DATA SCIENCE ENABLED REHABILITATION

IOANNIS VOURGANAS

Submitted For The Degree Of Doctor Of Philosophy

Department of Electronic & Electrical Engineering
University Of Strathclyde
Glasgow Scotland UK

2020

Declaration of Authenticity and Author's Rights

‘This thesis is the result of the author’s original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.’

Signed: ioannis j vourganas Date: 21/05/2021

Στη Λητώ.

Acknowledgements

Even though this page cannot convey the gratitude and thanks owed to all mentioned, it is a hamble attempt to recognise their significant contributions in materialise this thesis.

First of all, I would like to take this opportunity to thank my supervisor Dr Vladimir Stankovic and the Department of Electronic and Electrical Engineering for giving me the opportunity of starting a Doctorate and believed in me. His contribution was of paramount importance for completing my researcher and writing this thesis without significant delays through his insightful guidance despite several personal obstacles faced in the way. He is a source of inspiration, patience and understanding.

I would like to extend a very special thanks to Dr Lina Stankovic for her significant contribution and help throughout my PhD journey. Although she was not assigned as my official supervisor, she was offering her support and feedback all the time when I needed her and helped me significantly in shaping my work through the publications with her guidance and advice.

Special thanks also to Dr Andrew Kerr who helped guide the initial stages of my research by providing guidance for important aspects in the biomedical engineering domain, which was an entirely new field for me.

During a significant milestone in my second year Prof. Anthony Gachagan and Prof. Campbell Booth provided very supportive feedback that helped me find the courage to keep moving forward and have supported and inspired me throughout my academic career.

An important thanks to Lito, to the most important person in my life, who was always there to offer her advice and it was her support that kept me going.

Finally, I would like to thank the University of Strathclyde for providing training, continuous professional development, and all the facilities for the experiments; particularly the centre of National Orthotics and Prosthetics for their valuable help during the experiments.

Table of Contents

Acknowledgements iii

Nomenclature xiii

Publications xv

Abstract xvi

1 Introduction 1

1.1 Types of Stroke 1

1.2 Stroke Effects 2

1.3 Differentiation of Gender, Age and Ethnicity 4

1.4 Phases of Stroke and Patient Differentiation 4

1.5 Rehabilitation 5

1.6 Motivation Analysis 7

1.7 Chapter Summary 10

2 Motivation and Research Questions 12

2.1 Motivation 12

2.2 Socioeconomical Background & Benefit 13

2.3 Contribution & Research Question 14

2.4 Aims and Objectives 16

2.5 Thesis Organisation 17

3 Review, Taxonomy and Criteria for Home-Based Rehabilitation Systems. 19

3.1 Systematic Literature Search Methodology 19

3.2 Systematic Literature Search Results 20

3.3 Overview of Engineering in Rehabilitation 22

3.4 Limitations of Rehabilitation Technologies. 25

3.5 Home Rehabilitation Systems 26

3.6 Taxonomy......29

3.7 Contribution of Monitoring in Home Rehabilitation 32

3.8 Critical Evaluation of Engineering Approaches 35

3.9 Chapter Conclusion 39

4 Design of Home-Based Rehabilitation System & Evaluation Methodology41

4.1 Introduction 41

4.2 Patient Evaluation Tests 42

4.2.1 Automated Timed Up and Go Test 43

4.2.2 Automated Five Time Sit to Stand Test 45

4.3 Methodology 46

4.3.1 Participants 47

4.3.2 Statistical Data Analysis Measures 49

4.3.3 Experiments 50

4.3.4 Timed Up and Go Test Experiments 51

4.3.5 Five Time Sit to Stand Test 53

4.3.6 Mapping of the System......54

4.4 Experimental Results 56

4.4.1 Questionnaire......60

4.5 Chapter Discussion 62

4.6 Chapter Conclusion 63

5 Machine Learning-Enabled Individualised Home-Based Rehabilitation ... 65

5.1 Background 66

5.1.1 State of the Art in Applied Artificial Intelligence 66

5.1.2 ART AI for Patient-Centric Home-Based Rehabilitation 68

5.1.3 Review of ART AI in Medical Systems 70

5.1.4 ART AI in Home-Based Rehabilitation Systems 71

5.1.5 ART Design Considerations 76

5.2 Methodology 77

5.2.1 Training Dataset Design 78

5.2.2 Experimental Dataset 80

5.2.3 Synthetic Dataset 82

5.2.4 Evaluation of ML Methods for Supervised Classification 89

5.2.5 Hybrid ML Approach for Individualised ART-Driven Rehabilitation ..100

5.2.6 ART AI Design Approach 102

5.3 Evaluation and Results 104

5.4 Chapter Conclusion 110

6 Discussion 113

7 Conclusion 116

7.1 Limitations 117

7.2 Future Work 117

References **120**

Appendices **160**

Table of Figures

Figure 1 Pathology of stroke subcategories defined by ICF; sourced from [3]. 2

Figure 2 Rehabilitation process, created based on information available in [3]. 6

Figure 3 Goal setting theory created based on information provided in [36]. 7

Figure 4 Factors which contribute to self-efficacy enhancement created based on [41], [42]. 8

Figure 5 Factors for motivation enhancement created based on information available [36], [43]. 9

Figure 6 The factors which can cause loss or decrease of motivation based on information available in [44][45]. 9

Figure 7 Thesis organisation and significance denoted by the circle diameter with respect to addressing the research questions, aims and objectives of the thesis. 18

Figure 8 PRISMA diagram of the review stages. 21

Figure 9 A simulation of the human system as an information system for motor control sourced from [47]. .... **Error! Bookmark not defined.**

Figure 10 Relationship between cost and complexity sourced from [28]. 22

Figure 11 Taxonomy of Rehabilitation Systems for stroke patients, VR: Virtual Reality; tDCS: Transcranial Direct Current Stimulation; TMS: Transcranial Magnetic Stimulation; WBSN: Wearable Body Sensor Network Systems 30

Figure 12 Monitored qualities for the health and care applications and the need for additional sensorial input to monitoring devices, source [135]. 33

Figure 13 Subject performing TUG with the deployed system at National Centre of Prosthetics and Orthotics. The head is covered for subject privacy. 47

Figure 14 Timed Up and Go Test Experiment 51

Figure 15 Timed Up and Go Test Experiment Simulating Difficulty to Stand 52

Figure 16 Timed Up and Go Test Experiment Simulating Difficulty to Walk.......... 52

Figure 17 Timed Up and Go Test Experiment Simulating Difficulty to Turn........... 53

Figure 18 Five Times Sit to Stand Test Stages Repeated Five Times For Each Experiment 54

Figure 19 Location and placement of system components for TUG. 55

Figure 20 Location and placement of system components for FTSTS. 56

Figure 21 TUG Normal Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression.......... 58
Figure 22 TUG Difficult to Stand Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression 58
Figure 23 TUG Difficult to Turn Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression 59
Figure 24 TUG Difficult to Walk Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression 59
Figure 25 TUG Fast Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression 59
Figure 26 FTSTS Slow Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression.......... 59
Figure 27 FTSTS Fast Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression.......... 59
Figure 28 Questionnaire Results in terms of start-rating for each of the questions 1 to 5. Category 5 corresponds to the highest (good) rating while category 1 corresponds to the lowest (bad). 61
Figure 29 Responsible AI.......... 67
Figure 30 AmI characterisation based on[234] 72
Figure 31. (a) Correlation matrices for TUG experiment, (b) TUG synthetic data, (c), FTSTS experiment and (d) FTSTS synthetic data 88
Figure 32 (a) TUG exp. data after PCA analysis reducing the features to 2, (b) TUG synth, (c) FTSTS exp., and (d) FTSTS synth. respectively. 88
Figure 33 Different SVM Kernels where RBF is Radial Basis Function. 89
Figure 34 XGBoost parameters. Parameters that were tuned using grid search are highlighted in yellow.......... 92
Figure 35 Impurity-based feature importance of the xgboost forests based on mean decrease in impurity as calculated by XGB Classifier.feature_importances_for (a) TUG exp., (b) TUG synth., (c) FTSTS exp., and (d) FTSTS synth. 99
Figure 36 Graphical representation of the hybrid stacking model combining Difficulty prediction and Condition prediction for a medical test. The data refers to experimental plus synthetic data. 100

Figure 37 ML Model Evaluation Methodology Flowchart where “Evaluation” stands for the critical evaluation of the results presented in this thesis. Phase 1 Results were presented in Section 5.2.4.6. Phase 2 is discussed in this Section. 104
Figure 38 (a) Prediction accuracy improvement over time for TUG with difficulty to walk (one of the difficulties that are miss classified in 30% of the cases) and Geriatric classification used 5000 times by individual 1, (b) Prediction accuracy improvement over time for TUG performed Normally (anon-difficulty that is miss classified as Fast in 71% of the cases) 1000 times by individual 3 classified as Dementia Severe..... 107
Figure 39 Confusion matrices for the hybrid model presented separately for (a) TUG difficulty, (b)FTSTS difficulty, (c) TUG conditions, and (d) FTSTS conditions.... 108
Figure 40 Chapter 5's contributions 112

Table of Tables

Table 1 Disorders which can be caused by stroke and body functions affected post-stroke; sourced from [3][6]. 3
Table 2 Participation and activity restrictions caused by stroke sourced from [3][6]. 3
Table 3 Results of the literature search before and after inclusion criteria were applied. 21
Table 4 Diagnostic capabilities for health conditions using smart meter data, source [135]. 34
Table 6 Summative assessment of the reviewed systems 38
Table 6 Medical Tests to Assess Patient Activities 43
Table 7 Participant recorded parameters. 47
Table 8 Timed Up and Go Test and Five Time Sit to Stand Experiments 53
Table 9 Location and placement parameters for each participant of TUG. 55
Table 10 Location and placement parameters for each participant of FTSTS. 56
Table 11 Timed Up and Go (TUG) Characterisation Of The Tests X (σ). The Results Are Given in Seconds 57
Table 12 Five Time Sit to Stand (FTSTS) Characterisation Of The Tests X (σ). The Results Are Given in Seconds 57
Table 13 Characterisation Of The Tests: Coefficient of Vitiation Percentage (%) 57
Table 14 Characterisation of The Tests: Percentage Error and correlation Results 60
Table 15 Data Quality dimensions[254] and our approach to addressing them. 79
Table 16 Summary of datasets with their respective features and labels 85
Table 17 Optimal hyperparameter tunined values for XGboost for all TUG and FTSTS datasets. 93
Table 18 5-fold cross validation accuracy results of ML models. Accuracy is calculated as the fraction of correct predictions using `sklearn.metrics.accuracy_score` 97
Table 19 Accuracy results of XGBoost, KNN, and hybrid models using the test dataset 106
Table 20 Accuracy results of XGBoost, & KNN models using the train dataset (self-evaluation) 106

Table 21 Hybrid Model Confusion Matrix Metrics 109
Table 22 Hybrid model most frequent association of difficulty and condition for the experimental test dataset of the 8 participants 110

Nomenclature

Acronym	Definition
3D	Three Dimension
5-CV (5-F)	Five Fold Cross Validation
ADL	Assistive Daily Living
AI	Artificial Intelligence
AmI	Ambient Intelligence
ARAT	Action Research Arm Tests
ART	Accountable Responsible Transparent
BESS	Balance Error Scoring System
BMI	Body Mass Index
BPNN	Back Propagation Neural Network
CNN	Convolutional Neural Network
COVID-19	Corona Virus Disease 2019
DSS	Decision Support System
ESD	Early Support Discharge
EU	European Union
EXP	Experiment
FN	False Negative
FP	False Positive
FTSTS	Five Time Sit To Stand Test
HCI	Human Computer Interaction
ICC	Intraclass Corelation Coefficient
ICF	International Classification of Functioning
IDS	Infusion Detection System
IEEE	the Institute of Electrical and Electronics Engineers
KNN	K Nearest Neighbours
IoT	Internet of Things
LOA	Limit of Agreement
ML	Machine Learning
NHS	National Health System
NILM	None-Intrusive Load Monitoring
NN	Neural Network
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PE	Percentage Error
PIR	Passive Infrared
RBF	Radial Basis Function
RQ	Research Question
SLT	Speech and Language Therapy
SMOTE	Synthetic Minority Oversampling Technique
SVC	Support Vector Classification
SVM	Support Vector Machine
SVR	Support Vector Regression
SYNTH	Synthetic
TEX	Ethylebenzene and Xylene
TIA	Transient Ischemic Attack
TN	True Negative
TP	True Positive
TUG	Timed Up and Go

tDCS	Transcranial Direct Current Stimulation
TMS	Transcranial Magnetic Stimulation
UK	United Kingdom
VR	Virtual Reality
WBSN	Wearable Body Sensor Network
XAI	Explainable Artificial Intelligence
XGBOOST (XGB)	Extreme Gradient Boosting

Publications

- Factors That Contribute to the Use of Stroke Self-Rehabilitation Technologies: A Review I Vourganas, V Stankovic, L Stankovic, A Kerr, *JMIR Biomedical Engineering,* 4 (1), e13732, 2019

  The above publication is a concise summary of Chapters 1 and 3, establishes the identified gap in home rehabilitation systems, and the criteria for an effective system.
- Evaluation of home-based rehabilitation sensing systems with respect to standardised clinical tests I Vourganas, V Stankovic, L Stankovic, AL Michala, *Sensors* 2020, 20, 26.

  The above publication is a summary of the main findings and methods presented in Chapter 4. This establishes that the system records reliable information and follows the criteria identified in the previous publication.
- Individualised Responsible Artificial Intelligence for home-based rehabilitation I Vourganas, V Stankovic, L Stankovic, *Sensors* 2021, 21, 2.

  The above publication is a concise summary of Chapter 5. It presents the novel hybrid Machine Learning approach and relates individualisation, monitoring and co-morbidity identification with ART AI design considerations. This concludes the proposed system and addresses the full scope of the criteria identified in this thesis.

Abstract

Stroke is a main cause of impairment/disability. More stroke survivors undergo unsupervised home rehabilitation. Autonomous self-rehabilitation systems using sensing and machine learning are not tailored to patients' needs.

Based on a systematic narrative literature review, home-based rehabilitation systems were taxonomized and new design criteria were formulated for increased patient engagement enhancement and individualism. No system that addresses all the criteria was found in literature. An in-house low-cost home-based rehabilitation Ambient Intelligence (AmI) system was deployed meeting the criteria, and an accuracy evaluation method proposed, in line with medically approved tests. The Timed Up and Go (TUG) and Five Time Sit To Stand (FTSTS) tests evaluate daily living activity performance in the presence/development of comorbidities. The AmI-driven system complies with Accountability, Responsibility, and Transparency (ART) requirements for wider acceptability. A method is presented for generating synthetic datasets complementing experimental observations mitigating bias present due to practical limitations. Also, an incremental hybrid machine learning algorithm is proposed. It combines ensemble learning and hybrid stacking using extreme gradient boosted decision trees and k-nearest neighbours to meet individualisation, and ART requirements while maintaining low computation footprint.

