Feature vectors where prepossessed by first standardizing features then processed using a Nystroem kernel map using a polynomial kernel. The best performing model for prospective falls using FLA data, as configured and selected by Bayesian Optimization algorithm, was based on a Stochastic Gradient Descent (SGD) classifier. Feature vectors where prepossessed by first standardizing features then reducing the dimension using PCA to only include 0.98 of variance.

A recent theoretical modelling analysis concluded that the maximal accuracy of a fall prediction model, attempting to identify people with at least one fall incident over the course of a year would not exceed 0.81. While the results reported in our experiments are far from perfect, when compared to the theoretical maximum of 0.81, sensitivity and specificity in the ranges of 0.63-68 are good considering the robust nature in which performance was evaluated.

2 Functional tests for frailty measurement

2.1 Muscle mass, physical activity, bone quality & risk of falling[32]

Background

Lower skeletal muscle density, indicating greater infiltration of adipose tissue into muscles, is associated with higher fracture risk in older adults. We aimed to determine whether mid-calf muscle density is associated with falls risk and bone health in community-dwelling older adults.

Methods

2214 community-dwelling men and women at age 70 were included in this analysis. Mid-calf muscle density (mg/cm3) at the proximal tibia, and volumetric bone mineral density (vBMD) and architecture at the distal and proximal tibia and radius, were assessed by peripheral quantitative computed tomography. Whole-body lean and fat mass, lumbar spine, and total hip areal bone mineral density (aBMD) were assessed by dual-energy X-ray absorptiometry. Participants completed seven-day accelerometer measurements of physical activity intensity, and self-reported falls data were collected 6 and 12 months later. Participants wore a triaxial accelerometer (GT3X+; Actigraph, Pensacola, FL, USA) for seven days following the clinic appointment. This solid-state accelerometer measures acceleration with a dynamic range of Å}6 g in the anterior–posterior (z), mediolateral (x), and vertical (y) axes. Participants were instructed to wear the accelerometer on their non-dominant hip and to remove it only when showering, swimming or in bed at night. They were also instructed to be normally active in accordance with their current lifestyle, to obtain representative accelerometer measurements. Participants who did not provide at least 4 days of at least 10 h per day of valid measurements had accelerometer data excluded. Accelerometer data were collected at a frequency of 30 Hz and data were transformed into “counts” of movement with an activity threshold of 100 counts per min (CPM). Collected data were downloaded using ActiLife 6.11.2 software (Actigraph, Pensacola, FL, USA) in epoch lengths of 60 s with subsequent wear time validation performed. Periods ≥60 min characterized by zero activity were marked as nonwear time, facilitating the exclusion of sleep time from further analyses. Sedentary time was classified as 1 to 99 CPM, while physical activity was classified as light (100 to 1,951 cpm), moderate (1,952 to 5,724 cpm).

Results

302 (13.5%) participants reported a fall at the 6- or 12-month interview, and 29 (1.3%) reported a fall at both interviews. After adjustment for confounders, each standard deviation decrease in mid-calf muscle density was associated with a trend towards greater likelihood of experiencing a fall (OR 1.13; 95% CI 1.00, 1.29 per SD lower) and significantly greater likelihood of multiple falls (1.61; 1.16, 2.23). Higher muscle density was not associated with total hip aBMD, and was associated with lower lumbar spine aBMD (B=-0.003; 95% CI -0.005, -0.001 per mg/cm3) and higher proximal cortical vBMD (0.74;

0.20, 1.28) at the radius. At the tibia, muscle density was positively associated with distal total and trabecular vBMD, and proximal total and cortical vBMD, cortical thickness, cortical area, and stress-strain index (all $P < 0.05$). Only moderate/vigorous (%) intensity physical activity, not sedentary time or light activity was associated with higher mid-calf muscle density (0.086; 0.034, 0.138).

Conclusions

Lower mid-calf muscle density is independently associated with higher likelihood for multiple incident falls over 12 months and may have localised negative effects on bone structure at the tibia. Further studies are required to determine whether these associations persist over the long-term, and potentially contribute to the greater fracture risk previously observed in older adults with low muscle density.

2.2 Using wearables to evaluate symptoms from hands in Parkinsons disease

Introduction

Parkinson's disease is a progressive neurological disorder, which so far cannot be cured. Parkinson's (PD) results from a shortage of dopamine, a chemical that helps instructions from the brain to cross from one nerve cell to the next, in a part of the brain called the substantia nigra, which has to do with controlling movement[33].

The symptoms of Parkinson's disease are multiple; the most identifiable are related with the motor degenerations. In particular, the most recognizable are considered four: tremor, rigidity, movements impairments (akinesia, bradykinesia and hypokinesia), and postural instability. Tremor is the most common; it typically occurs at the distal parts of the limbs and it generally affects a single arm or leg, becoming bilateral with the degeneration of the disease[34]. There are many different types of tremor that affect people in different ways; the most common are:

- Rest tremor: occurs when the muscles are not being voluntarily moved.
- Postural tremor: occurs while maintaining a position such as out stretching your arm.
- Kinetic tremor: occurs when the limb or body part is being moved.

Over a period of around twenty years, the Unified Parkinson's Disease Rating Scale (UPDRS) has been the most used clinical rating for PD[35]. In 2001, it was updated and revised by the Movement Disorder Society (MDS), and, as a result, the MDS-UPDRS was released. Due to the heterogeneity and the complexity of PD symptoms, standard clinical rating may be challenging in the detection of severity levels. Consequently, clinicians with different degrees of experience and background may provide different results.

Wearable sensors give the opportunity to improve the evaluation of the PD motor symptoms, matching the obtained objective results with the clinical assessment.

Aim

The aim of the following study is to implement an automatic indoor evaluation of severity levels in Parkinson's subjects via a supervised machine learning approach. Several devices are used; in particular, inertial measurement unit, surface electromyography, and bend sensors. The protocol for the data collection consists of hand function tests, adapted in part from the MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). The PD severity for each exercise, is computed and compared with the clinician evaluation.

Materials and method

Subjects

Twenty PD patients ≥ 50 years of age and diagnosed with Parkinson's disease and two aged match healthy volunteers participated in the study. The subjects were recruited through the Cork Parkinson's Association of Ireland branch, through word of mouth of these members to non-members, and via

movement disorder clinics. They varied in age, PD stage, and functional ability. We did not include anyone with severe or late-stage PD because they would not have been physically able to travel to the study site and complete the hand movements. Eight PD specialists rated the videos. They included neurologists, geriatricians, and specialist registrars in geriatrics. All patient provided written informed consent. This research protocol has been approved by the local ethical committee.

Protocol

Participants completed six movements with their hands. We adapted the choreography of the movements from the MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). Each movement took about 30 seconds per hand to complete. The testing protocol consists in two repetitions of the following exercise:

- Resting tremor (duration 30 s): the patient sits quietly in a chair with the hands placed on the arms of the chair.
- Postural tremor (duration 30 s): the patient stretches arms out in front of his/her body with wrist straight and palms facing down
- Kinetic tremor (5 times): the patient executes a finger-to-other object movement.
- Finger tapping (30 times): index finger to thumb
- Hand Opening & Closing (15 times)
- Wrist Pronation / Supination (15 times)

All testing is done with the patient sitting in a chair with arm rests, and good back support, with both feet supported on ground.

Devices overview

Participants wore several devices to measure muscle response (electromyography (EMG)) and hand speed (accelerometers). They also wore a prototype of the PD monitoring glove (right hand only). The following research-grade/prototype devices were selected for the study. Details are illustrated as follows:

Xsense 3D motion tracking (Xsense, Enschede, Netherlands): A research-grade motion tracking unit which provides inertial raw data (accelerometer, gyroscope, and magnetometer) as well as position and orientation computed via proprietary validated algorithms.

BTS FREEEMG (BTS Bioengineering Corp., Quincy, USA): is a research-grade 4G technology for surface electromyography (EMG) analysis. The wireless nature, lightness and size enable users to perform free movements without any constraints.

Actisense Glove (licensed technology from the University of Ulster, Northern Ireland): a glove designed to assess joint angles in the fingers of patients suffering from various forms of morbidities. It works by using bend sensors on the joints to feed data back to a control unit on the wrist, which then sends the data wirelessly to a remote station for analysis.

Device set up

First repetition - Devices position (Figure 7.):

2 Xsense units, 1 strapped to the back of each hand.

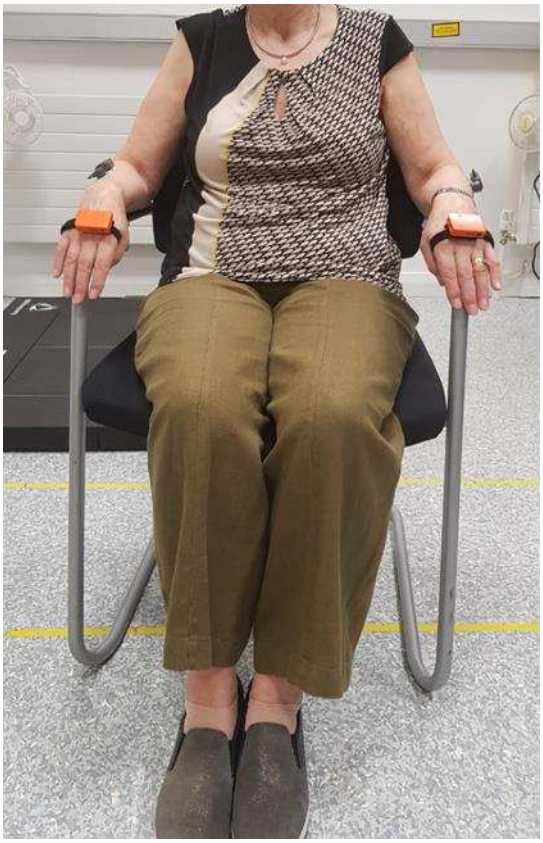
4 FREEEMG units, 1 attached to each underside of the palm and back of each forearm (specifically on the Abductor Pollicis Brevis Muscle and on the Extensor Digitorum Muscle)

Second repetition - Devices position (Figure 8.):

1 Action Sense Glove (Right-Hand only)

1 XSense IMU on the back of the hand (Right-Hand only)

Both the repetitions are recorded by two video cameras (BTS VIXTA50) for the clinicians' evaluation of the disease.