The proposed approach was based on the criteria: nonintrusive, nonwearable, motivation and engagement enhancing, individualized, supporting daily activities, cost-effective, simple, and transferable. The motivation method, suitability for elderly, and intended use were examined as supplementary criteria. Indicators of enhanced motivation and engagement, through questionnaire responses, demonstrate that >83% of participants support the proposed system's motivation and engagement enhancement. The system is fit for purpose with statistically significant ($\varrho c>0.99$, $R^2>0.94$, ICC>0.96) and unbiased correlation to the gold standard. The model reaches up to 100% accuracy for FTSTS and TUG in predicting associated patient medical condition, and 100% or 83.13%, respectively, in predicting area of difficulty in the segments of the test. Results show an improvement of 5% and 15% for FTSTS and TUG, over previous intrusive approaches.

Keywords:
Home-based rehabilitation systems, Stroke rehabilitation, Telerehabilitation, Patient participation, Motivation, Comparative effectiveness research, Automated timed up and go test, Automated five time sit to stand test, Self-evaluation, Evaluation of sensor systems, Non-intrusive sensing, Sensing for health, Accountable Artificial Intelligence, Responsible Artificial Intelligence, Transparent Artificial Intelligence, Hybrid ensemble learning, Patient-centric individualised rehabilitation

1 Introduction

In this chapter, an introduction will be presented regarding various types of stroke and its effects. A variety of patient's profiles and behaviours will be examined as well as their approach and reactions toward this particular condition. Different stages of stroke will be presented, existing methods of psychological support as well as various theories for psychological support and motivation enhancement towards goal-oriented rehabilitation for motor control impairments, which are significant contributors to patient engagement and rehabilitation progress. This chapter will inform the review in Chapter 3 and the identification of criteria for engagement and motivation enhancement.

1.1 Types of Stroke

Stroke is one of the main causes of death and permanent impairment [1]. Based on the recent research worldwide, stroke is responsible for 6.7 million deaths annually [1]. Particularly in the United Kingdom (UK) there are about 100.000 stroke cases each year. However, there has been in UK a decrease between 1990 and 2010 of 19% [1].

There are two main categories of stroke. The first one, ischaemic stroke, affects a higher percentage of the population, around 80% - 85% [1][2][3]. Ischemic stroke can happen in two ways embolic and thrombotic [2]. The second category, haemorrhagic stroke, affects a lower percentage of the population, around 15% [1][2][3]. There are other types of stroke, less common, which can occur from different causes such as cardiac arrest and hematomas adjacent to the brain. These types are not categorised directly and affect around 5% of the population [2][3]. Figure 1 presents the classification of pathology into subcategories as defined by International Classification of Functioning, Disability and Health (ICF).

Transient Ischemic Attack (TIA) is also known as mini stroke, manifest itself with stroke symptoms that last twenty-four hours before disappearing. Whereas TIAs typically do not cause permanent brain injury, they are a significant wake-up call of stroke and should not be ignored. Forty percent (40%) of the people who have had this experience will experience a real stroke. The period from TIA to stroke development varies from 2 days to 3 months.

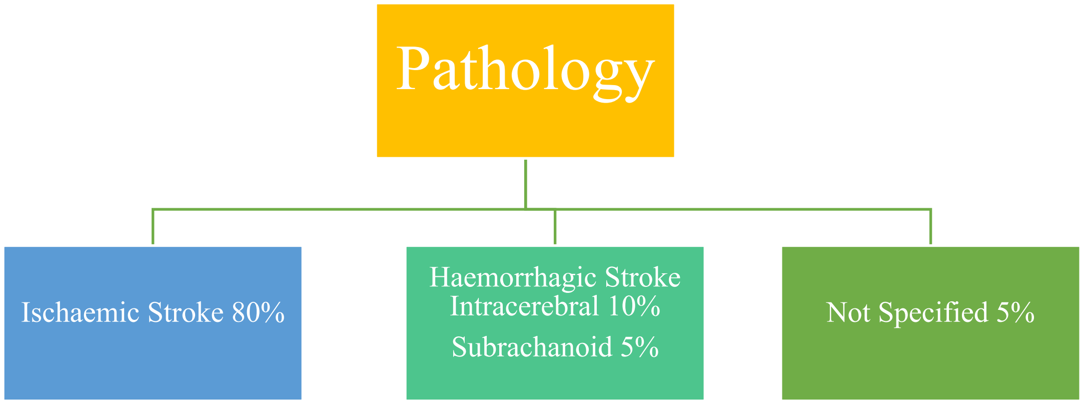


Figure 1 Pathology of stroke subcategories defined by ICF; sourced from [3].

1.2 Stroke Effects

The nature of stroke recovery is heterogeneous. The position and size of the lesion and the extent of recovery influence the residual effects of stroke. Recovery could be a complicated process that most likely happens through a mix of unprompted and learning depended process. This includes the restoration of broken neural tissue, rearrangement in order to regain lost abilities, and improvement of inequalities between the impaired skills of a patient and the demands of their environment [3].

In order to understand the level of functionality after stroke and the human body's physical restrictions as well as disability effects after various medical conditions, a multipurpose classification scale [4] which includes the condition of stroke has been established. The reasoning behind the use of the ICF classification presented in Figure 1 is, for example, the qualitative collection of data and information for various illnesses in which the medical diagnosis by itself is not adequate. Some examples include: the level of physical limitation and the limitation of space use after the illness, the illness duration and rehabilitation, the time limit in which a patient can return to work, the level of social interaction, integration to the society, as well as the level of performance at work [5]. Table 1 and Table 2 present impairments and restrictions respectively which are interlinked given that the level and the number of impairments of stroke survivors varies based on which part of the brain has been affected. Stroke generates analogous restrictions which reflect on several limitations on daily activities.

Table 1 Disorders which can be caused by stroke and body functions affected post-stroke; sourced from [3][6].

Most Relevant Body Functions Affected
Consciousness Orientation And Intellectual
Temperament And Personality
Energy And Drive
Sleep, Attention And Memory
Psychomotor And Perceptual
Cognitive And Seeing
Proprioception And Touch
Voice And Articulation
Ingestion, Defecation, Urinary And Sexual
Mobility And Stability Of Joints
Muscle Power Tone And Reflexes
Muscle Endurance
Control Of (In) Voluntary Movement
Gait Pattern Functions

Table 2 Participation and activity restrictions caused by stroke sourced from [3][6].

Most Relevant Restrictions In Participation
Acquisition Of Good And Services
Doing Housework
Preparation Of Meals
Basic Interpersonal
Recreation And Leisure Activities
Remunerative Employment
Most relevant activities affected
Communicating With And Speaking
Reading Writing And Calculating
Solving Problems
Undertake Single And Multiple Tasks
Transferring Oneself
Maintaining Body Position
Walking
Mobility
Toileting
Dressing
Moving Around Driving And Transportation
Washing And Self-Care
Hand And Arm Use

1.3 Differentiation of Gender, Age and Ethnicity

Stroke affects all people regardless of age and ethnicity, although it appears to be affecting a higher percentage of elder people. There is a slight age difference between gender and race as well. Women are most likely of being affected in elder age than men [7]. Black and South-Asian people have higher percentage of being affected in younger age than white people although different life styles must be taken under considerations such as smoking, alcohol consumption and drug abuse [8][9][10]. Recent research has shown that the age of an individual who can experience stroke has been decreased in some cases from 74 years old to 55 years old. However average age is for men 74 years and for women 80 years [7]. Moreover, there are cases of stroke in children. Children develop mostly haemorrhagic stroke instead of ischaemic [11][12][13][14][15][16].

In any case, stroke can create permanent disabilities. The disabilities caused to the human body are directly relevant to the part of the brain and the level at which it has been disposed [17]. Around 60% of survivors face visual impairments in the post stroke period which in most cases will be reduced further to 20% after a period of three months. On some cases there could be communication problems, perception problems, speech problems, as well as some confusion – a condition which is well known as aphasia. Lacking strength of legs and arms are quite common which can lead to mobility and balance problems. All these could affect the patient's daily life [18][19][20]. Although in the UK there are 100,000 strokes every year, the survivors have increased by 50% in a period of 20 years (1990-2010). Thus in the UK at the moment there are more than 1.2 million survivors [21][22][23].

1.4 Phases of Stroke and Patient Differentiation

After stroke onset, two different phases can be distinguished; the hyper acute treated as hospitalised emergency (3 to 24 hours) and the acute focusing on stabilization. Each patient is unique and their needs as well as the time needed in order to get stabilized might differ [24]. After stabilization, normally there would be a discharge plan, for the patient to continue his/her rehabilitation at home or in special care homes based on their needs [7]. Discharge planning is important for patient progress. It is an interdisciplinary way to deal with progression of care and a procedure that incorporates identification, appraisal, goal setting, arranging, execution, coordination, and continuing assessment [7]. When the patients return home, normally they will be facing some difficulties in order to accomplish their daily activities. Problems such as: toileting, dressing, washing by themselves, brushing their teeth's, shaving, preparing meals, communicating with other family members, even the replacement of toilet tissue roll in the bathroom could be challenging (Table 2).

Different abilities might have been lost (Table 1). Moreover, the personality plays a vital role. For example, there were patients with balance problems that were feeling weakness to move one leg and were finding helpful the use of a stick. While others with the same problem did not want to use a stick, because they thought that they might look older or disabled [25].

Another case could be the wheelchair. For some survivors, using a wheelchair works well because they enjoy the freedom of moving around. While some others with the same impairment disliked it due to the loss of autonomy, shame, and an inclination that individuals overlooked them [26][27].

Due to different psychological factors and the differentiation of personalities one device regardless effectiveness it could work for one patient while it could be useless to another [28]. Obviously, some of the problems could be overcome with the house modification for example by installing a chair lift or with the installation of bathtubs (Table 1).

1.5 Rehabilitation

Although the diagnostics for stroke have been improved significantly (immediate brain scan and consultancy) especially in hyper acute centralized services, due to different reasons for example either clinical or practical only 10-15% of the patients take advantage of thrombolysis [29][30]. Thus, rehabilitation is an important factor in order to help patients. Its main objective is to re-establish as much freedom as could reasonably be expected by enhancing physical, mental, and emotional functions [29].

In order to achieve the best outcome in patient recovery, multidisciplinary teams in hospitals and stroke units as well as Early Support Discharge teams (ESD) use a cyclic process (Figure 2) which includes different stages for patient's recovery [3].

An organised review is presented in [3] regarding the intervention of rehabilitation that actually can be effective in patient progress. Although there are different approaches for different problems, there are 3 main areas that would be beneficial for the quick recovery [3].

1) Rehabilitation must initiate early after the incident [31].
2) According to [29][32][33][38] it is showed that patients demonstrates better progress when their training activities are repetitive. This is especially true when these activities have specific goals and or are tailored to meet the needs of the patients and take place in a familiar environment such their home.
3) The gradually increased level of exercise is a great advantage [35].

Figure 2 Rehabilitation process, created based on information available in [3].

Although there is significant recovery progress with the multidisciplinary teams and ESD, the motivation of the patient [36] including family support [29] plays a vital role on their improvement.

1.6 Motivation Analysis

A significant factor which has a major impact in rehabilitation is the motivation of the patient. The improvement of the patient after the stroke demands personal commitment and high level of personal effort and contribution. This is a complex and fragile procedure which can be easily affected by the recovery progress of the individual as well as by their psychological condition, depression for example [36].

The motivation of the patients varies according to their needs and their achievements during the recovery procedure. Thus, for the individual who has achieved entirely or partially their targets through a set of physiotherapy exercises for example, these targets are no longer a stimulus for recovery [36].

There are different approaches in order to increase the level of patient's engagement:

1) Goal setting theory (Figure 3): this particular method sets small, realistic, manageable and well-defined targets. Challenge is a characteristic that these goals must include in order to keep motivation and engagement at a high level. This approach demonstrates high success rates in stroke survivors [36][37][38][39].

Figure 3 Goal setting theory created based on information provided in [36].

2) Building the self-confidence: through successful implementation of daily care activities was introduced by Bandura in 1997 [40]. Higher level of self-efficacy empowers people in stressful situations, increases confidence in implementing and accomplishing a task, and introduces comfort in overcoming challenges. Techniques of this approach have been used by nurses in hospital in order to increase patient's motivation for rehabilitation. It is a fact that the better the task is performed the higher the motivation is for further commitment and effort. This procedure involves 4 main areas of implementation in order to achieve the desired result [41][42] (Figure 4):

a) Mastery experience: small tasks accomplished successfully increases confidence and shelf-efficacy to address complicated tasks.
b) Vicarious experience: Observation of other patient's improvement.
c) Verbal persuasion: receiving verbal appraisal increases self-confidence and courage to perform the task.
d) Physiological feedback: support and empathy increases ability of learning new behaviours and motivation for recovery [41][42].

3) Possible selves' theory: Patients can increase their motivation for rehabilitation by providing an extend of a future self. This is directly relevant with the psychological stability in order to use positive future extension of their selves in order to imagine the development and success of recovery. This procedure creates an optimistic environment, and it can contribute to faster recovery. However, it can raise a high rate of uncertainty especially when the individual imagines mostly negative pictures [43].

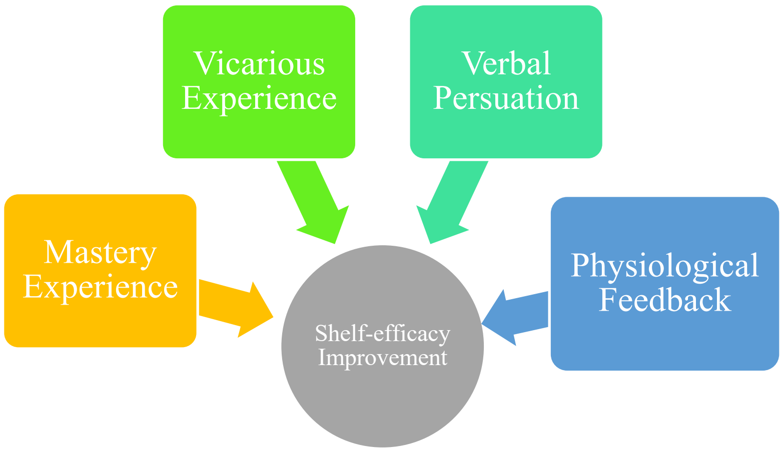


Figure 4 Factors which contribute to self-efficacy enhancement created based on [41], [42].

The motivation increases when there is no lack of information (Figures 5, 6). Survivors could be more motivated if they can retain control of their actions by increasing the level of their choices regarding their daily activities. The achievement of realistic and slightly optimistic goals could be another source of motivation because it increases the level of patient's engagement. The patient can operate and put higher level of effort when their needs and reasons that create an uncertain psychological condition such as anxiety can be sufficiently overcome [43]. However, it must be noted that motivation is quite complex and there is a significant variation between patient characters and personalities. For some of them a motivation for recovery could be something that has not been described above for example a pet who is depended upon them [36].

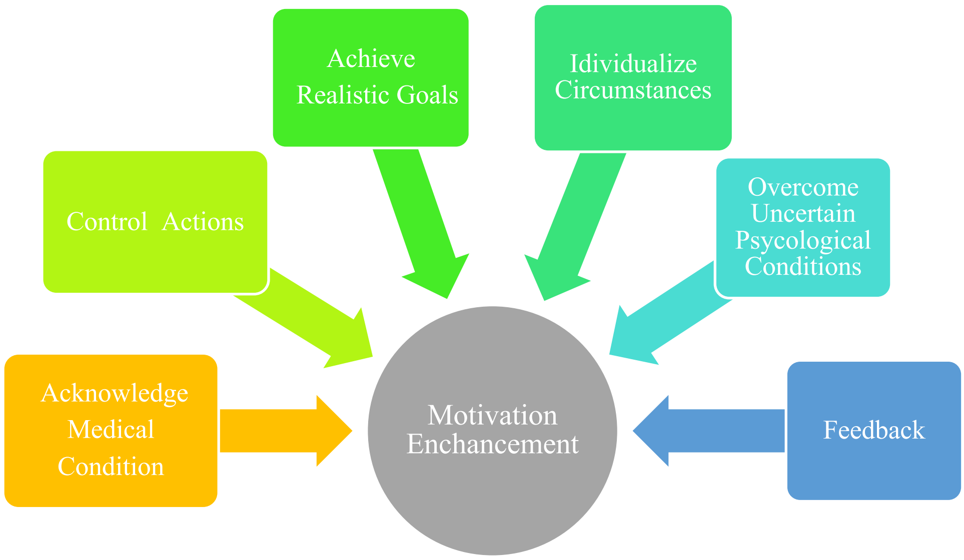


Figure 5 Factors for motivation enhancement created based on information available [36], [43].

Figure 6 The factors which can cause loss or decrease of motivation based on information available in [44][45].

Patients can demonstrate intrinsic or extrinsic motivation. Intrinsic motivation is sourced from within the patient. Extrinsic motivation is provided to the patient from the medic or carer. There are two categories of extrinsic motivation [43].

Feedback contributes significantly in motivation enhancement. There are different forms of feedback and they are mostly based on the patient's needs. Due to the variety of impairments that stroke can cause the approaches and feedback which leads to motivation enhancement might vary. There are three main categories among others. Oral or visual feedback, for example, a smile or a face expression in order to express that sufficient improvement has been achieved. This type is called motivational feedback. The second type is called knowledge of performance [36][46]. The third is physiological feedback through empathy and understanding of the patient's condition [41]

Motivation can be easily affected by different causes [44][45]. These include:

1) Different approaches for different individuals. Highly motivated patients may not respond to specific rehabilitation programmes [44][45];
2) Age or severity can negatively affect motivation. Elder people are less motivated than younger. Young survivors engage on a higher level with rehabilitation because they aim for a faster recovery [44][45];
3) Family unbalanced behaviour can negatively affect motivation e.g. setting unrealistic targets or unnecessary protection [44][45].
4) Cultural and religious backgrounds can affect motivation and outcome [44][45].
5) Environment especially progress of others [44][45] and home-based rehabilitation improve engagement [47].
6) Labelling and poor behaviour of staff can affect engagement and motivation [44][45].

1.7 Chapter Summary

In this chapter stroke was introduced and how the illness can cause significant changes was presented in different aspects of daily life by introducing several limitations. An introduction on how the reduced ability of carrying out daily activities could be linked to lack of motivation and affects rehabilitation goals and progress was presented. The importance of motivation for rehabilitation was stressed and various factors which are linked with the patient's progress and contribute on a patient's psychological condition which is vital for further recovery and strong engagement with predefined goals were analysed.

In the next chapter potential links with respect to home rehabilitation and the socioeconomical background will be investigated. The chapter will present the novelty of the thesis through the identified research questions.

2 Motivation and Research Questions

In the previous chapter, Chapter 1,an introduction regarding various types of stroke and its disastrous effects was presented. A variety of patient's profiles and focused on theories for motivation enhancement, goal-oriented rehabilitation for motor control impairments and engagement with the rehabilitation progress was examined.

In Chapter 2 the links between the home rehabilitation and the socioeconomical background motivating it will be investigated. The contribution of this particular research in terms of novelty as well as research questions will be presented. Aims & Objectives will be defined and they will be accomplished in the remaining chapters. Publications that have stemmed from this thesis and their relations to the chapters has been included along with the organisation of this thesis.

2.1 Motivation

Some of the main issues with rehabilitation have been analysed in this thesis in Chapter 1. The conclusion is that although a lot of technologies have been developed, they do not simultaneously meet criteria such as individualised rehabilitation, high level of engagement, lack of stimulus and motivation as analysed in the review in Chapter 3. Thus, most of the developed systems would be inappropriate for an extended period of use. This could result in poor rehabilitation outcomes. Moreover, most of the current systems combine high level of technological complexity making them difficult to use and understand and utilise intrusive means as well as a number of wearable sensors which are not welcomed by all patients at all time. The complexity of some of the systems make it difficult for the user to understand how to use the device and how to interpret the feedback. This is critical while at the same time the specialists rely entirely on the technology for the patient's progress which could prove to be inadequate for the particular subject and hence could lead to permanent impairment.

Some of the systems analysed in this thesis are strictly for clinical use (Chapter 3). However, findings have shown that patients respond better, and engage more with rehabilitation in the home environment, if they can be supported and surrounded by family members (Chapter 1). The successful completion of daily activities such as stand up, walk around in the house by their selves, make a coffee, increases their confidence and this could increase the level of motivation as well as engagement, given that they will be able to experience direct results.

Hence, a low cost, non-intrusive home rehabilitation system which can monitor patients progress, combine more than one daily activity and set up some goals and tasks which will be tailored to the subject needs by providing a clear and sufficient feedback and increasing the difficulty gradually based on their progress could be a successful approach to rehabilitation (Section 3.8).

2.2 Socioeconomical Background & Benefit

Stroke has become a global problem [1]. The number of stroke patients is predicted to increase by 59% over the next 20 years [48]. In the UK alone, more than 100,000 stroke cases are reported annually [1], with impairment or disability affecting two-thirds of the 1.2 million stroke survivors [1]. Due to the high number of patients, in England, for example, the social care costs are almost £1.7 billion per annum. Thus, cost is one of the main drives for service delivery practices. Early discharge units have been used, consisting of specialized personnel who offer an intensive rehabilitation program. Afterwards, the patient continues the rehabilitation at home. This is expected to reduce costs by £1600 over 5 years for every patient, according to a 2017 report [1].

Due to increasing pressure to discharge patients early [49], the need for home rehabilitation systems that are not dependent on specialist or clinician operators has increased [1], [50], [51] while providing service similar to a clinical environment. Technological advances in home rehabilitation have been mainly focused on motor control impairments due to their prevalence (85% worldwide [1]).

Rehabilitation in a home environment can prove more efficient than that in a clinical environment (Chapter 1), as it supports patient self-efficacy [52], [53] particularly through cooperation or in competition with family [54]. However, home environments have limitations that can affect the use of clinical devices. The most prevalent limitations are related to space and the lack of qualified personnel to operate devices. For example: number of occupants; patient's mobility, individual personality, and mood disorders; sound insulation, home modification requirements, and cost [3], [6]. Finally, different age groups react differently to technology and devices; for example, elderly survivors often do not engage with wearable devices or video games [55]. As a result, stroke rehabilitation requires a person-centric approach that is suitable for the home environment and that does not require infrastructure change in the home.

2.3 Contribution & Research Question

In this thesis criteria for successful rehabilitation are identified and a low-cost rehabilitation system proposed which will be simple to the user, it will involve low cost of development and manufacturing and hence will make it approachable to all subjects. It utilises modified non-intrusive low-cost sensors, covering a variety of different actions and provides sufficient feedback to the user which is easy to interpret. The system will be able to support patients with their rehabilitation goals as well as encourage patients to engage more with their goals through a simple and well understood feedback mechanism. The proposed system has a high level of transferability and is not focused on just one condition. Furthermore, could contribute on their initial evaluation of their condition by classifying them on particular clusters.

By taking this under consideration, people with different conditions, for example, different types of stroke, cognitive impairments, and dementia, will be able to use the same system. Some of the main settings, like identification of the condition, goal setting for rehabilitation, as well as the targets to be achieved, will be set automatically, given that the proposed system is able to classify the subjects based on their condition and learn from their performance over time. This provides targeted and tailored rehabilitation, which addresses major concerns regarding patients' individualisation.

When referring to either home monitoring or home rehabilitation of subjects, various sources of noise have to be taken under consideration that can be introduced during the monitoring or testing procedure. Such noise could be sourced either from other residents of the house or from different objects. Moreover, for the evaluation of this system, certified NHS tests have been used. These tests help to evaluate the proposed system in terms of accuracy, complexity and transferability to a variety of daily activities and other prognostic and diagnostic applications.

The proposed tests combine more than two daily activities which can increase self-efficacy of the subjects which in return will lead to a higher level of confidence and level of engagement. The particular tests could be carried out within the home environment. The validation of the device will be performed by comparing to a "ground-truth" dataset which will be collected manually in the fashion that it is recorded in NHS. Ethics approval will be obtained for testing and validation of the proposed system to take place.

Additionally, the feedback method will follow the goal-oriented feedback guidelines as identified in literature and will motivate the monitored subject to achieve a higher completion outcome. "Disaggregating" the moving patterns into various components is challenging. The data collected will be used to identify the stages and to describe the subject's ability in completing each stage.

The analysis performed on the timing for each building component of the moving pattern will use a clustering algorithm which will cluster the subjects based on the result and also analyse one single subject over time. According to the results of the analysis, individualised feedback will be displayed to the subject at each stage. The feedback will aim to stimulate better performance in the next repetition of the assessment test. Finally, to identify the efficacy of the device in encouraging and motivating patients, each subject will be requested to complete a questionnaire at the end of each session. The data from the questionnaires will be analysed to validate the performance of the device in relation to individualised home rehabilitation.

Moreover, the time needed for patients in order to complete the test will be significantly longer in comparison with a healthy subject. This will differentiate automatically between the patient and healthy family members causing interference in the home environment, given that the system is not designed to capture fast activities and the tests will be automatically cancelled.

Specifically, this thesis addresses the following research questions:

RQ 1: Can stroke patients and elder people be motivated and encouraged in order to engage with different daily activities for life quality improvement as well as for home rehabilitation purposes using non-intrusive ambient intelligence through low-cost smart devices?

RQ 2: Given a constraint of using a pair of communicating sensors only (and low-cost constraint and easy of installation) what kind of NHS tests can be supported at a sufficient level of accuracy which are relevant to home rehabilitation of stroke patients?

RQ 3: Given a number of underling conditions, what can be achieved in terms monitoring rehabilitation activities and monitoring of patient improvement over time (with a fixed accuracy desirable)?

RQ 4: Can Accountable, Reliable and Transparent (ART) Machine learning be utilised in order to classify patient's condition (diagnosis) and progress, giving interpreted feedback?

RQ1 and RQ2 will be addressed in Chapter 4, while RQ3 and RQ4 will be addressed in Chapter 5.

The novelty of this thesis lies in utilising and deploying a low-cost ambient intelligence sensor system and providing explainable and interpretable feedback through an applied machine learning hybrid approach, based on body-metrics and patient parameters, to provide home-based rehabilitation support tailored to the subject's needs.

2.4 Aims and Objectives

The aim of the thesis is to address the identified research questions presented in the previous subsection. This will be achieved through the following objectives:

1) Examining the state-of-the-art in-home rehabilitation systems and assessing their suitability and functionality from a patient engagement perspective in (1) combining research from 3 research domains: motivation enhancement as part of patient psychology, home rehabilitation technologies, and monitoring technologies and (2) identify a list of comparative criteria and successful device requirements.
2) Developing a low-complexity and easy to use system (hardware and software) for in-home rehabilitation that meets the set criteria
3) Designing appropriate experiments with subjects in order to evaluate the robustness and the accuracy of the system.

4) Develop, train and test a hybrid machine learning approach utilising the data recorded through the experiments and additional datasets available to support clustering subjects according to underling conditions and disaggregated motion patterns.

2.5 Thesis Organisation

The remainder of this thesis is organised as follows. Chapter 3 presents the literature search methodology. Related work and further analysis of existing technologies is covered and a taxonomy of the systems reviewed is presented along with a summative assessment, addressing all the strengths and weaknesses of the reviewed systems and demonstrating successful criteria for rehabilitation systems.

In Chapter 3 a systematic narrative literature review for systems which are used for home rehabilitation will be conducted. A taxonomy of systems for home and clinical rehabilitation will be presented as well as a set of criteria of various systems in order to evaluate the successfulness in rehabilitation accomplishments.

In Chapter 4, which medical tests are relevant for home-based rehabilitation monitoring will be evaluated. A literature review for a variety of systems which have been designed for monitoring the Timed Up and Go tests as well as Five Time Sit to Stand Tests will be conducted with respect to the criteria identified in the previous chapter. Design of the proposed system to meet the criteria, the evaluation methodology, and the descriptions of the experiments will be presented as well as analysis of the results and findings with respect to accuracy, engagement and motivation enhancement.

To address the need for patient-centric individualised solutions, in Chapter 5, a literature review regarding machine learning algorithms will be conducted. Different experiments will be carried out in order to identify the most suitable algorithms for high accuracy predictions. A hybrid algorithm will be presented, and results of the experiments will be analysed and presented later in this chapter.

In Chapter 6 a review of the findings of all the chapters will be presented, the research questions of the thesis will be revisited and the thesis novelty will be outlined. Finally, in Chapter 7 conclusions as well as future work of the thesis will be presented.