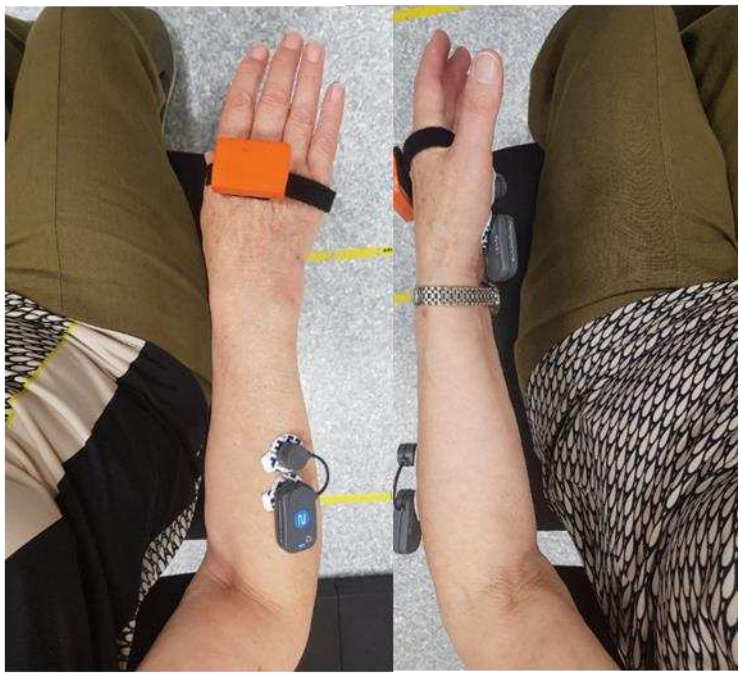


Figure 7. First repetition - Device positions

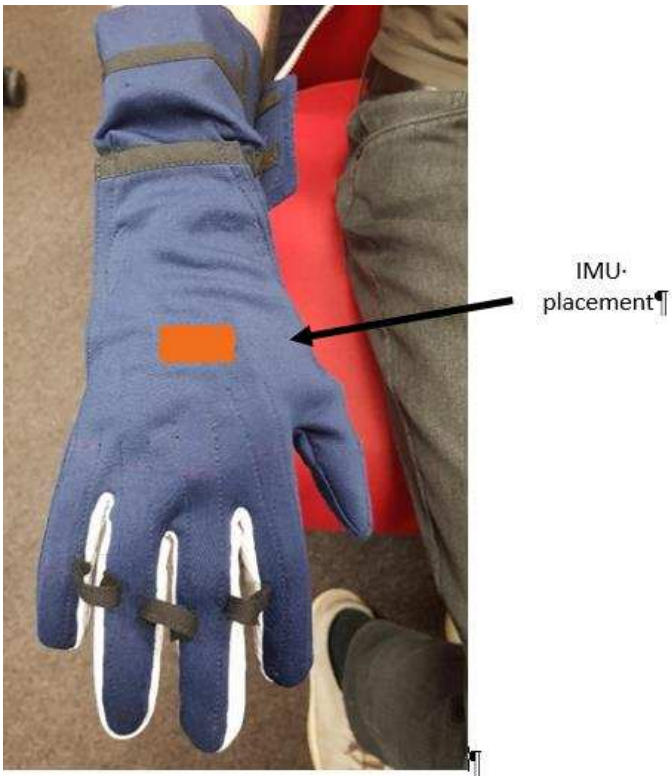


Figure 8. Second repetition - Device positions

Video recording

Two cameras recorded participants to provide the raters with multiple angles of the movements. Participants' faces were blurred to protect their privacy and to reduce potential bias from the raters, some of whom personally know some of the participants.

We compiled two videos for each participant (#1 with EMG sensors, #2 with glove). The videos presented the six movements in a random order. We created 40 videos (20 participants x 2 sets of movements).

The raters rated the 40 videos in a random order.

Rating

Eight PD specialists rated the videos based on their understanding of the MDS-UPDRS criteria. We provided a document with a 10cm line for each movement (total of 240 movements). See an example of the lines below in Figure 9. We asked the raters to place a mark on the line that represented their interpretation of that participant's disease severity. We intentionally left the end of the line open in case the raters wanted to give a rating of 0 or 100. We measured the marks to the nearest millimetre. Scores were recorded as whole number, from 0 to 100.

Figure 9

Analysis

Spread of data

Below are two examples of the spread of data. Each of these box & whisker plots (Figures 10 and 11) corresponds to one participant. The six hand movements (right and left hand) are along the x-axis. The PD specialist ratings from 0-100 are along the y-axis.

Greater spread (i.e. taller box) indicates variability among the ratings and, therefore, possible disagreement among the raters.

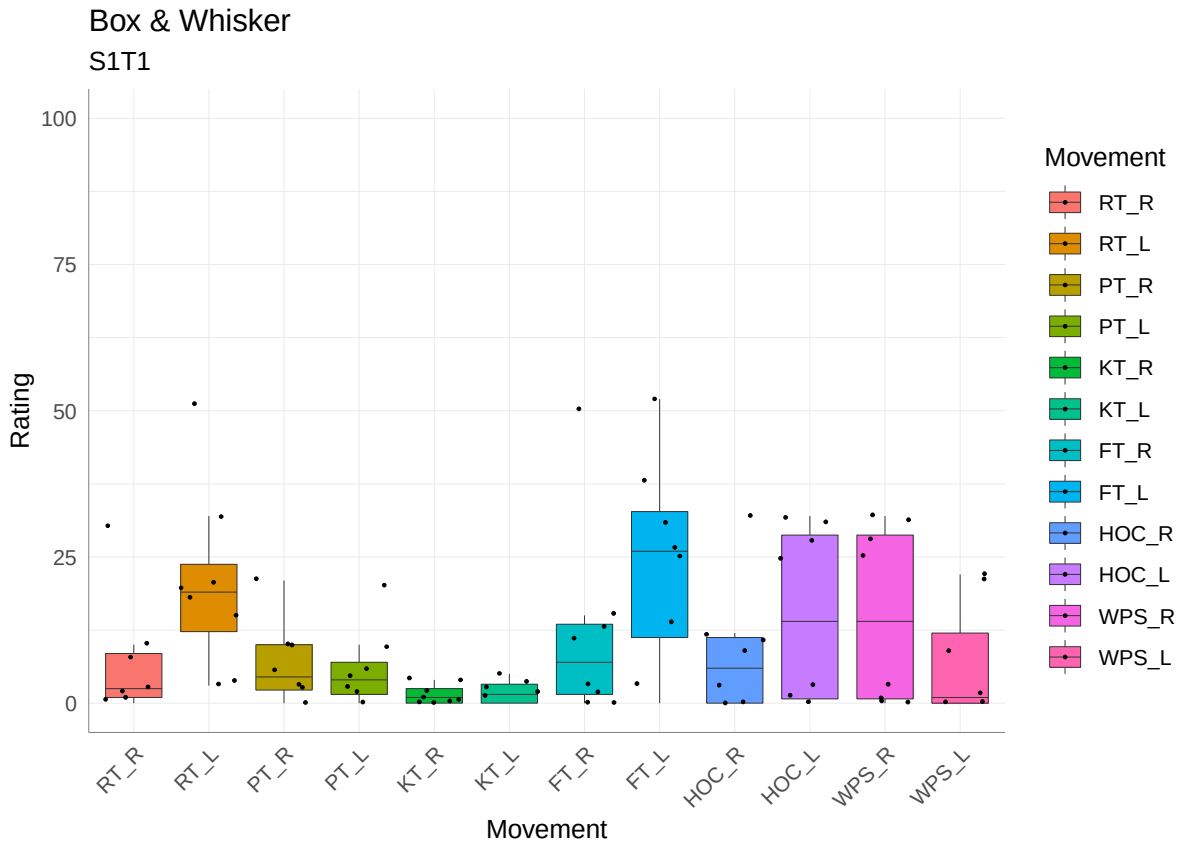


Figure 10

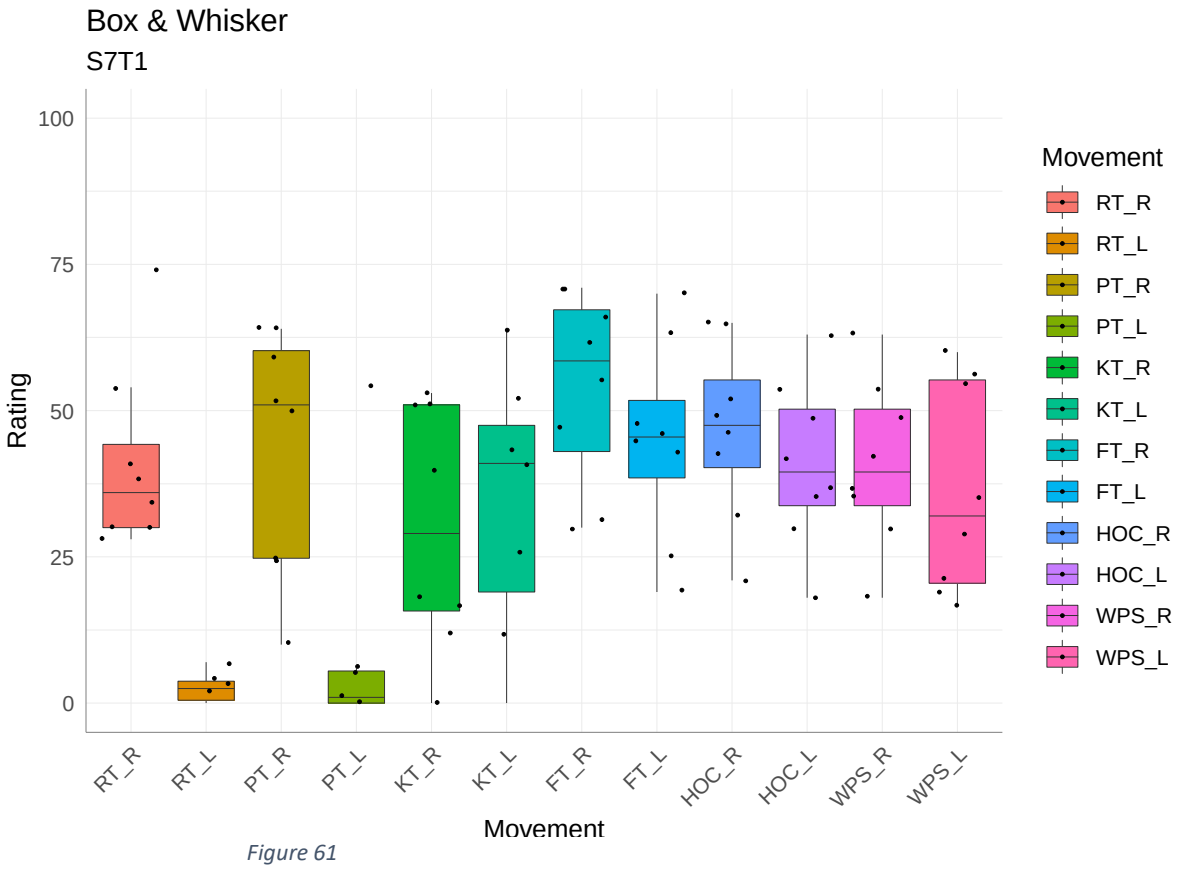


Figure 61

Intra-class Correlation Coefficient (ICC)

The ICC is a measure of reliability, specifically the reliability of two different raters to measure subjects similarly. Inter-rater reliability is important as it demonstrates that a scale is robust to changes in raters. Hence, scales with high inter-rater reliability are less prone to measurement error such as caused by variation in human judgment.

A high ICC close to 1 indicates high similarity between values from the same group. A low ICC close to zero means that values from the same group are not similar.

A low ICC could be caused by several factors: errors and limitations of the measurement tool; improper training of raters; errors by participants; errors or misinterpretations by raters.

ICC Scores

- > 0.90 = excellent
- 0.75 – 0.90 = good
- 0.50 – 0.75 = moderate
- < 0.50 = poor

We calculated 18 ICC scores:

- 6 with the EMG/accelerometer on the right hand
- 6 with the EMG/accelerometer on the left hand
- 6 with the glove on the right hand

Three ICC summary tables are presented below. Most ICC scores are poor. The moderate ICC scores of Resting Tremor could be the result of frequent zero, or “normal,” ratings. Many participants showed no tremor when resting and were therefore given a zero rating. However, the tremors materialised in the subsequent movements and the raters varied in how they scored those tremors.

This study shows that, when presented with videos of hand movements, PD specialists interpret motor symptom severity differently. It remains to be determined what causes these discrepancies. Standardised training, a modified measurement tool, and in person ratings could improve ICC scores and will the topic of future research.