In Figure 7 the thesis organisation is presented. The circles' diameters correspond to the chapters' significance in addressing the aims and objectives of the thesis and presenting contributions. For example, Chapter 5 and Chapter 6 which answer particular research questions appear of greater significance in comparison to Chapter 1. Chapter 3 where the criteria for a successful rehabilitation system are presented is more important than Chapter 1 but less important of Chapter 5 and 6.

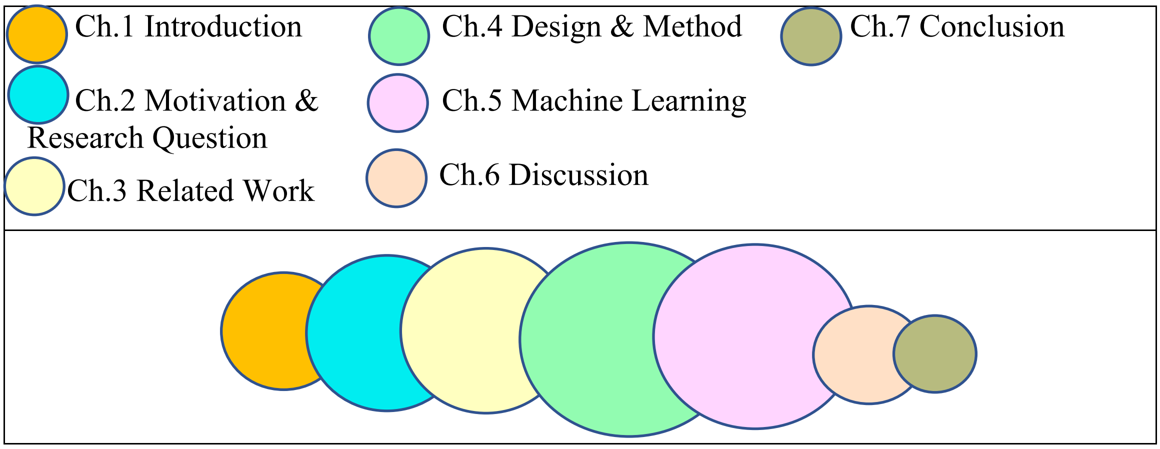


Figure 7 Thesis organisation and significance denoted by the circle diameter with respect to addressing the research questions, aims and objectives of the thesis. .

3 Review, Taxonomy and Criteria for Home-Based Rehabilitation Systems.

In Chapter 2 the links between the home rehabilitation and the socioeconomical background motivating the work was investigated. In this chapter a review methodology that was followed for the conducted literature review will be established. The resources and results of the review methodology will then be presented. Technologies relevant to stroke rehabilitation for clinical as well as for in-home use will be critically reviewed. The importance of monitoring through smart meters and ambient intelligence as well as pattern identification will be presented. Finally, the findings will be summarised and criteria for a successful device presented; thus, identifying requirements to address RQ1 (Chapter 2). The chapter is based on the material published in [56].

3.1 Systematic Literature Search Methodology

Although several review (narrative and systematic) articles have been published on rehabilitation technologies focused on particular areas of the taxonomy (e.g., wearable sensor systems review [57] and robotic systems review [54]), to our knowledge, no extensive narrative review of existing home-based rehabilitation technologies to identify criteria for designing future solutions has been done. Thus this thesis focusses on: (1) extending the state-of-the-art in assessment of home-based rehabilitation by combining research from 3 research domains: motivation enhancement as part of patient psychology, home rehabilitation technologies, and monitoring technologies through an interdisciplinary approach; (2) providing an in-depth narrative review of home rehabilitation systems that addresses both information and communication technologies and mechanical engineering solutions; (3) developing a patient motivation and engagement analysis of the reviewed technologies; and (4) identifying a list of comparative criteria and successful device requirements to address patient motivation and engagement designed based on research findings from all 3 research domains.

A list of articles and references for review of home rehabilitation systems and monitoring systems to be included in the comparative analysis were selected. The data sources used to search for items to be included in this review were the following databases of academic references, journals with a particular focus on stroke rehabilitation, and web sources: (1) PubMed, (2) Elsevier, (3) IEEE, (4) Springer, (5) Hindawi.com, (6) Journal of Neuro-Engineering and Rehabilitation, (7) websites of stroke-related institutions and foundations presenting articles on rehabilitation found through a generic Google search, and (8) Google Scholar (including ResearchGate). The database search was conducted between November 2017 and February 2019.

The search criteria included the following keywords and combinations thereof: stroke; devices for stroke rehabilitation; home rehabilitation; rehabilitation engagement; rehabilitation motivation; stroke rehabilitation; tele-rehabilitation; smart meter; pattern recognition; kinematic analysis; robotic systems; exoskeleton systems; virtual reality; games; mobile applications; individualization; gait analysis; upper limb rehabilitation; balance rehabilitation and/or training.

As the above combination of data sources and keywords returned a vast amount of results, we selected the following inclusion criteria to identify the most relevant sources. (1) Language: English. (2) Date range: within the past 20 years (1996-2018). The majority of articles were published within the past 5 years to reflect the state-of-the-art (since 2014). Older references were made to technologies that substantially shaped the future direction of home rehabilitation systems. (3) Relevance: home or self-rehabilitation was necessary.

3.2 Systematic Literature Search Results

The literature search returned a total of 307,550 results after the inclusion criteria were applied as presented in Table 3. The following exclusion criteria to identify the most relevant sources and reduce the number of literature search results, were used: (1) no relevance to stroke rehabilitation in the home environment, (2) trained personnel required to operate the technology; (3) medication or other clinical intervention required, (4) no report of engagement or motivation as a result of using the technology or other form of patient feedback, (5) no description of the technology, (6) no report of usability especially for older people, and (7) no additional contribution to the review findings compared with the previously reviewed articles.

Table 3 Results of the literature search before and after inclusion criteria were applied.

Topic	*Results of topic search*	*Results after inclusion criteria*
Devices for stroke rehabilitation	325,000	6800
Home rehabilitation	1,150,000	36,200
Rehabilitation engagement	651,000	17,100
Rehabilitation motivation	128,000	17,300
Stroke rehabilitation	1,640,000	45,800
Stroke; telerehabilitation	8180	3110
Smart meter; pattern recognition	83,200	18,100
Stroke; kinematic analysis	105,000	15,700
Stroke rehabilitation; robotic systems	43,700	16,900
Stroke rehabilitation; exoskeleton systems	15,300	4440
Stroke rehabilitation; virtual reality	41,000	14,100
Stroke rehabilitation; games	47,100	16,900
Stroke rehabilitation; mobile applications	46,500	17,400
Stroke rehabilitation; individualized systems	35,800	17,300
Stroke rehabilitation; gait analysis	112,000	16,000
Stroke; upper limb rehabilitation	138,000	17,200
Stroke; balance rehabilitation	398,000	15,600
Stroke; balance training	799,000	11,600
Total literature search results	5,766,780	307,550

Overall, 420 sources we studied. The remaining sources were excluded after reading the abstracts. A total of 96 sources remained for analysis (Figure 8) after meeting the inclusion criteria and having not been eliminated through the exclusion process.

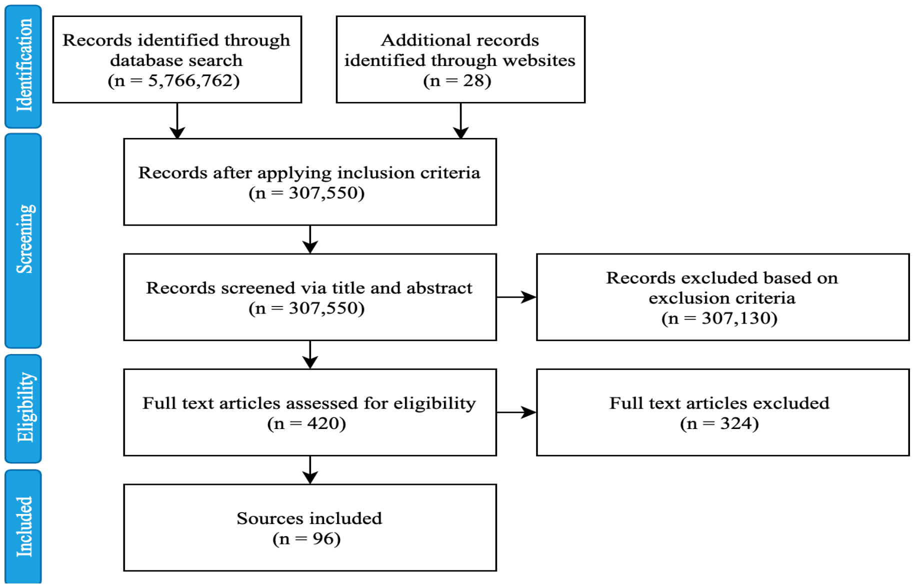


Figure 8 PRISMA diagram of the review stages.

3.3 Overview of Engineering in Rehabilitation

After a disability occurs [28][58], technological solutions might not be suitable in all cases. For example, for a phase of stroke called aphasia – described in earlier sections – the engineers have been unable to find an effective way of communication [28].

There is a difference between rehabilitation and assistive technology. The term rehabilitation is used in order to describe the utilization of scientific knowledge and sciences in order to provide improvement to people with disabilities. On the other hand, assistive technologies are the tools used to achieve the rehabilitation goals. These include devices, strategies, services that help a patient in a functional activity [28].

The complexity of the technology and the cost of manufacturing most times are directly relevant (Figure 9). Thus, complex devices with higher level of scientific knowledge will be more expensive than devices which combine low level of technology and they are simpler for development and use.

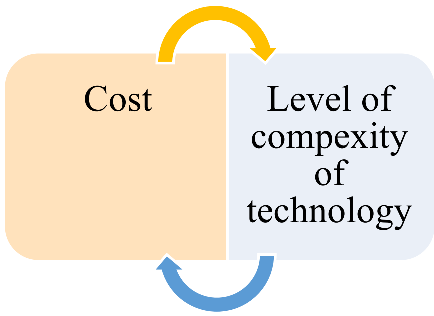


Figure 9 Relationship between cost and complexity sourced from [28].

Assistive technology devices demand complex design due to their multidisciplinary nature. Thus, often a wide range of engineering science (biomedical, electronic, electrical, mechanical etc.) will be combined with the knowledge of specialists from different disciplines such as medical or physiotherapists.

In every case the main scope is to assist and improve an individual's daily life. Thus, one of the criteria that we would need to take under consideration for a well-accepted design is the acknowledgement of the different stages that the person who has developed a particular disability has gone through. These stages can be related to either the individual's emotions or alterations of the mind state [47], the patient's needs and the differentiation of the personalities, the restrictions, the environment, the selections and the capabilities. The good understanding of the subject states and the differentiation of the conditions could be determined if the assistive or rehabilitation device used is well accepted by the individual [28].

Due to the rapid development of the technology lately, there are devices that use different technologies and approaches in order to improve disabilities after stroke. Although the design criteria for people with disability should remain the same, the philosophy and the technical approach for the patients in order to improve and eliminate the impairment after stroke is more complicated. This has to do mostly with the large amount of disabilities that scientists and engineers would need to address (see Table 1 on page 3).

Although the impairments of stroke survivors can vary, it has been noticed that a larger number of survivors suffer from problems of motor control [59][60]. Thus, a large number of rehabilitation approaches and devices based on the literature has focused on the improvement of motor control impairments but the proposed approaches differ [61].

This review focuses on: Loco Motor Training [60][62][63][64][65][66], exoskeleton approaches [67][68][69][70]; kinematic analysis [71]; robotic systems [72][73][74][75][76][77][78][79]; Biofeedback electromyography; wearable systems; and software approaches.

Robotics and exoskeleton devices are inappropriate for home use [72][73]. Although they are quite promising for rehabilitation of the individuals [74][75][76][77][78][79], they are cost-effective in a way due to the reduction of specialised personnel in the clinic. However, they are still mostly targeted to applications in the clinical environment [80][81]. On the other hand, the cameras for kinematic analysis etc. have been considered as an intrusive mean for most of the patients and individuals [55].

Biofeedback Electromyography [82][83] offers a variety of devices that can help patients to recover in their own environment [84]. However, because they demand the usage of wearable electronic equipment and removable sensors, it might not be suitable for all patients and particularly elderly [55].

Moreover, Virtual Reality and Video games are quite promising methods for rehabilitation, like Nintendo WII or XBOX Kinect [85][86] [87][88][89][90][91][92] [93][94][95][96][49]. However, there are challenges to overcome such as the individualization of the game to the patient needs in order to keep them highly motivated [50][36][97][98] as well as the engagement level with new technology for elder people which is low [55].

Non-invasive brain stimulation techniques such as Transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) [99], and regeneration of neural tissue with Stem Cell Therapy [100] will not be expanded in this thesis because they require intervention and clinical support. Thus, they are out of scope.

Mirror Therapy could contribute in post stroke motor recovery and with further instructions through video and audio-visual means could be implemented in home. However, most times the therapy must be done under the supervision of specialists and with combination of conventional treatments for better results [101][102].

Different academic and commercial devices have been published which aim to help with home rehabilitation such as, robotic systems for home use [103][104][105] [106][107][108] [109][110], balance measurement [111], wearable body sensor network systems (WBSN's) [112], cell phone balance training [113], rich to grasp training with extensive feedback [51], low cost systems with resistive elements [114], music glove which motivates patients with the help of music [115], rehabilitation systems with continuous monitoring or exercise [116]. These will be examined in detail in Section 3.5.

In the case of Cognitive Impairments after stroke, several methods have been used for rehabilitation. There are not sufficient patient outcomes regarding the means that can improve cognitive rehabilitation [60]. The problem of memory is quite challenging for stroke patients although there is improvement to the patients who receive escitalopram. The methods that have been used in [117][118][119] will not be expanded in this thesis because they are based on medical intervention and are out of the scope of this thesis as they are not addressed through engineering solutions.

Similarly, rehabilitation has also focused on improvement of Aphasia. Aphasia is a condition which affects one third of stroke patients. There is significant improvement with SLT (Speech and Language Therapy) [60][120]. However, further techniques which have been used are out of the scope of this thesis thus will not be expanded further.

3.4 Limitations of Rehabilitation Technologies.

Most of the methods for rehabilitation present some significant disadvantages. The majority are mostly deployable in the clinical environment. Wearable sensors or kinematic analysis with camera implementations are consider intrusive or not suitable for all patients. Similarly, video games cannot be used by all patients at the same level and do not provide high engagement from all the patients. Thus, individualisation for some of the video games has been proposed in order to keep the patient engaged and maintain their motivation [72][50].