Glove Right Hand

Movement	ICC	Lower CI	Upper CI	Raters	Subjects
Resting Tremor	0.605	0.438	0.778	8	20
Postural Tremor	0.467	0.280	0.680	8	20
Kinetic Tremor	0.282	0.139	0.500	8	20
Finger Tapping	0.355	0.181	0.581	8	20
Hand Opening & Closing	0.315	0.161	0.536	8	20
Wrist Pronation Supination	0.234	0.104	0.444	8	20

Table 7

EMG & Accelerometer Right Hand

Movement	ICC	Lower CI	Upper CI	Raters	Subjects
Resting Tremor	0.661	0.502	0.816	8	20
Postural Tremor	0.575	0.403	0.758	8	20
Kinetic Tremor	0.432	0.263	0.645	8	20
Finger Tapping	0.393	0.228	0.611	8	20
Hand Opening & Closing	0.414	0.244	0.631	8	20
Wrist Pronation Supination	0.342	0.186	0.562	8	20

Table 8

EMG & Accelerometer Left Hand

Movement	ICC	Lower CI	Upper CI	Raters	Subjects
Resting Tremor	0.660	0.500	0.815	8	20
Postural Tremor	0.142	0.040	0.330	8	20
Kinetic Tremor	0.445	0.265	0.659	8	20
Finger Tapping	0.289	0.144	0.506	8	20
Hand Opening & Closing	0.342	0.186	0.562	8	20
Wrist Pronation Supination	0.458	0.289	0.667	8	20

Table 9

Classification

From the data collected, a number of features were extracted: concerning resting, postural and kinetic tremor as well as wrist pronation/supination the features will be mainly extracted from the Xsense; while for finger tapping and hand opening and closing will be considered Xsense, FREEEMG and Glove.

Preliminary data processing

Concerning inertial sensors, data were resampled from 100 Hz to 1000 Hz. Subsequently, the signals were filtered by means of a moving average filter with windows of five samples; the main aim of this step was to eliminate the influence of undesired low frequency components. Then, values were scaled by subtracting their mean and dividing the absolute value of their max, to guarantee that data oscillate evenly around zero.

The EMG signals (sample frequency = 1000 Hz) were firstly filtered by a high-pass Butterworth filter of order 4 (cut-off frequency of 10 Hz) and secondly by a moving average filter with windows of three samples; the purpose was to eliminate the influence of undesired low frequency components. Afterward, the EMG signals were scaled by subtracting their mean and dividing the absolute value of their max, to ensure that data oscillate evenly around zero.

Preliminary data processing was carried out in MATLAB R2019b (Mathworks, Inc., Natick, MA, USA).

Feature Extraction

Several time and frequency domain features were calculated for both inertial and EMG data. All the features were computed for both right and left side of the signals: ax, ay, az, atot, gx, gy, gz, gtot, jerkx, jerky, jerkz, jerktot, abductor pollicis brevis muscle (APB), extensor digitorum muscle (ED), evelope of APB, and envelope of ED.

Time domain

- Mean absolute value
- Root mean square
- Variance
- Interquartile range
- Skewness
- Kurtosis
- Wave form change
- Wave form change squared
- Energy

Frequency domain

- Dominant frequency
- Magnitude dominant frequency
- Spectral centroid
- Spectral edge frequency
- Harmonic ratio
- Index harmonicity
- Median frequency
- Entropy

To enhance the extraction, signals were divided into observation widows of 700 samples with an overlap of 500, in exercise one (resting tremor), two (postural tremor), four (finger tapping), and five (hand opening and closing). While movements were segmented in exercises two (kinetic tremor) and six (wrist pronation/supination). Figures 12 and 13 show how data were segmented during kinetic tremor (finger to object) and wrist pronation/supination.

Feature extraction was carried out in MATLAB R2019b (Mathworks, Inc., Natick, MA, USA).

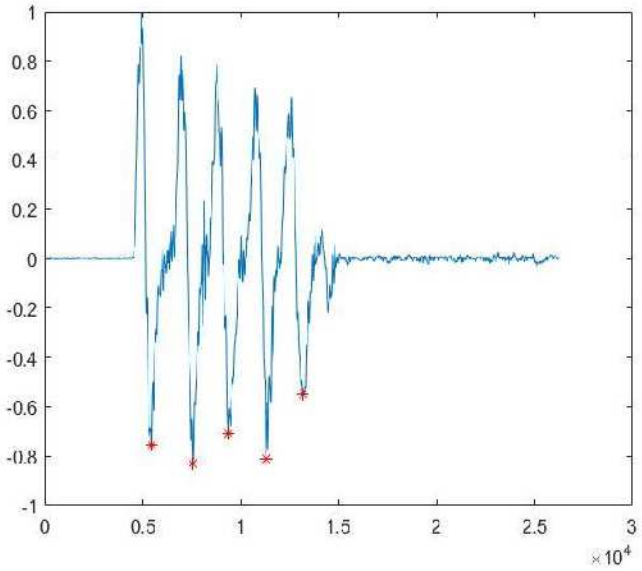


Figure 12. Kinetic tremor - number of samples (x-axis) vs signal amplitude (y-axis)

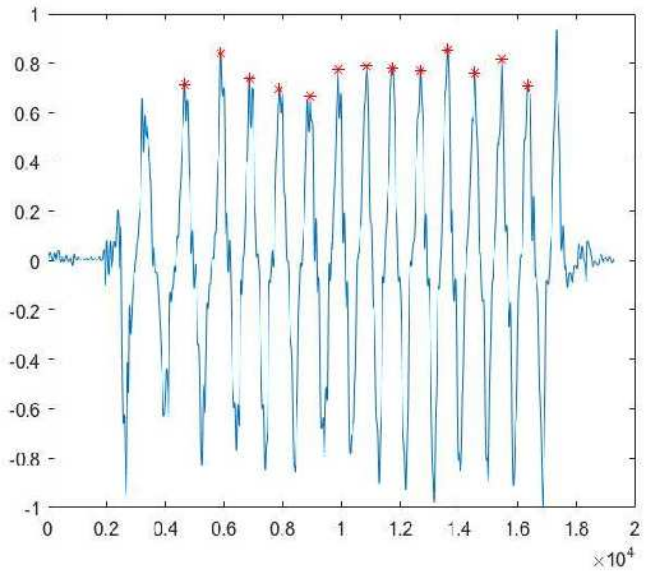


Figure 13. Wrist pronation/supination - number of samples (x-axis) vs signal amplitude (y-axis)

Learning Model

Eight PD specialists rated the exercises based on their understanding of the MDS-UPDRS criteria; scores varied from zero to 100. To assess the presence of tremor, each observation window (for resting tremor, postural tremor, finger tapping, and hand opening and closing) and movement (for finger to object and wrist pronation/supination) was labelled 1 (absence or mild tremor – score from zero to ten) or 0 (presence of tremor – score from 10 to 100).

All features were fed to a supervised-based classifier developed in Python 3 (Python Software Foundation, Delaware, US). Models considered for the classifier were k-Nearest Neighbors (kNN), Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF). F1 score was used as a metric to quantify the goodness-of-fit comparing the predictions of the classifier with the real labels each observation belongs to.

To build the model, the dataset was split into training set, 80%, and test set, 20%. Stratification was considered during the partitioning to insure the same percentage of 0 (presence) and 1 (absence) in both sets.
Subsequently, a k-fold cross-validation (with k = 10) and a grid search was employed on the training set to obtain optimal values for the model hyper-parameters. Fitting and feature selection were deployed simultaneously. The F1 score for train, validation, and test set was used to evaluate the performance of each classifier.

Results and Discussion

The results for each exercise and classifier are synthetized in Table 10.

Exercise one - Resting tremor

The four classifiers had good performances with F1 test higher than 0.96 (Table 10). The classifiers were neither overfitting nor underfitting (Table 10). KNN had the best results; Figure 14 shows the related confusion matrix.

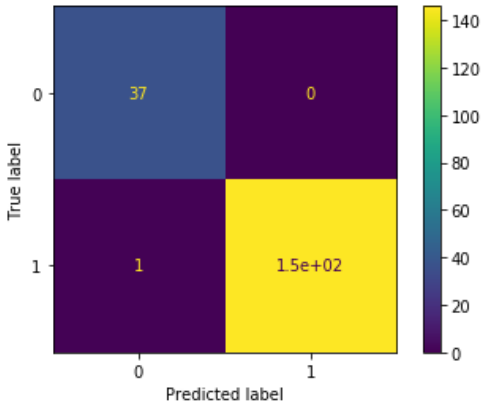


Figure 14. Exercise one – KNN - Confusion matrix

Exercise two - Postural tremor

The four classifiers reported an F1 test higher than 0.96 with neither overfitting nor underfitting (Table 10). KNN had the best results (Figure 15).

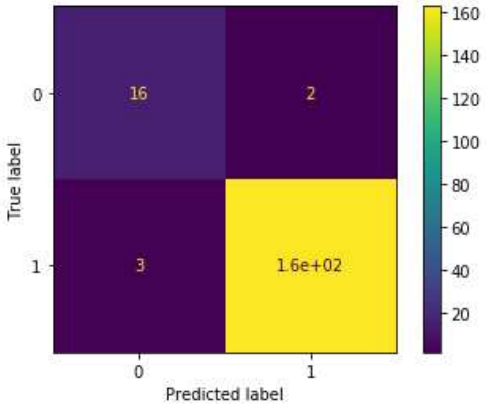


Figure 15. Exercise two – KNN - Confusion matrix

Exercise three - Kinetic tremor

KNN achieved an F1 test equal to 0.94 without overfitting or underfitting (Figure 16 and Table 10). LR, DT, and RF suffered from overfitting.

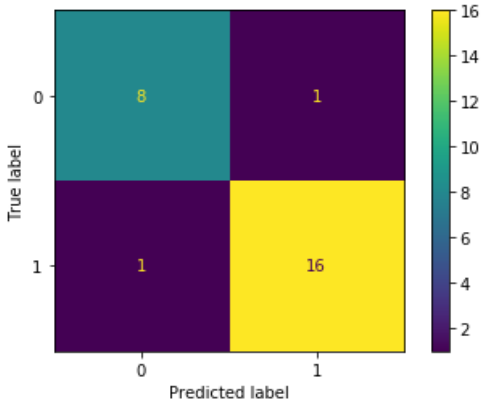


Figure 16. Exercise three – KNN - Confusion matrix

Exercise four - Finger tapping

KNN achieved an F1 test equal to 0.96 without overfitting or underfitting (Figure 17 and Table 10). LR, DT, and RF suffered from overfitting.

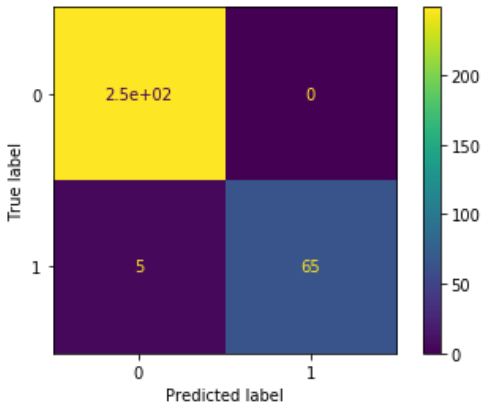


Figure 17. Exercise four – KNN - Confusion matrix

Exercise five - Hand opening and closing

KNN achieved an F1 test equal to 0.88 without overfitting or underfitting (Figure 18 and Table 10). LR, DT, and RF suffered from overfitting.

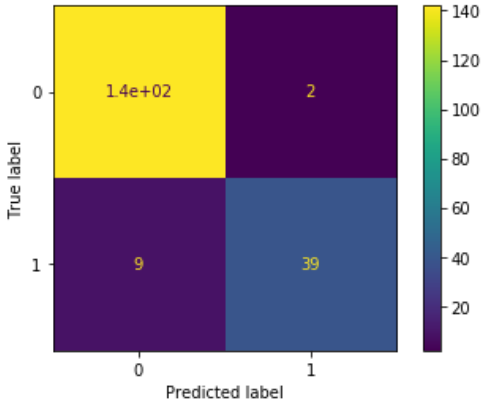


Figure 18. Exercise five – KNN - Confusion matrix

Exercise six - Wrist pronation/supination

KNN, LR, and RF had good performances with F1 test higher than 0.83 (Table 10). The classifiers were neither overfitting nor underfitting except for Decision tree (Table 10). KNN had the best results; Figure 19 shows the related confusion matrix.

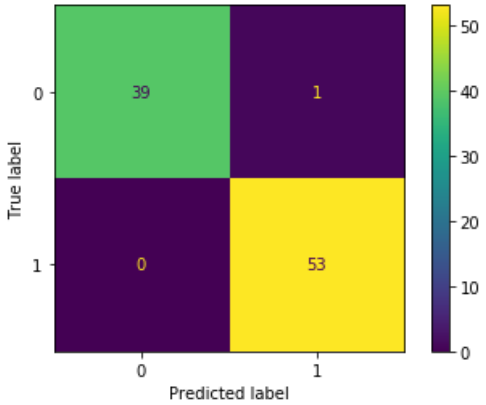


Figure 19. Exercise six – KNN - Confusion matrix

Exercise	Classifier	Hyperparameters	F1 traning	F1 validation	F1 test
1	KNN	k =1	1.00	0.99	1.00
	Logistic Regression	C = 0.40	1.00	0.98	0.99
	Decision Tree	Max depth = 5	0.98	0.94	0.96
	Rondom Forest	Criterion = entropy Max depth = 7 Number of estimators = 60	1.00	0.98	0.99
2	KNN	k =1	1.00	0.99	0.98
	Logistic Regression	C = 0.40	1.00	0.98	0.98
	Decision Tree	Max depth = 5	0.99	0.96	0.96
	Rondom Forest	Criterion = entropy Max depth = 7 Number of estimators = 40	1.00	0.98	0.98
3	KNN	k =1	1.00	0.95	0.94
	Logistic Regression	C = 0.20	1.00	0.94	0.88
	Decision Tree	Max depth = 5	1.00	0.77	0.77
	Rondom Forest	Criterion = gini Max depth = 6 Number of estimators = 40	1.00	0.90	0.87
4	KNN	k =2	0.97	0.94	0.96
	Logistic Regression	C = 0.20	0.93	0.75	0.77
	Decision Tree	Max depth = 6	0.91	0.72	0.70
	Rondom Forest	Criterion = entropy Max depth = 7 Number of estimators = 70	0.98	0.79	0.83
5	KNN	k =2	0.96	0.88	0.88
	Logistic Regression	C = 0.1	0.94	0.75	0.71
	Decision Tree	Max depth = 7	0.95	0.72	0.66
	Rondom Forest	Criterion = entropy Max depth = 7 Number of estimators = 80	0.99	0.71	0.59
6	KNN	k = 3	1.00	0.98	0.99
	Logistic Regression	C = 0.1	1.00	0.95	0.96
	Decision Tree	Max depth = 5	0.99	0.84	0.83
	Rondom Forest	Criterion = entropy Max depth = 6 Number of estimators = 40	1.00	0.96	0.95

Table 10. Performance of KNN, LR, DT, and RF.

2.3 Measuring spinal flexibility with wearables

SUPERVISED IMU RELIABILITY IN DERRY/LONDONDERRY, NORTHERN IRELAND, UK

Introduction

In 2016, the highest prevalence of spondyloarthritis was in North America and Europe, 1.35% and 0.54% respectively [36]. A key outcome measure for axial spondyloarthritis (axSPA) - a form of arthritis causing inflammation of spinal joints - is spinal mobility, but this is highly variable and subjective when assessed using conventional tools, e.g., goniometer [37]. Cervical rotation is the only movement test measured in degrees in the Bath Ankylosing Spondylitis Metrology Index (BASMI). Inertial Measurement Unit (IMU) sensors can attain accurate measures of body motion.

Aim

Our main goal was to evaluate the reliability of IMU attained measurements (from movement tests performed by axSpA participants) against measurements attained from conventional BASMI, the 'Truth' dimension of the OMERACT filter [38].

Materials and methods

Our methodology was as follows: Spinal movements were tracked and recorded using IMU sensors in 40 axSpA participants, who have a wide range of disease severity and a mean disease duration of 13 years, at Altnagelvin hospital in Derry/Londonderry, Northern Ireland (2018-2019). Dorsa Vi Move IMU sensors were used in this study and set specifically in two positions (Figure 20):

1. At the lumbar level over T12 and S1 bony landmarks (Vi Move custom adhesives and height-specific templates were employed to achieve position accuracy).
2. At the cervical level at occiput and T3.

Apart from monitoring back and neck movements on patients, there were also collected patient reported outcomes and conventional metrology.

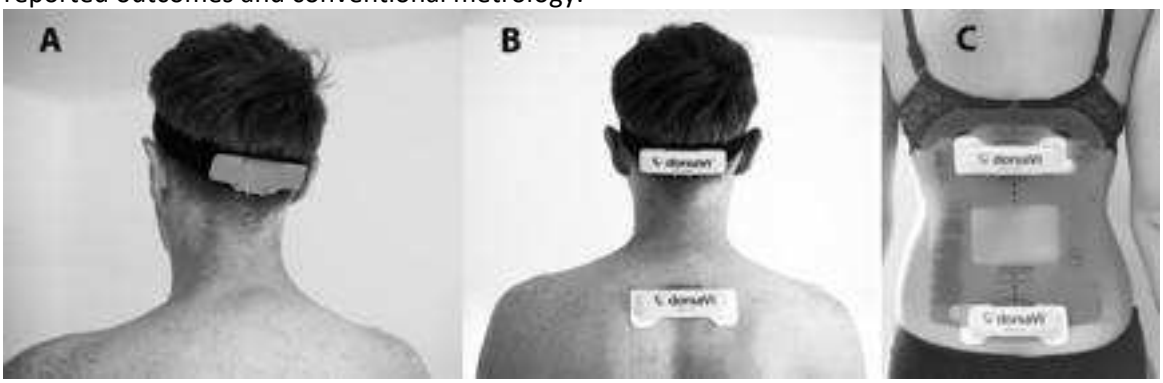


Figure 20. A: Clip-in Baseplate for sensor: B: Positioning of cervical sensors (occiput, T3) C: Positioning of Lumbar sensors using height specific template (T12, S1)

Intra-rater, inter-rater and test-retest reliability with a two-week gap between tests were performed as follows:

Day1:
Physio A takes the Range of Movement (ROM) measurements from the participant wearing the sensors twice: early in the morning and then again after a couple of hours.
Physio B takes the ROM measurements from the same participant wearing the sensors after a couple of hours and after Physio A completed his/her set of measurements.
Day2:

Physio A takes the ROM measurement from the participant wearing the sensors again. The ROM angles at lumbar and cervical regions of each participant visit were analysed using IBM SPSS v 25. The technique employed was Intraclass Correlation Coefficients (ICCs). For both regions, the maximum ROM angles for anterior flexion/extension (Flex), lateral flexion (Left+Right) and rotation (Left+Right) were analysed. Also, these six attained values were employed to calculate the IMU-ASMI composite score. Pearson correlation coefficients were calculated for each component and the overall score and with Bath Ankylosing Spondylitis Functional Index (BASFI).

Results

Person correlations for IMU and conventional measurements were obtained and shown in Tables 11 and 12, respectively.

	Cervical IMU			Lumbar IMU			
Movement	Flex	Lat	Rot	Flex	Lat	Rot	IMU-ASMI
Intra-rater	0.95	0.87	0.97	0.89	0.84	0.76	0.96
Inter-rater	0.96	0.96	0.97	0.94	0.94	0.77	0.98
Test-retest	0.92	0.84	0.96	0.91	0.94	0.81	0.97

Table 11. Reliability ICCs for IMU test

	Cervical IMU			Lumbar IMU			
Movement	Flex	Lat	Rot	Flex	Lat	Rot	IMU-ASMI
Intra-rater	N/A	N/A	0.96	1.00	0.94	N/A	0.97
Inter-rater	N/A	N/A	0.91	0.68	0.96	N/A	0.98
Test-retest	N/A	N/A	0.79	0.73	0.91	N/A	0.96

Table 12. Reliability ICCs for conventional spinal test

As can be observed in Tables 11 and 12, sensor tests show excellent reliability. Potential 'floor' effects were seen with cervical tests, and 'ceiling' effects were seen with some lumbar tests. No such issues were seen with composite IMU values. IMU measures showed reliable equivalence with the comparable BASMI measurements (correlating closely). Cervical rotation, side flexion, lumbar flexion, and cervical flexion with values of r equals to 0.85, 0.84, 0.62 and 0.65, respectively. The mean BASMI was 4.8 (range from 1.2 to 8.4). The mean IMU-ASMI score was 4.0 (range from 0.1 to 9.3). The correlation between BASMI and IMU-ASMI was 0.88. The correlation between BASFI and BASMI was 0.68 and between BASFI and IMU-ASMI is 0.71.

Lumbo-pelvic restriction unforeseen patterns, Figure 21, were found: The mean contribution to the lumbar flexion by the pelvic movement was 52.6%. Fifteen out of the forty participants had an abnormal Lumbo-Pelvic Rhythm (LPR): There were identified 7 and 8 individuals with lumbar and pelvic restrictions, respectively. Of these identified individuals, 12 had trunk flexion less than 70 degrees and in those with trunk flexion larger than 70 degrees, 22 out of 25 had a normal LPR.

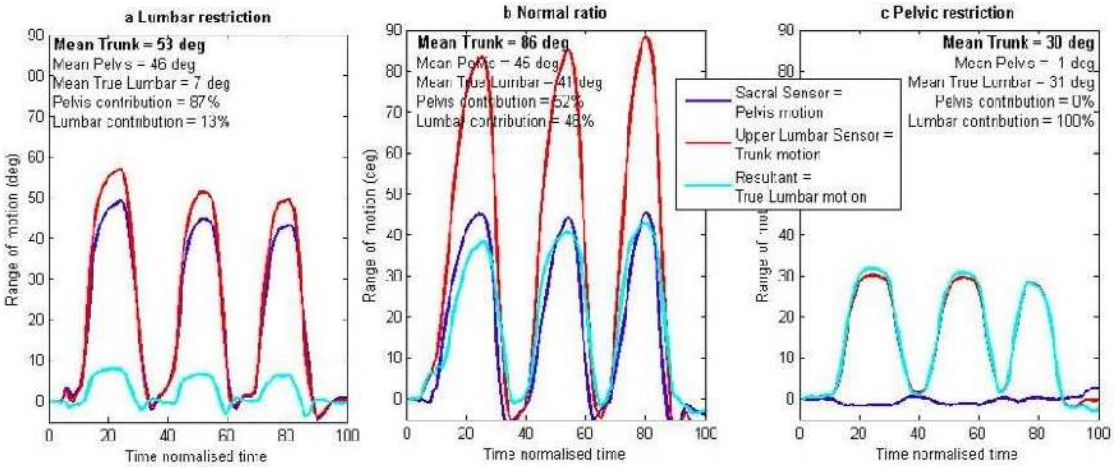


Figure 21. Examples of Lumbo-pelvic restriction observed on this study

Discussion and conclusion

Test-retest reliability of individual cervical movement-test was good to excellent (ICCs>0.8), superior to those reported by Theobald, Jones and Williams [39]. Lumbar movements tests had slightly lower test reliability (ICCs>0.7), like the findings reported by Ronchi et al.[40] and Laird, Kent and Keating [36] using the same sensor setup. Combining the right and left or flexion movements improved reliability, probably, because it is difficult for assessors to appraise the return to the exact midline point.