Furthermore, it is well understood from the engineering society that:

- One assistive technology cannot provide a solution for all the impairments.
- The cost of a successful device sometimes could be low given that devices with lower complexity of technology could prove more suitable and easier to use in comparison with others which are more expensive and complex.
- People with impairments could be facing changes to some daily activities during time, thus, the assistive means can prove less accurate and incapable to assist fully the patient.
- Devices which do not provide high level of engagement or motivation enhancement are more likely to be abandoned after 90 days [28].

The successful design of assistive technology or rehabilitation device should take under consideration what the individual will be trying to achieve during rehabilitation; for example, all the goals and tasks which the therapy involves. A multidisciplinary team which combines different experts could achieve a better outcome [28].

Quantification and further analysis of present and future condition of the patient could overcome difficulties and unforeseen circumstances and could result to a better design of the assistive technology. Data and patterns from electronic databases are quite important because they can be used in order to identify and propose the appropriate system for rehabilitation. The choice of the individual must strongly be taken under consideration because the patient will be the user of the device or the assistive technology, thus, he/she must feel comfortable with it. The technology would be operating more successfully if the patient's environment has been fully informed about its operation. It appears that the continued adjustment of the technology/device based on the particular patient is more beneficial.

3.5 Home Rehabilitation Systems

Given that the amount of the stroke patients will be increased in coming years, there is an excessive need for new systems for home rehabilitation [50][1][51]. The rehabilitation in home, however, is very challenging due to different difficulties, such as design and technical limitations [72]. The following methods have been concentrated strictly on home rehabilitation of the patients after stroke. In this section, the sources identified through the systematic literature search are discussed through a narrative review. A taxonomy, comparative analysis and critical evaluation is then presented in rest of this chapter.

In [51] a home rehabilitation system for upper limb recovery after stroke is proposed. A specially designed desk and chair were used to monitor the patient's movement through sensors and cameras (kinematic analysis). However, the system is intrusive using cameras, has a wearable component and has a high technological complexity while is not tailored to the needs of individual patients. Similarly, in [103] an exoskeleton device has been proposed for upper limb rehabilitation. However, the system does not provide sufficient feedback to the subject.

In [104] a home wrist and fingers rehabilitation system is proposed. It comprises a robotic glove and software including different games in order to keep patient highly motivated. However, the system utilises wearable sensors, is not tailored to the needs of the user, and engagement and motivation are not investigated.

In [105] the researchers have identified that home rehabilitation under the supervision of a specialist is quite beneficial. A home gait rehabilitation approach is proposed focusing on dynamic force analysis [121]. Although the device is quite promising, and it is suitable for gait rehabilitation at home, it does not provide any feedback and the level of patient's engagement is unknown. Moreover, the device is mostly for supervised rehabilitation and that would increase the overall cost.

In [106] a robotic system has been developed for upper limb rehabilitation. The proposed system is portable. Although the device is inexpensive it does not provide any feedback and there is no sufficient evidence that it keeps the patient engaged with his daily goals. Moreover, there is no adjusted level of difficulty and this could lead to a patient steady state condition of recovery without further improvement.

In [109] a robotic system has been proposed for home upper limb rehabilitation without supervision. The software contains 8 different games in order to engage the subject more efficiently. The games through an interactive algorithm can increase the level of difficulty gradually and contributes toward a sufficient recovery. However, the device does not provide sufficient feedback and it has not taken in account some difficulties such as the activity of other people in the house at the same time where the exercises are taking place. Furthermore, it did not keep high the level of patient's engagement.

In [110] a system has been proposed for upper limb rehabilitation. The system can individualise the therapy in order to increase the motivation of the patient evaluated through the method of Fugl-Meyer [122]. The system can provide extensive feedback. However, the population which has been used in order to test the system is small. Note that previous studies, reviewed in [110], have shown that game environments or usage of additional equipment on the subject quite often leads to aversion from treatment.

In [111] a system has been proposed for home use in order to help patients to regain their balance. However, it would be more helpful if clinical trials with subjects are presented. Moreover, older patients' engagement with new technology and particularly cameras and games is not always successful.

Researchers in [112] have proposed a Wearable Body Sensor Network system (WBSNs) for home rehabilitation to help people with upper limb stroke impairments through the method of Brunnstrom in 6 stages [112]. However, the paper discusses a 6 stage method, while according the literature the Brunnstrom method consists of 7 stages [123][124]. Furthermore, wearable systems are not ideal for all patients.

In [114] many conventional methods for rehabilitation such as exoskeleton devices and robotic arms have been analysed. Most were found to be either very high cost or unsuitable for home rehabilitation. A simple system for goal-oriented training that consists of special modified resistive elements is proposed. The system can provide a feedback as sinusoidal signals. It is proven that graphs such as signals are not sufficient to deduct information [125]. Although the system is low cost, the feedback is not individualised.

In [115] the device which has been proposed for hand rehabilitation has taken into account the positive influence of music. In this paper, it has been found that the device will be more functional and beneficial when it can be used for daily activities. The device is a wearable glove and provides musical and graphical feedback to the user. Although the device is low cost, it has shown that it can efficiently help only patients with moderate to low level of impairment. Moreover, it is wearable. The individualisation of the device might provide better results.

In [116] an individualised system for physiotherapy is proposed. It has been found that the continuous monitoring of the patient is important. Although the system has understood the philosophy for individualised therapy, the devices that have been used are not tailored to the patient needs, they use the same mean for all patients. The level of engagement has not been tested especially with older people and there is no further evaluation of the device.

In [88], researchers have identified the importance of motivation through challenging, and simple yet gradually increasing in difficulty tasks. A virtual reality system for upper limb rehabilitation is proposed using wearable sensors. The therapy and the goals provided to the patients are tailored. Although, the paper presents successful rates after the clinical trials, it has not taken under consideration the reaction of elder patients. Furthermore, elder patients present other neurological disorders as well in addition with stroke which might contribute to further limitation on the system usage.

In [89] a game for upper limb home rehabilitation has been proposed. The researchers have identified that the subjects can have better chance for successful rehabilitation in home when they choose the parameters of their rehabilitation. It has been found that there is a direct link between the self-control in therapy and the enhancement of motivation [126]. The game allows an initial calibration in order to assess and understand the patient needs, evaluated with Fugl-Meyer [122], Action Research Arm Tests (ARAT) [127] and ABILHAND [128]. Although the system presents good results and does not employ any wearable sensors or intrusive means like cameras, it provides therapy through games, and it is a question how the system helps the patients to accomplish daily tasks.

In [95] a home device, it has been suggested that it must be small and portable with sufficient feedback to the patient which utilises different games which help to improve upper limb functionality. The system adjusts the level of difficulty based on needs. Although through the evaluation in this study the results are promising regarding the motivation of the patients, which can be maintained on a high level, and the recovery results are encouraging, there is a significant variation on the device usage. Moreover, it employs wearable sensors.

In [49] a iPad app game is proposed. for upper limp rehabilitation. The application does not provide any significant evaluation of the subject and it has significant limitations for individuals. The paper does not explain the level of the feedback which the application can provide to the patient as well as the level of the motivation and engagement.

In [97] a system which has been proposed for upper limb rehabilitation, combines wearable sensors. The important finding is the way to increase the motivation of the individual. Along with the common games that the authors have offered which are four in total, three of them can be used against or with healthy individuals in a competitive or cooperative way. Through the clinical trial, it has been found that the motivation and enjoyment increase significantly. Yet, more trials must be carried out and for a longer period of time because in this paper the time of subject involvement was limited. Furthermore, if the device is individualised for patient needs then this might not be suitable for the healthy individual.

3.6 Taxonomy

Following from the above systematic search and comprehensive narrative review, we propose a taxonomy of rehabilitation systems, shown in Figure 10, based on the type of technology presented in the reviewed articles. We develop the taxonomy on the basis of the therapeutic effect in combination with sensing technology. Home rehabilitation mainly focuses on motor control impairments due to minimal or no clinical and medical intervention [59], [60]. On the other hand, most clinical systems (see left-hand side of Figure 10) have dependencies and are difficult to implement at home. Therapy that requires either clinical or specialist personnel to assist in execution includes transcranial magnetic stimulation and transcranial direct current stimulation [99], regeneration of neural tissue stem cell therapy [100], and mirror therapy [101], [129]. Similarly, treatment of aphasia and cognitive impairments is predominantly within a clinical environment or through specialist intervention [60], [120] . As a result, these approaches would require regular home visits or would be impossible to perform away from the clinical environment.

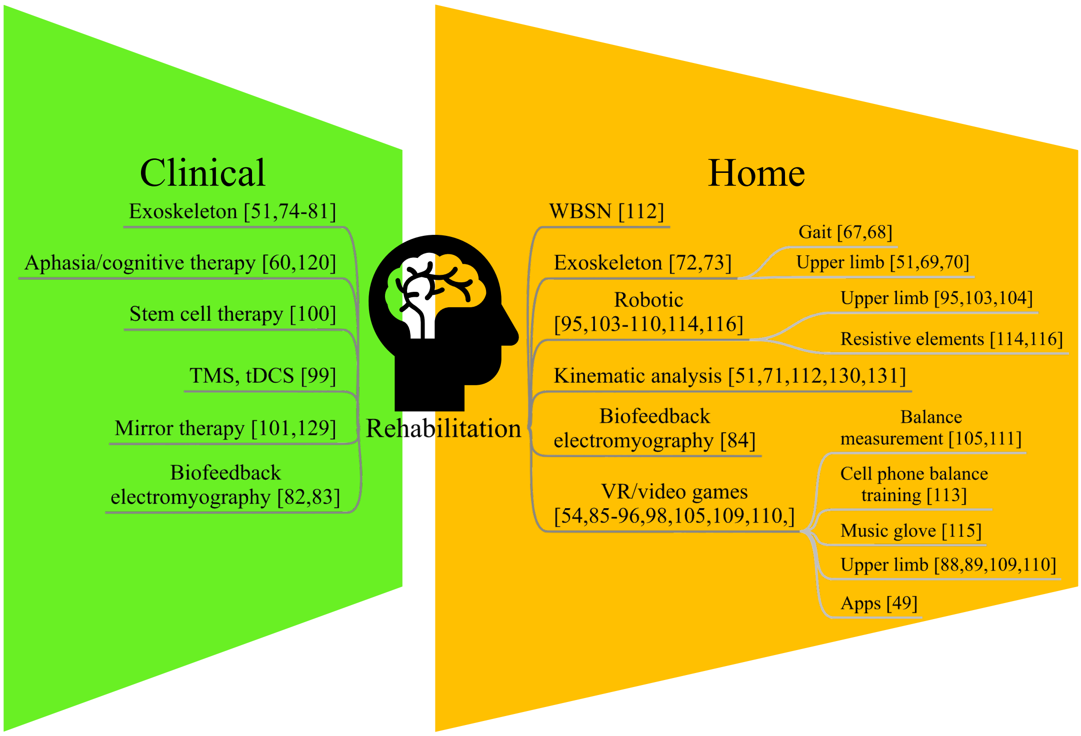


Figure 10 Taxonomy of Rehabilitation Systems for stroke patients, VR: Virtual Reality; tDCS: Transcranial Direct Current Stimulation; TMS: Transcranial Magnetic Stimulation; WBSN: Wearable Body Sensor Network Systems

The right-hand side of Figure 10 shows a variety of methods and approaches developed to support home rehabilitation focusing on locomotor training. They differ based on the individual’s situation and disabilities [61].

Locomotor training [60], [62]–[66] can be implemented through various methods. One approach is through the use of exoskeleton devices [72], [73] for gait [67][68] or upper limb [69], [70] training. Most large exoskeleton devices reduce clinic personnel costs [74]–[79] but are inappropriate for home use [80], [81]. Some devices in this category have started to have feedback mechanisms incorporated, such as that described in [51]. However, these are still very expensive systems requiring a caregiver to guide and support training. Thus, we did not review these systems.

Biofeedback electromyography is based on feedback systems [82], [83]. Though mainly designed for clinical use, some devices using this approach have been designed for home use, such as Biomove [84]. However, the disadvantage of this method [84] is the use of wearable equipment, which is not suitable for all patients and particularly for the elderly [55].

The same challenge is faced by wearable body sensor network systems [112]. Additionally, observation by expert or clinical personnel is often needed and, thus, we did not investigate these 2 categories further in this review.

Another approach is to use cameras or wearable sensors for motion or kinematic analysis [71], [112], [130], [131]. Cameras and wearables, however, are considered too intrusive for home use by many patients and individuals [55]. Many applications of cameras and wearables in home rehabilitation systems exist; thus, we reviewed these in detail.

Robotic systems have been heavily investigated [103]–[110] for home use. However, they face the same challenges of high complexity and cost. This includes systems such as low-cost resistive elements training [114]. However, these systems still do not avoid the requirement for supervision of the exercise. We reviewed systems in this category to identify their ability to enhance motivation and patient engagement.

Another area of research interest is the virtual reality and video game domain [49], [85], [87]–[96]. Although this is a promising area for home rehabilitation, there are still many challenges. The games are not individualised to the patients' needs; hence, patients lose motivation easily and are not engaged with the activities they need to perform [36], [50], [97], [98]. In particular, elderly patients demonstrate very low engagement with this technology [55]. This category can be expanded to include balance measurement [111], cell phone balance training [113], and even a music glove, which motivates patients with the help of music [115]. We further analysed systems in this category.

We critically evaluate home-based rehabilitation technologies with a focus on patient engagement as the widely recognized key indicator of success of rehabilitation systems in the reviewed articles. In contrast with usability, which is a measure preferred in human-computer interaction studies, engagement is not the singular measure of the usability of an interface, but rather of the perpetual retention of the user's interest over a prolonged period of time as defined in [132]. Engagement can be the effect of a successful human-computer interaction design in combination with the psychological motivation of stroke survivors for rehabilitation [132]. Based on the literature, engagement is more likely when the feedback is sufficient and well understood by the patient, and the system, apparatus, or device is easy and convenient to use without employing intrusive means and without complex requirements from the user [133].

3.7 Contribution of Monitoring in Home Rehabilitation

The literature review identified challenges in incorporating the rehabilitation systems in the home environment. Retrofitting is more challenging than designing smart homes with embedded technology. To avoid these issues, research has mostly focused on smart home environment or monitoring devices that stand alone and do not require redesign of the home. Such systems mostly focus on monitoring generic parameters and provide individualization through pattern recognition algorithms, but do not contribute to rehabilitation activities. To support rehabilitation, their scope would need to be altered to encompass rehabilitation goals, and patient motivation and engagement, while at the same time being transferable (supporting different application domains). This section explores the potential benefits such systems could bring to the home-rehabilitation domain.

Smart meters are devices used in order to measure energy consumption and contribute to a better and sufficient usage of appliances, which will be providing the additional benefit of lower cost reflected on the electricity bills. The smart meters will be playing an important role on the design of future Non-Intrusive Load Monitoring (NILM) [134]. Although smart meters are not directly related to stroke rehabilitation, they can play a vital role in pattern identification. For example, monitoring is crucial for some cases of stroke patients who can develop a second illness such as dementia thus further analysis in monitoring patterns might be useful. Due to the potential growth of usage in the near future an advantage could be taken through the identification and analysis of power consumption patterns. With an indirect way daily activities could be monitored and data collected can be used for health related monitoring when certain qualities and additional sensors are used as presented in Figure 11[135].

A future NILM system will be having the ability to identify the particular appliance usage and take samples of electricity in sufficient time frame. The usage of particular appliances could create individualised patterns for the home users and this can help for early detection of different illnesses such as dementia and mental disorders. Most importantly, NILM does not utilise any intrusive means like cameras or wearable sensors, it is integrated in the environment of the user and it has good transferability on various illnesses detection. Developers would need to take under consideration that for non-intrusive monitoring system designed for a particular illness, changes on behaviour and Assistive Daily Leaving (ADL), or faster sampling time, and complimentary sensors might be necessary [135][57].

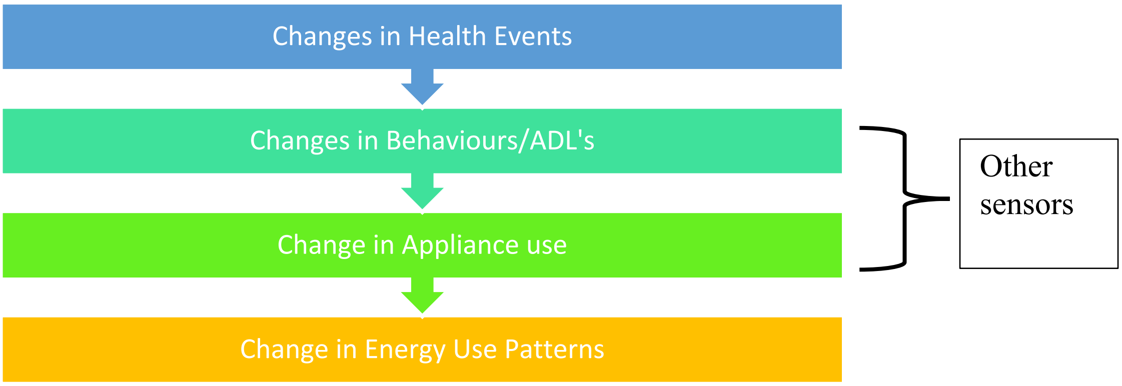


Figure 11 Monitored qualities for the health and care applications and the need for additional sensorial input to monitoring devices, source [135].

The smart meters can be used mostly for telehealth and telecare, and the patterns of the users will be able to detect possible illnesses. Possible illnesses detection can be viewed in Table 4.

It is worth noticing that research has focused on the development of algorithms using machine learning in order to understand user behaviour in a more accurate way [55]

[136][137][138][139][140][141][142][143]. Obviously, it might raise concerns to the users to share personal data. In addition, it raises the question of who is responsible in case the device does not operate in the promised way [135][143]. This is further examined in Chapter 5.

Table 4 Diagnostic capabilities for health conditions using smart meter data, source [135].

	In-activity	Sleep Distur-bances	Memory Problems/ Confusion	Changes in other behaviour or activity patterns	Occupancy/ Absence	Low activity/ Non-compliance	Unhealthy Living conditions
Dementia including Parkinson	✓	✓	✓	✓	✓		✓
Elderly	✓		✓	✓			✓
Anyone in need for home treatment						✓	✓
Alcohol abuse	✓		✓				✓
Common Mental disorders		✓					✓
Post - Traumatic Stress disorder		✓					
Attention deficit hyper-activity disorder		✓					
Mania		✓		✓	✓		

In addition, there are commercialised assistive technologies helping patients and their relatives. Some of them can be plugged in the wall socket and are able to monitor the power consumption of a particular device/appliance. These commercial devices can notify the carer of the patient that someone used the device and this can be an indication of the patient's or elder person's condition [135][144][145][146][147][148][149]. The main benefit of these systems is the low cost in comparison with other systems. However, there is no application of such systems in the rehabilitation domain, all of them are focused mostly on monitoring.

To the best of the author's knowledge, there is only one system in literature whose aim is to bridge monitoring systems and rehabilitation. In [150], the proposed system is based on the use of wearable sensors along with distributed sensors and cameras within the patient's home. The paper promises to support tele-rehabilitation but only through monitoring by a trained caregiver from a remote location. As previously identified, wearable sensors are not preferable. In practice the claim of bridging monitoring with rehabilitation is not implemented in this system. It is instead, a monitoring system that enables remote management of rehabilitation.

3.8 Critical Evaluation of Engineering Approaches

Table 6 demonstrates a summative comparison of all the systems discussed in the previous sections. Only the systems that are suitable for home use were presented, clinical systems were not taken under consideration. The first three criteria based on motivation and engagement aspects were selected (Section 1.6). The next four criteria were selected according to commonly used and evaluated metrics in the majority of the reviewed articles (Sections **Error! Reference source not found.** to 3.7) informed by the conclusions in Chapter 1. The remaining criteria were selected to meet other acceptability and economic aspects, as well as a separate category for the application area. The presented analysis identifies 3 aspects of technologies that we use for comparison: (1) motivation, (2) acceptance of technology, and (3) technological aspects. These aspects were selected for their importance in supporting patients' motivation and engagement (motivation) and in being incorporated into patients' rehabilitation routines (acceptance, technology).

For each aspect, several comparison criteria were identified. Regarding motivation, the criteria are (1) the motivation method used, (2) the patient's engagement with the technology, and (3) whether the technology supports daily activities as an additional measure of motivation. There are 3 motivation methods: cooperative, supportive, and constructive. When the method used in a technology was not specified, the characterisation made was general. With respect to acceptance, the criteria are (1) individualization of the device to meet patients' needs, (2) suitability of the device for elderly patients, (3) the use of wearable components, and (4) the use of intrusive monitoring methods (e.g., wearable sensors, on-body sensors, cameras, microphones). Wearable and intrusive methods have a negative impact on acceptance. Technological aspects are (1) intended use for the technology (monitoring, rehabilitation, diagnosis),

(2) cost, (3) complexity, and (4) transferability to other domains. Regarding intended use, besides our focus of rehabilitation, we also included 2 systems that perform monitoring and diagnostics.

Based on Table 6 there is lack of systems for home rehabilitations due to various limitations in house design. Strengths and weaknesses of the systems are identified based on several criteria. It can be seen from the table, that what has been missing, is a system which can combine high level motivation and engagement by helping the individual through rehabilitation exercises to accomplish daily tasks. The system must be not intrusive and not wearable as well as tailored to the needs of the user in order to increase the level of usage and the number of the users regardless the age. Due to the complexity of stroke illness it has to be taken under the consideration that the subject has not been affected by other illnesses which affect elder people, thus the monitoring of the patient and the transferability of the device which will be capable of diagnosis or prevention of other threat is crucial in several occasions. Sometimes due to the relationship between cost and complexity (Figure 9), the devices which are low cost appear to be less complex, but they are not efficient enough due to the lack of features such as feedback etc. Thus, the system should maintain all the above criteria and at the same time must maintain simplicity at a low-cost basis.

We use the extracted information from the reviewed articles to establish the criteria and to identify whether the criteria were met by the proposed systems. For the engagement and motivation criteria, as well as acceptance, all of the reviewed articles reported results on a common basis; thus, we needed no additional steps to cross-validate the reported results.

Table 6 presents a detailed comparative analysis of all the aforementioned technologies applicable for use in the home environment. We select the technologies as representative examples of each category we analysed. The system is assessed to be individualized or personalized or person centric when it learned or adapted to the needs of a particular patient by incorporating some type of feedback loop mechanism where the device adjusted the requested task(s) to the ability of the patient. Examples of such mechanisms are various machine learning approaches and increasing task difficulty. Suitability for the elderly was assessed based on Debes et al [55]. Nonwearable (on-body sensors, wearable components) and nonintrusive (cameras, microphones) systems were classified according to the system inputs that were used. The intended use of the system can be for rehabilitation, smart home monitoring, or smart home diagnosis of a health condition.

Table 5 Summative assessment of the reviewed systems

	Motivating				Engaging	Supports Daily Activities	Individualised	Suitable for elderly	Not-wearable	Not-intrusive	Intended use			Cost Effective	Technologically Complex	Transferable
	Cooperative	Supportive	Constructive	General							Monitoring	Rehabilitation	Diagnosis			
[51]				✓								✓			✓	
[103]										✓		✓			✓	
[105]								✓		✓		✓				
[106]								✓		✓		✓		✓		
[109]				✓								✓			✓	
[110]				✓		✓	✓			✓		✓			✓	
[111]								✓		✓		✓		✓		
[112]										✓		✓				
[114]								✓		✓		✓		✓		
[115]				✓		✓		✓		✓		✓				
[116]								✓		✓		✓		✓		
[88]				✓						✓		✓			✓	
[89]				✓				✓	✓	✓		✓				
[95]				✓	✓	✓				✓		✓			✓	
[49]									✓	✓		✓		✓		
[97]	✓	✓	✓	✓	✓			✓		✓		✓			✓	
[135]							✓	✓	✓	✓	✓		✓			✓
[144]							✓	✓	✓	✓	✓			✓		✓
[150]							✓	✓			✓			✓	✓	✓

In the analysis, systems were considered that could be purchased by an average household in the United Kingdom to be cost-effective. Systems that would require a high capital investment were considered, and thus reimbursement from the health care provider, to be not cost-effective. Technologically complex systems were considered to be those that had a significant number of components, required significant training before use, or required extensive installation to be usable in a household. Finally, transferable systems were those that could be used for other rehabilitation purposes and were not restricted to stroke rehabilitation.

As the tables show, no technology met all the selected criteria. Most of the technologies were suitable for the elderly and were nonintrusive. However, most technologies lacked motivation and engagement enhancement through the use of a variety of motivation methods. The developed approaches were technology centric, whereas a person-centric approach is necessary to keep patients engaged and motivated in achieving their rehabilitation goals. Several devices claimed to enhance motivation but produced little or no evidence of patient engagement [51], [97], [109], [110], [87], [93], [111]. None of the devices intended for rehabilitation were transferable to other uses. Devices intended for monitoring or diagnosis had the desired transferability features [135], [150]. Only 1 of the reviewed technologies proposed for rehabilitation supported individualization [110]; however, it did not meet the requirements for elderly patients and it used wearable components. On the other hand, individualization was supported by monitoring devices that were not intended for rehabilitation use [135], [142]. Several technologies we reviewed were inappropriate for home rehabilitation, as they were technologically complex and expensive.

3.9 Chapter Conclusion

The first rows of Table 6 list all the selected criteria which we review in this paragraph, drawn from our extensive literature review. The system needs to avoid wearable or intrusive components. It needs to support enhanced motivation and engagement by being incorporated into the daily activity routine. It must be cost-effective and not complex to install, maintain, and use. It needs to support the needs of all patients, regardless of age and background. Moreover, it needs to be portable and transferable to other domains such as diagnosis of co-morbidities.

The successful design of an assistive technology or rehabilitation device should take under consideration what the individual should and can achieve during rehabilitation [28]. Quantification and further analysis of the present and future conditions of the patient could overcome difficulties and unforeseen circumstances and could result in better assistive technology design.

It is important to tailor rehabilitation to the patients' requirements and goals, adapt to their individual needs, and provide suitable challenges. Individual choice and personal control are mandatory for success. Technology design has to follow a person-centric approach considering technology ability levels. Given the developments in smart devices, algorithms, and information extraction, devices can adopt a person-centric approach while meeting the requirements for cost and complexity.

Thus, a system catering to every occasion, individualised and adapted to support the patient's daily activities in their home environment, has a higher potential for successful acceptance and engagement, but is a challenging prospect. Hence, the successful system should focus on supporting specific daily activities that have measurable outcomes specified in recognised health care rehabilitation tests. In summary the contributions of this chapter are addressing RQ1 as identified in Chapter 2 and are:

a) The criteria for a successful home-based rehabilitation system;

b) The identification of the gap in literature.

4 Design of Home-Based Rehabilitation System & Evaluation Methodology

In Chapter 3 a review of the relevant literature was performed and a taxonomy was presented (Figure 10). The importance of monitoring and the pattern identification was discussed and criteria for a successful device, along with summarised findings presented, published in [56].

Through Chapter 3 a significant gap in the literature was identified. This chapter firstly will review clinical tests widely used worldwide and the selection criteria for the tests that are most relevant to home-based rehabilitation, daily activities, and could support a system that will satisfy the criteria discussed in Chapter 3. Then systems which offer automated solutions for the selected tests will be reviewed.

These tests will be used to evaluate a home-based rehabilitation system through experimental patient case studies, thus, addressing both RQ1 and RQ2 as defined in Chapter 2. The methodology will be presented along with the participants' profile. Then the analytical description of the experiments will be presented. Finally, the chapter will be concluded with the presentation of experimental and questionnaire results. These will provide the basic dataset for further development of pattern identification to build an intelligent system that satisfies the full extent of the criteria presented in Chapter 3, directly addressing RQ3 and RQ4 (Chapter 2). This intelligent system will be further discussed in Chapter 5.

4.1 Introduction

Since the patients are meant to interact with rehabilitation systems alone after receiving some training, without support from a specialist, acceptance and lasting engagement are crucial. However, home-based rehabilitation equipment that fulfils the aforementioned criteria usually cannot meet the specifications of clinical rehabilitation systems in terms of rigor. Therefore, home-based rehabilitation equipment must be rigorously evaluated against specific and measurable medical tests [151] in order to meet medical standards. These tests combine multiple daily activities such as walking, sitting and standing. A detailed literature search performed in Chapter 3 reaffirms the previous findings of [152] that existing automated self-evaluation systems do not meet the above identified criteria, particularly in terms of acceptance and low-cost requirements.

This chapter proposes practical methodological steps for evaluating new home-based rehabilitation systems in terms of meeting the medical specifications and the acceptance criteria (Chapter 3). To demonstrate the proposed evaluation methodology, we evaluated a home-based rehabilitation system that satisfies the four criteria of patient acceptance, and evaluate its performance against medically accepted standard tests, as discussed next.