Both intra-rater and inter-rater reliability were excellent. Cervical tests were more reliable than lumbar-spine movement tests. The variability in lumbar measurements was due to biological variability rather than sensor error: Laird, Kent and Keating [41] suggested that it was due to inherent variability in the "lumbo-pelvic rhythm". IMU sensor tests are of clinical relevance because they give detailed overview of movement limitations in degrees by region and movement type.

IMU measures showed reliable equivalence with the comparable BASMI measurements, correlating closely. Lumbo-pelvic restriction patterns that were unexpected were found. IMU sensors were highly reliable in measuring spinal movement for axSPA patients.

Future work

A mobile application is currently on development, which will be used in combination with IMU sensors that the participants can bring with them at home. A usability assessment study will be conducted with the App.

2.4 Usability of wearable sensors in the rehabilitation hospital and home rehabilitation - Emphasis on two sensor systems analyzing gait

Purpose

The purpose of the study was to analyze *Usability and Utility* of two wearable sensor systems of measuring gait in real life rehabilitation for elderly. Evaluated was smart insoles (MoveSole®) and one sensor system (G-WALK®). These attributes were in two phases qualitative protocol evaluated: Questionnaire (System Usability Scale) after each use and Theme interview after longer using period.

Methods

This research used a qualitative approach, and the content analysis was made with theoretical/thematic analysis. Twelve participants were interviewed. The interviews were face-to-face, semi-structured theme interviews and the themes were planned in advance. In semi-structured interviews the participants have opportunity to raise issues even though there are ready-made themes in the interview [16] (pp. 78.), and this was also considered in the interviews.

We compiled the interview themes based on Nielsen´s criteria of usability [42] (pp.25) and in addition the context theme. System acceptability is divided into social acceptability and practical acceptability. Practical acceptability is further divided usefulness and usefulness to utility and usability. Our theme interview included social acceptability, and usability from the practical acceptability.

Preliminary results

Qualitative analysis and reporting is currently on going.

The insole system (MoveSole®) could be used regularly, and it is easy to learn and use. Major variation was reported about the experience of the inconsistency of the insole system. Experiences from G-WALK® -system seems to be uniformly positive. Therapists had varied thoughts about their need for development and support of new skills and knowledge before the permanent use of wearable sensors. Most of the patients were interested in the achieved results and were willing to use sensor systems, but they need assistance in the use and interpretation of results.

2.5 Validation process of one wearable sensor system product

Abstract

Sensor insoles usually measure gait ground forces or pressure using piezoelectric phenomenon. Micro voltage (µV, µC) results are used as the basics of counting pressure and/or force values, and various algorithms are used to achieve necessary results from measures. Typical results, that are wanted for research and especially clinical work, are for example movements of center of mass during steps, distribution of forces or pressure to different areas of the foot during steps, recognition of properties in different step phases – heel strike, toe off, mid stance, effect of pronation and supination of foot, time of step etc. MoveSole StepLab is a mobile measurement system for instant underfoot force measurements during gait, which includes a MoveSole Smart Insole and a MoveSole Smart Device wirelessly connected to each other. Unique, electromagnetic film (EMFI) based sensor technology and printed electronics production technology is integrated in the MoveSole StepLab measurement system. The MoveSole StepLab measures plantar ground force distribution over the sensors and views an estimation of the maximum total ground force.

Our goal was to analyze if the MoveSole is valid equipment for clinical use in rehabilitation processes. We made three phases of validation process in order to extract relevant parameters and compared the results to a Kistler force plate using Bioware analyzing program as a gold criteria / standard measure. We also used other measuring equipment and clinical examination to find impacts from spatiotemporal and functional capacity connection to results.

Phase one

The purpose of the first phase was to compare the ground force results of vertical (Y-) direction of Kistler force plate and MoveSole smart insoles and the impact of walking speed on those values. Participants trained a few times to walk through the 10 m distance, so that they could be certain that their left foot would land on the first Kistler plate (5th step) and the right foot on the second Kistler plate (6th step). The two Kistler force plates were placed about the middle of the total distance, so that the walking speed was possible to accelerate to selected level. Participants walked two times with

their own natural walking speed and two times with their maximum speed. The walking speed was measured from the middle four meters. (Face, criterion and concurrent validity).

Phase two

The purpose of the second phase was to compare Kistler and MoveSole results when the insole was inside a thin but tight sport sock and on top of those, a slipping preventing Instant Grip Sock® (CareCare ltd). Original algorithm was developed through wrestling shoe measures. The set-up was developed to find out, if the algorithm is accurate compared to - as close as possible - bare foot measurements and if the foot elasticity properties could be detected in MoveSole values.
Participants trained a few times to walk through the distance, so that they could be certain that their left foot would land on the first Kistler plate (3rd step). Each participant walked with their own natural walking speed seven meters and third (left foot) step ground force was measured. After this, each participant walked four times, with their own natural walking speed. (Criterion validity)

Phase three

In the third phase, three older male volunteers, with different lower limb problems, due hip or knee arthrosis, participated. We analyzed the maximum ground force with Kistler and MoveSole and distribution of weight and force to different MoveSole sensors. We studied what kind of impact Sievi-working shoes had on maximum ground forces, weight and force distribution, together with of-shell individually selected or individually designed and structured FootCare® insoles. Comprehensive physiotherapeutic clinical examination of lower limb was also done. GaitRite® and G-walk® analysis of kinematic properties of gait was done to check if there are some severe spatiotemporal changes in gait. The purpose was to find out if the results of Kistler MoveSole behave logically, concerning clinically evident changes of structures and movements of foot and lower limb. This third phase was considered as a case study, i.e., could this equipment detect possible impacts in gait (Content validity). First, an experienced physiotherapist made a comprehensive examination of lover limb structures. Secondly, the length of the lower limb was measured (highest palpated point of trochanter major – floor). Thirdly, the height and weight of the participant was measured. Fourth foot width, length, shoe size and suggested type of Sievi special insoles were analyzed with a Sievi Scanner. Fifth, an Elinvision iQube 500 (JSC Elinvision) (20) scanner was used to measure the foot properties for constructing an individually designed Footcare® insole. Sixth, each participant placed a G-walk® sensor on top of the sacrum S2 – level. Each participant walked first 10 meters with MoveSole insoles inside a tight athletic sock, and on top of it IGS grip socks, and after that they walked 15 meters through GaitRite®. Next, participants walked similar way with two different types of Sievi working shoes with basic insole, individually selected insole and individually designed and structured insole. One type of shoe is a safety shoe with shock absorption structure and the other without. Both have toe and nail protection. This was done to detect possible impact of shoes and insoles on foot posture and behavior during step.

Results generally

Totally 46 voluntary persons participated, in these 3 phases of validation. They walked 4-8 times, 7-15 m distances, with personal normal or personal maximum speed. Mainly participants were healthy adults, to find out the reliability and accuracy compared to golden standard /system within easier situation – healthy. Three participants were older 55-61 years, who had various gait problems caused by arthrosis and body posture and structures. Key finding was that correlation/accuracy of Movesole reached almost clinically relevant level (Pearson correlation about 0.822 - .875). The mean difference of Kistler and Movesole values was too big for clinical use varying 9.1 N / 130 N, 130 N /82.3 N depending on gait speed and shoe or sock outfit. According Bland-Altman analysis reliability should be developed. Important factor to develop within Movesole, was the time that highest vertical ground

force was evident. I.e., does it occur either during heel strike or toe off (Pearson correlation .351 - .430 - .504). Movesole seems to be able recognize changes to ground forces of foot caused different structures and ROM, and compensatory actors (shoe, Insole). Manufacturer is considering developments of insole design, working instructions and the algorithm according validation process results.

2.6 Demonstrator project: older people use of wearable sensors in home or homelike environments with remote connections and equipment

Sencom remote connection equipment with Movesole smart insoles was used with 14 rehabilitees both female and male. MPower emg system was used with one rehabilitee. Totally 84 remote rehabilitation sessions with sensors. They were aged 55 – 78 years with various problems of functional capacity and diagnosis. Each trial period was 2 – 4 weeks, during that approximately 6 contacts /person was conducted. Key factors: technical demands, interaction demands, key attributes of both persons and equipment and instruction / guidance:

1. Some participants had problems remembering what to do in different phases, so the equipment was too difficult to use. Conclusion is that the rehabilitee selection and usage guidance and support is important.
2. In a couple of cases the connections didn't work well enough (for example stone house in customer end and University heavy loaded wifi etc. connections). The connections and support must be reachable.
3. Generally, rehabilitees experiences were positive, and results - when explained/understood - were interesting and motivating, while using the equipment during updating exercise according online reports or instructing exercise while doing them. Some older persons reported that the results activated them to do more exercise.
4. When instructing exercises online with the sensor data online clearly gave more important information for therapist, more than just plain picture.
5. Environment – space, lights etc. must be considered and timetables agreed. responsibility of everything working is always "in the therapist end" so you must be able to guide on some level both IT and sensor/application technology during session.

3 Gait analysis for exercise

3.1 Measuring of physical activity as potential risk factor for diabetes[43]

Insufficient physical activity (PA) is an important risk factor for obesity and noncommunicable diseases (NCDs) such as cardiovascular disease (CVD) and type 2 diabetes. For this reason, the World Health Organization (WHO) has issued recommendations for different age categories, stating that 150 minutes of moderate-intensity PA (or 75 minutes of vigorous PA) per week, performed in bouts of at least 10 minutes each, will provide significant health benefits for individuals aged 18–64 and ≥65 years44. Moreover, reaching 300 minutes of moderate-intensity PA (or 150 minutes of vigorous PA) per week is suggested to confer additional health benefits. These recommendations are based on several extensive reviews assessing the relation between predominantly self-reported PA and different aspects of health. Therefore, in the present study, we investigated objective and self-reported PA in a large cohort of 70-year-old individuals, with emphasis on PA patterns and amounts in individuals with and without diabetes (type 1 or type 2) in relation to the WHO guidelines.

Participants

The investigated cohort comprised 1,872 participants with complete measurements of PA patterns, blood parameters, and body composition. Body composition was assessed by waist circumference and by visceral adipose tissue (VAT), quantified by dual-energy X-ray absorptiometry using a Lunar iDXA

and the CoreScan application (GE Healthcare, Wauwatosa, Wisc.). No exclusion criteria were applied because the aim was to investigate a sample reflecting the general population. Data were collected between June 2012 and December 2015, and data analysis was conducted in 2016. The study was approved by the Umeå Regional Research Ethics Committee (Dnr 07-031M with extensions).

Diabetes Criteria

Participants enrolled in the HAI were instructed to be in a fasting state, with no intake of food, calorie-containing beverage, or nicotine for 4 hours before the clinic visit. A capillary blood sample was analyzed on a HemoCue portable glucose analyzer (HemoCue AB, Ängelholm, Sweden) to determine the fasting blood glucose (FBG) concentration. In addition, participants' self-reported information on prior diabetes diagnosis was obtained. Information of self-reported diabetes had been shown to have very high specificity (99.7%) compared to medical records (16). Participants were divided into four groups based on FBG and prior diabetes diagnosis: no diabetes (group 1; FBG <_6.1 mmol/L (<_110 mg/dL) and no known diabetes, n = 1,396), prediabetes (group 2; FBG ≥6.1 to <_7 mmol/L [≥110 to <_125 mg/dL] and no known diabetes, n = 266), diabetes detected in the study (group 3; FBG ≥7.0 mmol/L [≥125 mg/dL] and no known diabetes, n = 52), and known diabetes (group 4; regardless of blood glucose concentration, n = 158). For most results presented, groups 1 and 2 and groups 3 and 4 were merged into "no diabetes" and "diabetes" groups.