4.2 Patient Evaluation Tests

Different tests exist for the evaluation of gait, balance and mobility of subjects. These tests are used to measure/evaluate specific characteristics relevant to the subject's clinical condition. The outcomes of these tests help identify underlying illnesses or support recovery after an illness has occurred (e.g., post stroke) [153]. Most well-known and used tests are presented in Table 6. The functional reach test would require a combination of several sensor systems, including wearable sensors to capture vestibular motion. Berg balance scale, performance-oriented mobility assessment, and balance evaluation system tests all assess static balance and posture; but require wearable or intrusive sensing techniques as well as a specialist being present during the tests. Hence, the aforementioned tests are not suitable for self-assessment and home rehabilitation, where specialists may not be present. Falls efficacy scale and balance confidence scale self-evaluation are carried out via a questionnaire to describe daily activities. To automate and monitor all the activities covered in the questionnaire, the system cost would increase significantly. The balance error scoring system (BESS) test is targeting the younger segment of adult population and particularly athletes, with tasks that could be challenging for the less mobile. Finally, the timed up and go (TUG) and five times sit to stand (FTSTS) tests can be characterised by their simplicity, accuracy and suitability for all adults. Furthermore, TUG and FTSTS cover multiple activities with one test, and can be monitored by systems meeting the four criteria of patient acceptance, both in home and clinical environments. Due to a variety of reasons, including socioeconomic [154], automated solutions for some of these tests have appeared. A significant motivator for the automation of these tests is the elimination of human error [152], [155], [156]. Indeed, in the majority of non-automated tests, the time is measured using a stopwatch, which inherently incurs human error [157].

Table 6 Medical Tests to Assess Patient Activities

Test	Measured Capacity
Functional Reach Test	Dynamic Balance [158]
Berg Balance Scale	Dynamic And Static Balance [159]
Performance Oriented Mobility Assessment	Dynamic/Static Balance And Gait Abilities [160]
Balance Evaluation System	Overall Balance. Tests Include Sit To Stand Test, Rise To Toes, Stand On One Leg [161]
Falls Efficacy Scale and Balance Confidence Scale Self-Evaluation	Subject Ability /Confidence In Carrying Out Daily Activities [162]
Balance Error Scoring System (BESS)	Static Postural Stability [163]
Timed Up and Go (TUG)	Mobility Static And Dynamic Balance [153]
Five Times Sit To Stand (FTSTS) Test	Lower- Limb Functionality, Durability And Balance [164]

The TUG test can be carried out in the home environment with a non-clinical assistance. However, for safety purposes either the system should automatically raise flags and specialist can be notified or a family member could be present. The only tool needed is a stopwatch to measure the time to complete the test [165]. The test algorithm is relatively simple, combines more than one daily activity and contains performance thresholds, as defined by the NHS of the United Kingdom. For example, completion time exceeding 15*sec* identifies a patient at risk of falling [165][166].

However, factors such as age, gender, different levels of impairment or other medical conditions, can affect the accuracy of this assessment. Thus, different thresholds have been proposed to incorporate these factors, as presented in a study of 2084 subjects in [167]. To investigate, validate and evaluate the transferability of the automated sensor system, a second test was selected. The FTSTS test also incorporates basic motion linked to daily activities [168], but it complements TUG in assessing strength of lower limbs and durability, is time based and identifies fall risk [169]. Both TUG and FTSTS tests are approved by the NHS.

4.2.1 Automated Timed Up and Go Test

In the literature, different methods have been proposed in order to automate the TUG test mostly using intrusive means such as cameras [170][171]. Such a system, in [172], is able to analyse stability and classify the subjects on susceptive or non-susceptive to fall, while in [173] the system monitors and analyses each phase of the test. Recent approaches use advanced webcam sensors [174][175][155] such as Microsoft and improve accuracy. These studies given the fact that their design is based on video technology, have different constraints such as camera positioning, people interference, lighting issues, floor quality for the Kinect sensors and most important privacy issues.

On the other hand, research has focused on wearable approaches [51] [112]. In [176], the TUG test was performed along 7 meters instead of 3 meters and has been divided in four phases for most accurate measurements. In [177][178][179] the focus is on fall prevention and tele-evaluation by specialists. In [180] the time needed for TUG stages completion is measured to distinguish fallers from non-fallers. Similarly, TUG test was used for hemiplegic subjects study in [181] and it was capable of distinguishing the hemiplegic subjects where supervision was necessary. In [182] a TUG test has been carried out with Alzheimer's illness and mild cognitive impairment subjects, in order to identify successfully differences in gait characteristics among two categories.

There are applications based on mobile phones that rely on wearable sensors [183][184]–[186]. For example, [187] places the phone on patient's waist. The approach proposed in [188] has been developed for self-assessment; it is able to capture different parameters and through a user interface all the results can be displayed on the mobile screen.

Although mobile and wearable sensors applications are quite promising and some of them easy to use by healthy subjects, for elder people as we have aforementioned, these types of applications are not always welcome and, in this thesis, will not be analysed further. This could happen for a variety of reasons for example: due to the requirement of specialists in order to place the sensors on the body or could be due to the complexity of displayed information on mobile screen/user interface.

Another way that has been developed in order to automate the TUG test is with the help of ambient sensors. Ambient sensors can be defined as two different main categories of sensors which could be either wearables or non-wearables and able to be integrated in subjects' environments. Although the wearable ambient sensor systems share the same disadvantages with other wearable systems which do not utilise ambient technology, ambient sensors which could be attached on transducer boards or being integrated on stand-alone systems appear to be an attractive solution given that they are not intrusive, could provide a solution tailored to the user needs, they are easily extendable, they can be integrated into daily environments (in the house, for example) and they can easily be readjusted to new needs of the individual [189].

Although, ambient systems have been developed in order to capture gait speed using PIR sensors [190], [191] they have not been designed in such a way in order to automate the Timed up and go test. However, in [192] one fully automated timed up and go test has been proposed which utilises ambient sensors. However, during the experiment procedure, the subjects were asked to wear white cuffs on their feet. Moreover, in order to transfer the data captured due maybe to high volume, a cable was necessary. Furthermore, this system does not provide any feedback to the user and thus a specialist should be present.

4.2.2 Automated Five Time Sit to Stand Test

In order to investigate, validate and evaluate the transferability of a device, the FTSTS were carried out. Due to a variety of reasons, such as human error when measurements are taken with the help of stopwatch [157], different technological approaches have been developed in order to provide automated solutions for better accuracy. There is a variety of different ways which have been proposed in literature and utilise different means in order to achieve test automation.

Wearable or camera-based systems have been mostly investigated for the automation of FTSTS [193][194][195]. In [196] cameras and coloured markers placed along left side of the subject are used for kinematic analysis. In [197] a Microsoft Kinect sensor has been employed along with VR. For validation of the test a stopwatch measurement and recordings for 2 video cameras were taken.

Moreover, ambient sensor systems have been utilised to monitor and automate the FTSTS test. In [198], [199] ambient sensors are used to evaluate different phases of the test, for example time of sit to stand, stand to sit etc.

4.3 Methodology

In this section, the methodology to assess the accuracy of home-based rehabilitation systems for the TUG and FTSTS tests were described , as illustrated in Table 6 and Section 4.2. For that purpose, a home-based rehabilitation sensor system was designed developed and deployed, which satisfies the four criteria of [56], namely non-intrusiveness, does not contain any wearable component, is easy to use and low cost. The system comprises two time-synchronised blocks, each assembled using a micro-controller and a modified BISS0001 passive infrared (PIR) sensor.

To identify suitable sensors for the experiments a variety of sensors were reviewed. The main reasons for selecting BISS0001 PIR sensor were: (a) the low cost of the sensor, (b) the capacity of the sensor to be altered or modified in order to comply with the extended needs of the system, (c) the size and footprint.

When it comes to home monitoring or home rehabilitation of subjects, during the design phase, further consideration of various sources of noise need be taken into account, which can be introduced during monitoring. Such noise could be sourced either from other residents of the house or from different objects that they can be moved and accidentally block either the capture angle of the sensor and/or the whole view. This is where the system is able to identify that something in the room or home condition has changed.

The sensors were used to capture the time a subject took to walk between two points as he/she performed the TUG test (through horizontally spaced sensor blocks) (Figure 12). Replacement of the electrical components as well as optimisation of the lens yielded sensitivity range of up to 1 m and capture angle of 30° x 30°. The digital signal was adjusted to remain high after the trigger for a variable time between 0.25 sec to 25 sec, depending on subject speed. The FTSTS test was performed through a vertical arrangement of the two aforementioned sensor blocks, which measured the time from sit to stand. The sensor allowed for vertical and horizontal arrangement. The two sensor blocks were optimally placed for both tests to minimise false positives (i.e., when a motion that was not part of the test was picked up), and false negatives (i.e., when a motion that was part of the test was not picked up).

Though applied to this system for demonstration purposes only, the proposed evaluation methodology described in the next sections is generic and can be followed for assessment of other similar home-based rehabilitation systems.

Figure 12 Subject performing TUG with the deployed system at National Centre of Prosthetics and Orthotics. The head is covered for subject privacy.

4.3.1 Participants

Participants were recruited and the experiments took place in the National Centre of Prosthetics and Orthotics at the University of Strathclyde, after ethical approval UEC16/52 was obtained by University of Strathclyde Ethics Committee.

In total, eight healthy subjects were recruited to take part in the experiment in order to evaluate the proposed system for TUG and FTSTS tests. The participant recorded parameters are presented in Table 7.

Table 7 Participant recorded parameters.

Participant	**Height (m)**	**Weight (kg)**	**Age (years)**	**Sex (Female (F)/ Male (M))**
1	1.5	56	34	F
2	1.7	75	25	M
3	1.82	84	41	M
4	1.7	86	25	M
5	1.84	80	22	M
6	1.85	77	28	M
7	1.74	62	27	M
8	1.8	92	40	M

The subjects were recruited to illustrate the proposed methodology and to demonstrate if a system (any system) meets the desired requirements. The subjects were seven males and one female aged 22 to 41 (mean age $\overline{X}$ = 30.25). The mean height of the participants was $\overline{X}$ = 1.74 cm, with standard deviation of σ = 0.12 and weight $\overline{X}$ = 76.5 kg with σ = 12.14. This range of heights and weights allows for evaluation of the technology in a variety of scenarios even though it is predominantly representative of characteristics relevant to male adults[200] [201].

The eight subjects were over the age of 18 with good vision (with or without corrective aids), the author of this thesis provide a participant information sheet explaining the procedure and were able to provide consent, and instructed them through the process. They followed instructions in English and attended a single appointment at the lab. Exclusion criteria were used for subjects that were pregnant, had a hearing and/or visual problems that was not corrected, subjects that were unwell or were taking medication potentially compromising their ability of mild physical activity, subjects with significant speech problems affecting the safe execution of the experiment and subjects with vestibular impairments, heart or respiratory conditions that limited their ability to walk. The study was approved by the University of Strathclyde ethics committee and a data management plan for data security was in place. Following a similar approach to [170], subjects were asked to simulate various disabilities. The simulated disabilities and number of repetitions are discussed in Section 4.3.3 for each test. The total number of individual experiments were 184 for TUG and 40 for FTSTS.

To demonstrate that the evaluated technology is able to record TUG and FTSTS results that are relevant to a wide range of adults, we compared the TUG and FTSTS completion time recorded during the simulations to the completion time reported for healthy and geriatric elderly (>65 years) and adults (>18 years) in the Shirley Ryan Ability Lab [201] international database. For TUG test, a total of 48 studies were analysed in the database of 6632 participants with a variety of conditions. Of those, only 10 studies report male/female populations; seven are predominately male at 67.3% on average, while three have male populations of 36.3% on average. For FTSTS test, a further 23 studies were analysed with a total of 7794 participants in the adult and elderly groups. Of those, only two studies report male/female populations at an average of 48.7% male predominance. A two-sample t-test analysis was performed with the hypothesis that recorded completion time distribution for each of the simulations was equal to completion time distribution reported in the database for conditions relevant to the simulated disabilities. The selected groups are presented in Section 4.4. As the female representation in the database is on average 49.23%, the hypothesis will also further support our experimental results not being overly biased towards male participants. The experiment procedure for the pilot validity study is described in Section 4.3.4 for the TUG and 4.3.5 for FTSTS, respectively. Alongside the system an assistant was present holding a stopwatch to record time of completion. Moreover, for the gold standard a camera recording was used capturing the whole scene. The camera was setup on a tripod for stability and was on eye level of the subject taking the experiment when standing. Ceiling lights were turned on to avoid blur and glare and there were no windows in the room. The angle of filming was setup to avoid participants blocking any segment of the test from being recorded. A professional camera was used to record high quality video.

4.3.2 Statistical Data Analysis Measures

To identify agreement between the results obtained by the automated sensor system and the video measurements the Bland–Altman 95% bias analysis was carried out. This method is widely used in the medical field when comparing two measurements of the same variable. For each pair of measurements, the x-axis illustrates the mean and the y-axis the difference. The method also provides the lower and upper level of agreement and establishes acceptable limits [202]. The percentage error (PE) of the measurements of the experiment is calculated following the Bland–Altman method based on the upper and lower limit of agreement (LOA) according to [203] (Equation 1):

$$PE = \frac{Upper_{LoA} - Lower_{LoA}}{\overline{X}} \quad (1)$$

where:

$PE \rightarrow$ Percentage error

$Upper_{LoA} \rightarrow$ Upper level of agreement

$Lower_{LoA} \rightarrow$ Lower level of agreement

$\overline{X} \rightarrow$ Mean value of set

Lin's concordance correlation coefficient (ρc) (Equation 2) was calculated to compare the proposed automated sensor system measurements against a "gold standard" or "ground truth" measurement as one of the most well-established methods to assess agreement [204], as follows:

$$\rho c \; = \; \frac{2 * \rho * \sigma_{system} * \sigma_{video}}{(\overline{X}_{system} - \overline{X}_{video}) + \sigma_{system} + \sigma_{video}} \quad (2)$$

where ρ is the correlation coefficient, σ_{system} and σ_{video} represent the standard deviation, of the automated sensor system and the video system, respectively, and $\overline{X}_{system}$ and $\overline{X}_{video}$ are the mean of the automated sensor system and video system data points, respectively.

The intraclass correlation coefficient (ICC) was calculated as assessment of the reliability of the measurements [205]. The ICC was evaluated after conducting analysis of variance of two factors without replication. Finally, linear regressions analysis was performed to obtain accuracy, quantification and data trends.

4.3.3 Experiments

For each of the TUG and FTSTS tests that was conducted, data were recorded through:

1. a non-intrusive, non-wearable, low cost, motivation and engagement enhancing system that can be individualized, is simple and transferable;
2. a stopwatch following the instructions for specialists according to the NHS suggested method [165];
3. and a standard video camera as golden standard to avoid human error in the stopwatch method.