PA Assessments

Self-reported PA data were obtained using the International Physical Activity Questionnaire–short form (IPAQ-SF) (17), which has been validated in several populations in different countries (18). Respondents were asked to recall the time spent (in at least 10-minute bouts) walking and engaging in moderate and vigorous PA during the past 7 days. Participants were also asked to report their adolescent PA levels on a scale from 1 to 5, in which 1 represented excluded from physical education and 5 represented training and competing at an elite level.

PA was measured objectively using triaxial accelerometers (GT3X+, Actigraph, Pensacola, Fla.) and subsequently filtered (normal frequency). Accelerometers are motion sensors that detect duration and intensity of PA by measuring accelerations in three dimensions. Sampled at 30 Hz, acceleration data were then transformed into counts representing activity. For this study, participants were asked to be normally active as they wore the accelerometers on their nondominant hip for 7 consecutive days, removing it only during water-based activities, including bathing, and nighttime sleep.

Upon participants' return, accelerometer data were downloaded with ActiLife software 6.6.3 (Actigraph) into epoch lengths of 60 seconds, and wear-time was validated in accordance with Troiano et al. 45. Briefly, participants were required to accumulate at least 4 days of PA measurements, with a minimum of 10 hours/day, for the data to be considered valid and eligible for further analysis. Non-wear time was defined as >60 minutes of inactivity, with a spike tolerance of no more than 2 minutes exceeding the cut point for sedentary time. PA patterns were investigated using predetermined cut points based on uniaxial counts as described by Freedson et al. (20)46, stating that tasks reaching a metabolic equivalent of ≥3 (e.g., walking, jogging, and cycling) are to be considered MVPA. A bout of MVPA was defined as at least 10 consecutive minutes, reaching the MVPA threshold of 1,952 counts per minute, with allowance for 1- to 2-minute interruptions. The total time per week in MVPA bouts and total time in MVPA without bout prerequisite were then recorded separately and used for further calculations. In addition, daily step-counts were also attained from accelerometers and retrieved by the ActiLife software. Finally, total registered counts from all three axes (with no bout limitation and no regard for intensity level) were combined and divided by total accelerometer wear time to form the total PA variable, which represents a measure of all PA conducted during the day and thus in addition to recorded MVPA also included light PA (e.g., lighter gardening and household work).

Summary

In this population-based study of men and women aged 70 years, participants with known diabetes at baseline had PA >20% lower than those without diabetes, as measured by objective accelerometers. A cut point of 6,000 steps/day discriminated best between subjects with and without diabetes. Number of steps per day also showed the strongest association with amount of VAT, a potential mediator of the effects of PA with respect to diabetes.

3.2 Exploratory Gait Analysis using Wearable Technology

Gait Analysis (GA) is an area of research that is continually expanding and evolving across a wide range of domains such as medical, healthcare, sport science and surveillance. There are a plethora of gait applications such as the evaluation of prosthetics, surgical procedures [47] [48], treatments plans, fall risk in the elderly [49], elite athletes [50], assessment of neuropathies [51] and identification of individuals for forensic biometric purposes [52] [53].

During the past four decades the measurement and assessment of gait has evolved rapidly, tools and technology now provide an objective more quantitative approach. Current clinical practice for motor assessment of the lower limb in stroke survivors are based upon assessments using a battery of tests such as two-minute walking test, timed-up and go, berg balance scale, fugl-meyer assessment, motor assessment scale, rivermead motor assessment of movement, motricity index and stroke rehabilitation assessment of movement. All the aforementioned motor assessment scales predate the year 1997 and have an average age of 31 years. Although they provide a quantitative score they are based upon human clinical observation and are subject to inter- and intra-rater variability. Additionally, most of these assessment approaches are not capable of detecting subtle changes in motor function particularly at the top end of the scales as a ceiling effect often occurs [54].

Advances in technology used to measure gait have been instrumental in the evolution of GA. Biomechanical movement of the human body is complex therefore effective GA requires information such as kinematics, ground reaction forces and muscle activity. Motion Capture (MC) strives to measure kinematic data in accurate, valid, and unobtrusive manner. There are two competing MC technologies which offer various advantages and have some disadvantages depending on the context of the application being considered. In this study we will explore the use of Inertial Motion Capture (IMC) technology to collect gait information.

Aim

The aim of the study is to create a computational gait model via to represent walking activity for the healthy population which encompasses some elderly participants. It is anticipated that this model will provide a baseline gait model from which further studies into gait analysis for Stroke survivors can be conducted.

Participants

Six healthy volunteers have participated in the study to date. Table 13 provides an overview of participant demographics. Four of these participants were recruited via Ulster University in Northern Ireland and were aged between 26-61. Two elderly participants were recruited via the Healthy Age Initiative in Umeå, Sweden and were 71 years of age, one male, one female (Figure 22).

Table 13 Demographic information of participants

Participant ID	Gender	Age	Height (cm)	Weight (Kg)	Country
1	Male	61	179	68.7	NIR
2	Female	52	158	N/A	NIR
3	Male	26	186	N/A	NIR
4	Male	71	172	73.0	SWE
5	Female	71	164	64.0	SWE
6	Male	N/A	177	N/A	NIR

The study will continue to collect data on healthy participants aged 18-75 with no previous conditions i.e. no neurological or knee/hip surgery history, so that a robust gait model can be formed.

Figure 22: Female elderly participant from Umeå Sweden.

Protocol

Initial feasibility testing was conducted on 6 participants to evaluate that the wearable systems were fit for purpose in terms of robustness, reliability, usability, comfort, repeatability, and set-up time.

Walking activities have been designed to include walking at differing speeds, walking on a treadmill versus over the ground. The study also included turns as often it can be overlooked in GA. Given that the potential use case scenario for this research will be typically elderly post-stroke survivors some slower walking speeds were included.

Participants performed two walking activities. The first involved walking indoors over the ground on a flat smooth surface within a gait laboratory environment for 10m and turning. This activity was repeated for 2.5 mins at a comfortable walking speed self-selected by the participant.

The second activity required participants to walk for 2.5 minutes at 4 different speeds on a treadmill for a total walking time of 10 minutes. The four speeds using the treadmill with zero incline were: very slow (0.5m/s), slow (0.75m/s), medium (a comfortable speed self-selected by the participant, either 1m/s or 1.25m/s) and fast (1.5m/s). The fastest walking speed test is like the 2-minute walking test which is a clinical assessment.

To configure the IMC system called Xsens several body measurements were taken to calibrate the system as shown in Figure 23.

Figure 23: Xsens body measurements

Note that the walking over the ground activity was not performed on the two elderly participants Umeå.

Devices overview

The following research devices were used to collect gait data/information for the study.

Xsens

The Xsens IMC system (Figure 24) is a wireless body area network of 17 Inertial Measurement Units (IMU) which consists of three sensors (accelerometer, gyroscope, and magnetometer). The software algorithms are capable of producing segment and joint information for the whole body however for this study the system was restricted to operate on the lower body using just 7 IMUs.

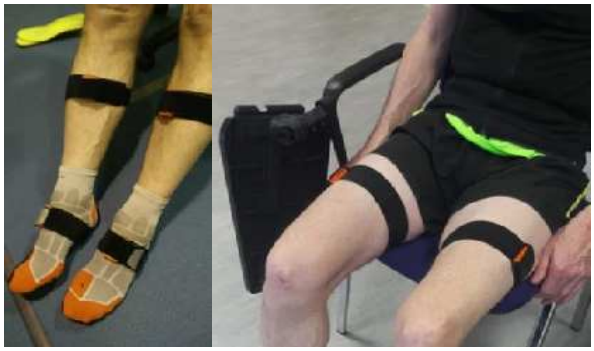


Figure 24: Xsens IMC system showing IMU positions for foot, lower leg and upper leg.

Walkinsense

The Walkinsense smart insole (Figure 25) consists of 8 force sensitive resistors to form a sensor network which is attached to a regular shoe insole. This device provides plantar foot pressure data which can be used to complement information from the IMC.

Figure 25: Walkinsense smart insole showing sensor layout, show insertion and leg attachment.

The device is worn around the ankle and provide wireless transmission of data via Bluetooth.

Device Calibration

The Xsens IMC requires calibration prior to each recording to ensure the data/information collected is of a high quality. Participants are required to stand in a T pose and when instructed walk forward for 2 metres and turn and walk back to their starting position and finally turn to face their original starting direction. If there is any electromagnetic interference within the recording environment, then the calibration may need to be completed more than once. Alternatively, if there is significant electromagnetic interference then the activities may need to be moved or equipment responsible for the inference switched off.

The Walkinsense devices do not require any calibration but do require both the hardware and software to be checked to ensure that data is streaming and being appropriately collected by the software.

Results

In this study only data/information from the Xsens IMC has been analysed. Please see the future work section to read about plans for the Walkinsense insole data.

The Xsens system has already been well validated against OMC systems and has been reported to have a coefficient of multiple correlation > 0.96 for all joints during flexion/extension for level walking activities [55]. However, configuration, calibration and positioning sensors can have varying effects on the quality of the data collected. Repeatability was explored for in-day testing of the Xsens without doffing/donning the sensors.

The analysis to date has looked at presenting gait information as phase portraits which typically plot a joint angle against its velocity. In this study we explored the knee joint but the same principles can be applied across many of the other joints or body segments.

Two phase portraits shown in Figure 26 are highly correlated and present knee angle against angular velocity of the lower limb for participant 1. These phase portraits show the dynamic nature of the knee during walking on a treadmill at a 5.4 m/s for 2 minutes. A single gait cycle is represented by one phase which can be seen as a closed loop. The phases are plotted on top of each other as each gait cycle is repeated, it shows high levels of correlation but with some dynamic and chaotic variations. The variation between gait cycles in the first test can be seen in red while the blue lines show the variation of gait cycles in the second test. Since both tests were recorded within 30 minutes and under the same conditions the variation which is expected to be minimal between walks can be observed by comparing red and blue lines.

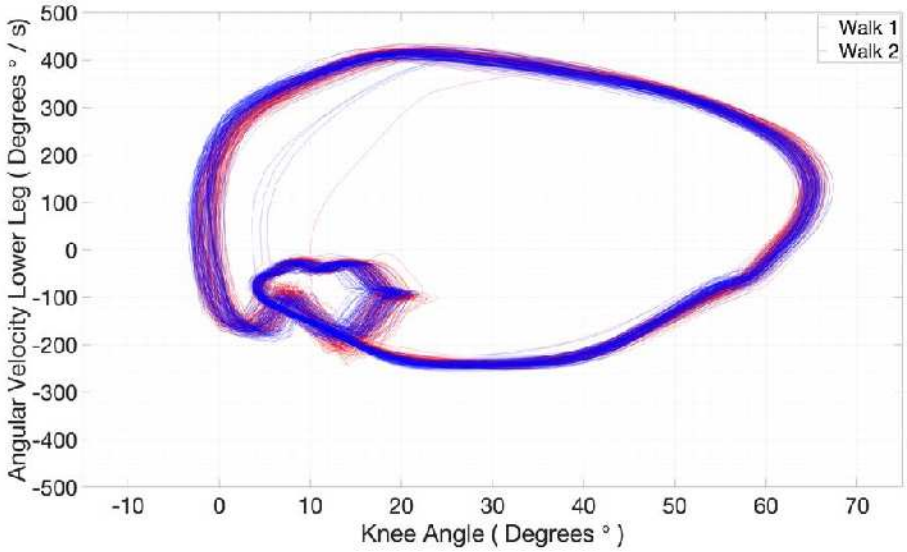


Figure 26: A highly correlated phase portrait of knee angle versus angular velocity of the lower right leg for Participant 1 during two separate tests while walking on treadmill at 5.4 m/s for 2 minutes.

To provide a statistical measure of correlation an average gait cycle was computed for both walks and a correlation coefficient calculated by comparing both average gait cycles. The average phase portraits can be seen in Figure 27, these are highly correlated as expected (r=0.9993).

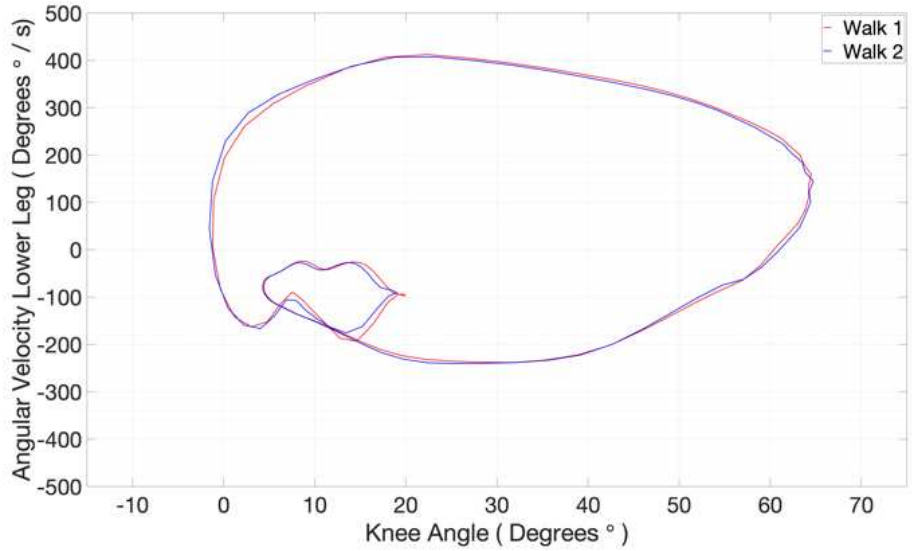


Figure 27: A highly correlated (r=0.9993) phase portrait, it represents an average gait cycle from two walks by participant 1.

The next stage of GA was to compare all of the walking activities for participants across a number of different walking speeds; this involved 2-minute treadmill walks at 1.8 km/h, 2.7 km/h, 3.6 km/h or 4.5 km/h and 5.4 km/h and a final 2 minute walk over the ground for 10 metres with a turn. The phase portraits of these walking activities can be seen in Figure 28 for participant 1. Initial observations show correlation between walking speed and the area enclosed within the phase portraits. A greater range of motion and increased angular velocity should result in an increased area within the curves. Additionally, as the speed increases the variability in the phase portraits reduces to produce a more rhythmic and stable gait cycle, this is particularly for true for walking activities on the treadmill. It seems that this effect is caused by a combination of the treadmill and higher walking speeds.

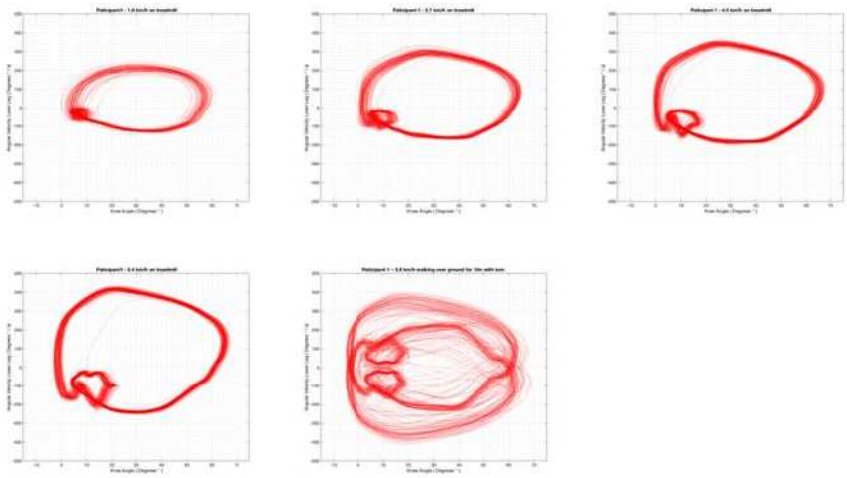


Figure 28: Phase portraits of knee angle versus angular velocity of the lower right leg for Participant 1 at four walking speeds (1.8 km/h, 2.7 km/h, 4.5 km/h, 5.4 km/h) on a treadmill walking over the ground at 3.9 km/h.

To quantify the relationship between the area of the phase portraits and speed an average phase portrait was computed for each walk. These were then used as a basis to calculate an average area of each phase portrait and to plot these against speed to quantify any relationship and how it may change across the cohort. Observation of Figure 29 shows a common pattern of increased area equating to increased speed. Participants 4 and 5 were elderly and it's interesting that they were able to maintain walking speeds of 1.8 km/h, 2.7km/h and 3.3km/h with a reduced area.

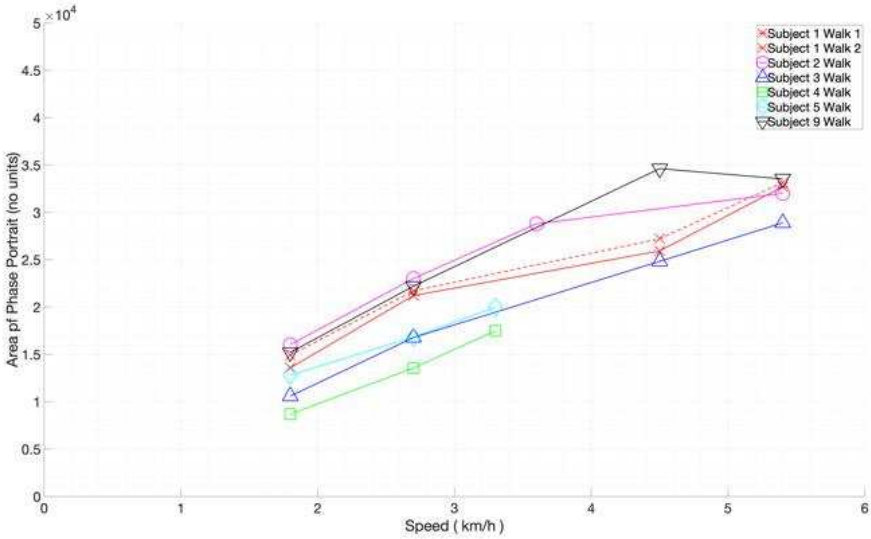


Figure 29:Walking activity for six participants showing average phase plot area against speed.

Comparing treadmill walking activities against over the ground walking in Figure 28 shows that the variation in knee angle is more apparent while walking over ground. As the walking over the ground activity included turning 180 degrees every 10 metres there are a significant number of turns (n~=16) within a 2-minute period. Therefore, it makes it more difficult to attribute the increased knee angle variation as a direct result of walking over the ground. The second noticeable feature of the phase portraits for walking over the ground is that there is a mirroring effect which results in two prominent distinct phases. These are a direct result of walking in two opposite directions and may be combined

into a single phase if it is possible to adjust the data to accommodate walking direction. This would be a useful analysis feature as it would allow all walking activities to be easily analyzed and compared irrespective of walking direction.

Summary

Gait analysis of the normal population and of different pathologies is an area of research that is expanding rapidly. There are several competing technologies that can provide gait information, two such competing sets are research gait lab technology and wearable technology. The former tends to be more expensive, less flexible and with longer setup times often requiring specialised training. With recent advances, wearable technology can offer a cheaper, more accessible, less restrictive, and easier to use option without comprising on the accuracy or quality of the information. This is particularly true of recent advances of IMC systems.

The Xsens IMC system captured kinematic data from walking. Due to the exploratory nature of this study only the dynamic nature of knee angle during walking was considered; the gait variation across a number of walking speeds on a treadmill and walking over the ground and gait variation across the population were assessed. Future research aims to build a computational model that can be used to assess a user's gait during ambulation. A large set of features will be generated to serve as input to the model and as such can be configured in multiple ways via feature selection to ascertain the optimum model and as a result what are the optimum features, technologies, and sensors.

Lessons learned and future perspectives

Frailty is an important health issue that affects older people, increases with increasing patient age, and is a multifactorial phenomenon associated with adverse health outcomes. We agree with Rockwood et al. that frailty is an umbrella term, a construct for understanding frailty conceptualizes the frail older person as a complex system on the threshold of breakdown[56].

During the project we have reached an understanding during our demonstrator trials that have in some part worked with the conceptualization of frailty as a complex system failure and how it might help us understand existing evidence regarding the aetiology and prevention of outcomes and, most importantly, how it might impact our clinical practice. The demonstrators have helped us to increase our understanding of frailty and its relation to co-morbidities and how wearables could aid in early detection and monitoring.

We have been able to develop sensory based frailty tests aiming on specific topics such as balance and the risk of falling, sedentary behaviour as risk factor for depression and gait analysis and fall risk. What becomes evident is that there is a plethora of knowledge on research level that needs to be developed to drive innovation and implementation of wearables in regular healthcare.
We have also investigated specific functional tests for frailty measurements such as physiological testing for falls and hand symptoms in Parkinson's disease. Wearables have a great potential to develop better testing and monitoring for frail patient group but there is a need to validate each subgroup thoroughly and a great need for research and SMEs to collaborate to get high quality products that target the need of the healthcare.

We have also investigated the validation processes for wearable sensor systems. To date, generally, there is a lack of validation and a lack of statement that data collection is secure and that the devices in play are working as expected in many of commercially available systems. There is thus a need to investigate validation of wearables, it is not easy to penetrate the market which has become evident in our testing.
We have focused on the process together with SMEs and see a need for more collaboration between researchers, SMEs and health care in the future to validate and develop wearable products.
We have also evaluated implementation strategies of use of wearables in home environment for home dwelling elderlies. There is potential in wearables as health care improvement strategies that are particularly interesting for rural areas. Rural public health centers could potentially service relatively larger areas of dispersed older adults with limited resources using wearables but there is cost associated with implementation both in product procurement and education for staff.

In summary, there are clear challenges to developing technologies for frailty. During the project time we have identified three major difficulties that needs addressing:

1. Objective definition linked to "hard" endpoints such as mortality, falls, fractures.
2. Agreement on appropriate outcomes for efficacy and registration trials for wearables used in the field.
3. Need to expand potential targeted areas of interest.

References

[1] Rodríguez-Mañas L, García-Sánchez I, Hendry A et al (2018) Key messages for a frailty prevention and management policy in Europe from the Advantage Joint Action Consortium. J Nutr Health Aging. https://doi.org/10.1007/s12603-018-1064-y

[2] Johansson J, Jarocka E, Westling G, **Nordström A**, Nordström P. Predicting incident alls: Relationship between postural sway and limits of stability in older adults. Hum Mov Sci 2019 10;66:117-123

[3] ickerill ML, Harter RA. Validity and reliability of limits-of-stability testing: a comparison of 2 postural stability evaluation devices. J Athl Train 2011;46:600-6.

[4] Johansson, J., Nordstrom, A., & Nordstrom, P. (2015). Objectively measured physical activity is associated with parameters of bone in 70-year-old men and women. Bone, 81, 72–79.

[5] Johansson, J., Nordström, A., Gustafson, Y., Westling, G., & Nordström, P. (2017). Increased postural sway during quiet stance as a risk factor for prospective falls in community-dwelling elderly individuals. Age and Ageing, 1–6.

[6] Forth, K.E., Fiedler, M., & Paloski, W. (2011). Estimating functional stability boundaries for bipedal stance. *Gait & posture, 33 4*, 715-7 .

[7] Johansson, J., Nordström, A., & Nordström, P. (2015). Objectively measured physical activity is associated with parameters of bone in 70-year-old men and women. Bone, 81, 72–79. https://doi.org/10.1016/j.bone.2015.07.001

[8] Holmquist S, Mattsson S, Schele I, Nordström P, Nordström A. Low physical activity as a key differentiating factor in the potential high-risk profile for depressive symptoms in older adults. Depress Anxiety. 2017 Sep;34(9):817-825

[9] Yesavage, J., & Sheikh, J. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*(1-2), 165–173.

[10] Berginstrom, N., Johansson, J., Nordstrom, P., & Nordstrom, A. (2015). Attention in Older Adults: A Normative Study of the Integrated Visual and Auditory Continuous Performance Test for Persons Aged 70 Years. *Clinical Neuropsychologist, 29*(5), 595-610.

Holmquist, S., Mattsson, S., Schele, I., Nordström, P., & Nordström, A. (2017). Low physical activity as a key differentiating factor in the potential high-risk profile for depressive symptoms in older adults. *Depression and Anxiety*, *34*(9), 817-825.

[11] Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., & Friedrich, F. (2006). The criterion validity of the Geriatric Depression Scale: A systematic review. *Acta Psychiatrica Scandinavica*, *114*(6), 398–410

[12] Almeida, O. P., & Almeida, S. A. (1999). Short versions of the geriatric depression scale: A study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. *International Journal of Geriatric Psychiatry*, *14*(10), 858–865.

[13] Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine & Science in Sports & Exercise, 30*(5), 777-781.

[14] Podsiadlo, D., & Richardson, S. (1991). The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, *39*(2), 142–148. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x

[15] Holmquist, S, Mattsson, S, Schele, I, Nordström, P, Nordström, A. (2017) Low physical activity as a key differentiating factor in the potential high-risk profile for depressive symptoms in older adults. Depress Anxiety. 34: 817– 825. https://doi.org/10.1002/da.22638

[16] Braun, V., & Clarke, V. (2013). Successful qualitative research: A practical guide for beginners. London: SAGE.

[17] T. Doi, S. Hirata, R. Ono, K. Tsutsumimoto, S. Misu, H. Ando (2013). Theharmonic ratio of trunk acceleration predicts falling among older people: results of a 1-year prospective study., Journal of neuroengineering and rehabilitation 10 (1) 7.doi:10.1186/1743-0003-10-7

[18] U. Laessoe, H. C. Hoeck, O. Simonsen, T. Sinkjaer, M. Voigt, Fall risk in an active elderly population - Can it be assessed?, Journal of Negative Results in BioMedicine 6 (2007) 2.doi:10.1186/1477-5751-6-2.

[19] M. Marschollek, A. Rehwald, K. H. Wolf, M. Gietzelt, G. Nemitz,H. Meyer zu Schwabedissen, R. Haux, Sensor-based fall risk assessment- an expert 'to go', Methods of Information in Medicine 50 (5) (2011) 420–426.doi:10.3414/ME10-01-0040

[20] K. Paterson, K. Hill, N. Lythgo, Stride dynamics, gait variability and prospective falls risk in active community dwelling older women, Gait and Posture 33 (2) (2011) 251–255.doi:10.1016/j.gaitpost.2010.11.014.

[21] R. Schwesig, D. Fischer, A. Lauenroth, S. Becker, S. Leuchte, Can falls be predicted with gait analytical and posturographic measurement systems? A prospective follow-up study in a nursing home population, Clinical Re-habilitation 27 (2) (2013) 183–190.doi:10.1177/0269215512452880.

[22] B. R. Greene, E. P. Doheny, C. Walsh, C. Cunningham, L. Crosby,R. A. Kenny, Evaluation of Falls Risk in Community-Dwelling Older Adults Using Body-Worn Sensors, Gerontology 58 (5) (2012) 472–480.doi:10.1159/000337259.

[23] J. Howcroft, J. Kofman, E. D. Lemaire, Prospective Fall-Risk Prediction Models for Older Adults Based on Wearable Sensors, IEEE Transactions on Neural Systems and Rehabilitation Engineering 25 (10) (2017) 1812–1820.doi:10.1109/TNSRE.2017.2687100.

[24] B. C. Kwok, R. A. Clark, Y. H. Pua, Novel use of the Wii Balance Board to prospectively predict falls in community-dwelling older adults, Clinical Biomechanics 30 (5) (2015) 481-484.doi:10.1016/j.clinbiomech.2015.03.006.

[25] M. Schwenk, K. Hauer, T. Zieschang, S. Englert, J. Mohler, B. Najafi, S. .Hauer, . Zieschang, . Englert, . Mohler, . Najafi, Sensor-Derived Physical 23 Activity Parameters Can Predict Future Falls in People with Dementia, Gerontology 60 (2014) 483–492.doi:10.1159/000363136.

[26] Y. Liu, S. J. Redmond, T. Shany, J. Woolgar, M. R. Narayanan, S. R. Lord,N. H. Lovell, Validation of an accelerometer-based fall prediction model, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 4531–4534.doi:10.1109/EMBC.2014.6944631.

[27] M. Gietzelt, F. Feldwieser, M. Govercin, E. Steinhagen-Thiessen, M. Marschollek, A prospective field study for sensor-based identification of fall risk in older people with dementia, Informatics for Health and Social Care 39 (3-4) (2014) 249–261.doi:10.3109/17538157.2014.931851.

[28] M. J. Mohler, C. S. Wendel, R. E. Taylor-Piliae, N. Toosizadeh, B. Na-jafi, Motor Performance and Physical Activity as Predictors of Prospective Falls in Community-Dwelling Older Adults by Frailty Level: Ap-plication of Wearable Technology, Gerontology 62 (6) (2016) 654–664.doi:10.1159/000445889.

[29] K. S. Van Schooten, M. Pijnappels, S. M. Rispens, P. J. Elders, P. Lips,A. Daffertshofer, P. J. Beek, J. H. Van Die¨en, Daily-life gait quality as predictor of falls in older people: A 1-year prospective cohort study, PLoSONE 11 (7) (7 2016).doi:10.1371/journal.pone.0158623.

[30] A. Nait Aicha, G. Englebienne, K. van Schooten, M. Pijnappels, B. Kr¨ose,Deep Learning to Predict Falls in Older Adults Based on Daily-Life TrunkAccelerometry, Sensors 18 (5) (2018) 1654.doi:10.3390/s18051654.

[31] J. Howcroft, E. D. Lemaire, J. Kofman, Prospective elderly fall prediction by older-adult fall-risk modeling with feature selection, Biomedical Signal Processing and Control 43 (2018) 320–328.doi:10.1016/j.bspc.2018.03.005.

[32] Scott D, Johansson J, McMillan L, Nordstrom P, Nordström A. Mid-calf skeletal muscle density and its associations with physical activity, bone health and incident 12-month falls in older adults: The Healthy Ageing Initiative Bone. 2019 Mar;120:446-451. doi: 10.1016/j.bone.2018.12.004. Epub 2018 Dec 8.

[33] Lang AE, Lozano AM. Parkinson's disease. First of two parts. The New England journal of medicine. 1998; 339(15):1044-53.

[34] Jankovic J. Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery, and psychiatry. 2008; 79(4):368-76.

[35] Fahn S. Unified Parkinson's disease rating scale. Recent development in Parkinson's disease. 1987.

[36] Stolwijk, C. et al. (2016) 'Global Prevalence of Spondyloarthritis: A Systematic Review and Meta-Regression Analysis', Arthritis Care & Research. John Wiley & Sons, Ltd, 68(9), pp. 1320–1331. doi: 10.1002/acr.22831.

[37] Mancini, M. and Horak, F. B. (2010) 'The relevance of clinical balance assessment tools to differentiate balance deficits', European journal of physical and rehabilitation medicine, 46(2), pp. 239–248. Available at: https://www.ncbi.nlm.nih.gov/pubmed/20485226.

[38] Wells, G. et al. (2014) 'Updating the OMERACT filter: discrimination and feasibility', The Journal of rheumatology. 2014/04/01, 41(5), pp. 1005–1010. doi: 10.3899/jrheum.131311.

[39] Theobald, P. S., Jones, M. D. and Williams, J. M. (2012) 'Do inertial sensors represent a viable method to reliably measure cervical spine range of motion?', Manual Therapy, 17(1), pp. 92–96. doi: https://doi.org/10.1016/j.math.2011.06.007.

[40] Ronchi, A. J. et al. (2008) 'A reliability study of the new Back Strain Monitor based on clinical trials', in 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 693–696. doi: 10.1109/IEMBS.2008.4649247.

[41] Laird, R. A., Kent, P. and Keating, J. L. (2012) 'Modifying patterns of movement in people with low back pain -does it help? A systematic review', BMC musculoskeletal disorders. BioMed Central, 13, p. 169. doi: 10.1186/1471-2474-13-169.

[42] Nielsen, J., (1993). Usability engineering. Boston: Academic Press.

[43] Hult A, Johansson J, Nordström P, **Nordström A**. Objectively measured physical activity in older adults with and without diabetes. Accepted for publication in Clinical Diabetes.

[44] World Health Organization. Global recommendations on physical activity for health. Geneva, Switzerland, World Health Organization, 2010.

[45] Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008;40:181–188

[46] Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998;30:777–781

[47] A. Heikkilä, N. Sevander-Kreus, A. Häkkinen, M. Vuorenmaa, P. Salo, P. Konsta and J. Ylinen, "Effect of total knee replacement surgery and postoperative 12 month home exercise program on gait parameters," *Gait & Poature,* vol. 53, pp. 92-97, 2017.

[48] P. Ornettia, J.-F. Maillefert, D. Laroche, C. Morisset, M. Dougados and L. Gossec, "Gait analysis as a quantifiable outcome measure in hip or knee osteoarthritis: A systematic review," *Joint Bone Spine,* vol. 77, pp. 421-425, 2010.

[49] M. Zagoa, C. Sforza, D. R. Bonardi, E. E. Guffanti and M. Galli, "Gait analysis in patients with chronic obstructive pulmonary disease: a systematic review," *Gait & Posture,* vol. 61, pp. 408-415, 2018.

[50] B. Hanley, C. B. Tucker and A. Bissas, "Differences between motion capture and video analysis systems in calculating knee angles in elite-standard race walking," *Journal of Sport Science,* vol. 36, no. 11, pp. 1250-1255, 2018.

[51] M. E. Fernando, R. G. Crowther, P. A. Lazzarini, K. S. Sangla, P. Buttner and J. Golledge, "Gait parameters of people with diabetes-related neuropathic plantar foot ulcers," *Clinical Biomechanics,* vol. 37, pp. 98-107, 2016.

[52] I. Bouchrika, M. Goffredo, J. Carter and N. Mark, "On Using Gait in Forensic Biometrics," *Journal of Forensic Sciences,* vol. 56, no. 4, pp. 882-889, 2011.

[53] Y. He, J. Zhang, H. Shan and L. Wang, "Multi-Task GANs for View-Specific Feature Learning in Gait Recognition," *IEEE Transactions on Information Forensics and Security,* vol. 14, no. 1, pp. 102-113, 2019.

[54] D. J. Gladstone, C. J. Danells and S. E. Black, "The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties," *Neurorehabilitation and Neural Repair,* vol. 16, no. 3, pp. 232-240, 2002.

[55] J.-T. Zhang, A. C. Novak, B. Brouwer and Q. Li, "Concurrent validation of Xsens {MVN} measurement of lower limb joint angular kinematics," *Physiological Measurement,* vol. 34, no. 8, pp. 63-69, 2013.

[56] Rockwood K. Why geriatric medicine? In: Rai S, Mulley G, editors. Elderly Medicine: A training guide. 2nd edn. London: Taylor & Francis; 2007.