The automated sensor system is able to capture motion and time. It is a portable system, easy to use and set up. The placement of the components depends on the participant's biometric characteristics in order to collect and extract accurate data during the experiments. For each participant, the system has to be calibrated to the individual, as presented for each test in the following subsections. The test completion time is crucial, given that slower time than normal could be an indication of a medical condition. In all of the experiments the time of completion was measured.

At the end of every experiment repetition, the subject was receiving simple feedback through a screen. The feedback was based on the performance and the time of each completion. If there was an improvement over time in comparison with the previous repetition the feedback was illustrated as a happy green face; for stable performance the feedback was a yellow face and for degraded performance in comparison with previous repetitions a red sad face. This simple feedback was considered easy to interpret according to the relevant literature review in Chapter 1.

4.3.4 Timed Up and Go Test Experiments

Inline with NHS instructions, subjects were seated on an armed chair, and on the word "Go", the subjects would stand, walk 3 meters on a straight line, make a 180°degree turn, walk back to the chair, turn and sit down (Figure 13).

For calibration, each participant was asked to complete the TUG test at their normal walking speed (own pace) three times. Next, to investigate the properties of the system under a wide range of conditions, the participants were asked to simulate three impairments, and motion at an accelerated pace. First, they were asked to simulate reduced ability, or difficulty, to stand (Figure 14). The subject was trying to stand up by spending time on various positions or by performing unsuccessful attempts.

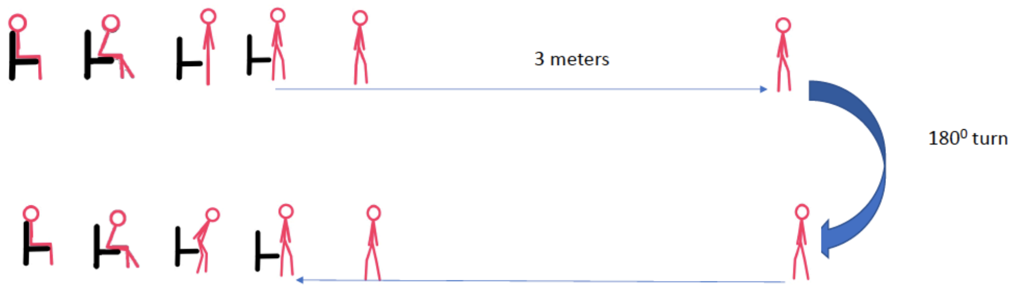


Figure 13 Timed Up and Go Test Experiment

This was a way of testing that the device and the motion sensor will be capturing and transmitting data accurately with the right angle and range without resetting and recapturing the particular subject multiple times. The task of sitting down was performed in a similar manner during this simulation.

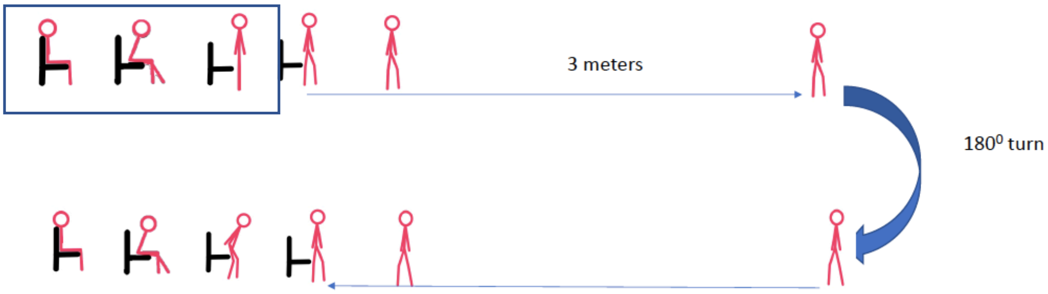


Figure 14 Timed Up and Go Test Experiment Simulating Difficulty to Stand

Next the participants were asked to simulate a reduced ability, or difficulty, to walk as it is often the case for patients with reduced mobility even if the distance is limited to 3 meters (Figure 15). Subjects were asked to slow down in order to ensure that the time captured will be the time of the worst-case scenario (i.e., the time a geriatric elderly would need to perform this test). The aim was to evaluate the system's ability to accurately capture the overall time, without system resets and without the sensor recording interference from the testing area.

Finally, the participants were asked to simulate reduced ability or difficulty to turn by delaying when they were performing the 180° turn. This was simulated as wobbling or assuming the need of a walking aid (Figure 16). Here, the subjects were asked to simulate imbalance while turning to capture the motion that are relevant to this stage.

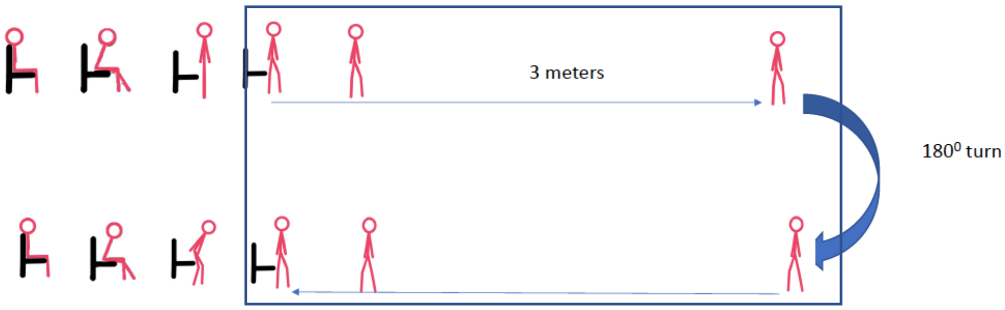


Figure 15 Timed Up and Go Test Experiment Simulating Difficulty to Walk

Figure 16 Timed Up and Go Test Experiment Simulating Difficulty to Turn

This part of the experiment was designed to demonstrate the ability of the automated sensor system to recognise and capture potential mobility problems by distinguishing between these stages. There were no time restrictions for the participants to carry out and complete each of the five repetitions.

Finally, the system was tested under fast walking conditions. The participants were asked to perform the TUG test as fast as they could without running. The aim of this set was to identify the limitations of very low-cost sensors, in accurately capturing fast motion. The number of times each stage was repeated is presented in Table 8.

Table 8 Timed Up and Go Test and Five Time Sit to Stand Experiments

Test	Stage	Repetitions	Simulations
TUG	1	3	Normal Walking
	2	5	Difficulty To Stand
	3	5	Difficulty To Walk
	4	5	Difficulty To Turn
	5	5	Fast Walking
FTSTS	1	3	Fast
	2	2	Difficulty To Stand And Sit

4.3.5 Five Time Sit to Stand Test

Instructions were initially given to the participants on how to perform the FTSTS test. Each subject was seated on an armless chair with hands crossed over the chest. On the word “Go”, the participant had to stand and then sit five times without support (Figure 17). The participants performed experiments at their own pace.

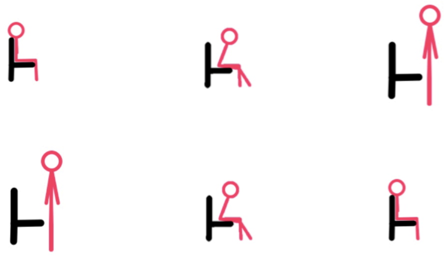


Figure 17 Five Times Sit to Stand Test Stages Repeated Five Times For Each Experiment

Next, the participants were asked to simulate difficulty to stand and sit, by performing the same task but with a delay in sitting and standing to validate if the sensor-system could accurately record these cases. Table 8 summarises the experiments.

4.3.6 Mapping of the System

Due to different body metrics of the subjects and given that the system is portable, the right placement of the system plays a vital role for the tests to be carried out successfully. Hence, for people with different heights and weight, sensors and the whole system should be placed accordingly to the subject's requirements. However, the initial system would need to be installed at the right position the first time before operation and should be tailored to the subject. Afterwards no further alterations or adjustments would need to be made. Moreover, when conducting the experiments, the system had to be adjusted frequently given that the participants had different heights and weights. The height of the participants was affecting the vertical placement of the sensors and the weight of subjects, i.e., the mass was affecting the sensors on horizontal line. Furthermore, the system consists of two different portable parts, which were located 3 meters apart.

Figure 18 presents the location of two portable components of the system (S1 and S2) and the parameters that were calculated for each subject. These were estimated based on the participant recorded parameters presented in Table 7. Then, experimenting with the device and sensitivity, the final placement was adjusted to acquire the most accurate results. Table 9 presents the final measurements after the adjustment.

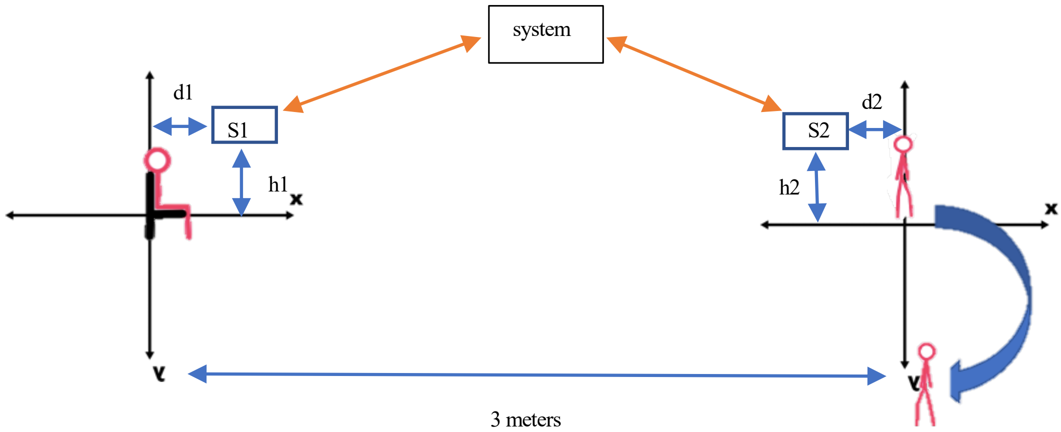


Figure 18 Location and placement of system components for TUG.

Table 9 Location and placement parameters for each participant of TUG.

Subject	d1 (m)	h1 (m)	d2 (m)	h2 (m)
1	0.4	0.915	0.60	0.5
2	0.5	0.915	0.68	0.7
3	0.6	0.915	0.73	0. 915
4	0.6	0.915	0.68	0.7
5	0.6	0.915	0.74	0.915
6	0.5	0.915	0.74	0.915
7	0.4	0.915	0.70	0.7
8	0.7	0.915	0.72	0.915

Analysing the resulting measurements presented in Table 9, through correlation analysis, the following relationships were identified (Equations 3 to 6):

$$d1 = 0.0084 * weight - 0.1037 \quad (3)$$

$$h1 = seat\ height + 0.315 \quad (4)$$

$$d2 = 0.4 * height \quad (5)$$

$$h2 = 0.8964 * height - 0.7456 \quad (6)$$

where:
weight →in kilograms (kg)
height → in meters (m)

In order to avoid any confusion, the user could select between 2 modes, the first selection enables the TUG test while the second selection enables the FTSTS test. The two parts of the system had a different placement arrangement for FTSTS as presented in Figure 19. The system was easy to operate given that the screen was a touch screen and the menu very simple for the user. However, a small remote control was available as well which was enabling distanced selection. The only thing that the user had to do in order to change the test option was pressing the number 1 or number 2 on the remote control.

A similar process was followed to identify the most suitable placement for FTSTS. Component S1 remained at the same location, while S2 was moved as presented in Figure 19. Thus, Table 10 presents the final measured location and placement parameters, for FTSTS. Equations 3 and 4 remained the same for $d1$ and $h1$. Analysing the remaining parameters, the following equations were identified (Equations 7 and 8):

$$d2 = seat\ back - 0.2 \quad (7)$$

$$h2 == 1.7277 * height - 1.5165 \quad (8)$$

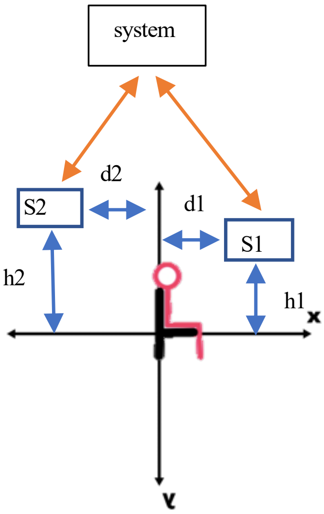


Figure 19 Location and placement of system components for FTSTS.

Table 10 Location and placement parameters for each participant of FTSTS.

Subject	d1 (m)	h1 (m)	d2 (m)	h2 (m)
1	0.4	0.915	-0.2	1.15
2	0.5	0.915	-0.2	1.35
3	0.6	0.915	-0.2	1.65
4	0.6	0.915	-0.2	1.35
5	0.6	0.915	-0.2	1.69
6	0.5	0.915	-0.2	1.72
7	0.4	0.915	-0.2	1.41
8	0.7	0.915	-0.2	1.65

4.4 Experimental Results

Table 11 and Table 12 present the $\overline{X}$ and σ values for each of the TUG and FTSTS tests, respectively. The last column presents results from studies with adult patients reported in [201]. As demonstrated in the last column, the ranges of completion time recorded in the experiments were within the reported ranges for those patient populations. A two-sample t-test analysis comparing the automated sensor system to the database supports the hypothesis that the two distributions are significantly similar for all cases, where α = 0.05 is the acceptance limit. The t-values were -2.14, 1.42, 2.51, 0.27, 2.25, 6.66 and 0.98, for each case. The probability of the two distributions being equal was calculated to $0.91 \leq P(|t| \geq t_{value}) \leq 0.99$ for all other cases, except the fast TUG set with $P(|t| \geq 0.273) = 0.61$ and the fast FTSTS set with $P(|t| \geq 0.982) = 0.83$.

Table 11 Timed Up and Go (TUG) Characterisation Of The Tests $\overline{X}$ (σ). The Results Are Given in Seconds

TUG Set	Automated sensor system	Stopwatch	Video	Patient
Walk	24.94 (9.41)	26.46(9.0)	25.54(9.31)	31.9(20.9) (geriatric age $\overline{X}$ 79.9)
Turn	26.43(8.25)	26.63(7.92)	26.78(8.1)	23.33(11.66) (Bilateral vestibular hypofunction age $\overline{X}$ 77.95)
Stand	22.32(8.65)	21.27(7.01)	22.31(8.6)	15.5(11.03) (Parkinson's fallers/ No medication age $\overline{X}$ 77.95)
Fast	8.13(2.47)	7.82(1.9)	7.73(1.73)	7.94(2.15) (Parkison's Non faller/Medication age $\overline{X}$ 66.64)
Normal	9.66(1.30)	9.80(1.8)	9.69(1.37)	8.13(2.34) (Parkison's Non fallers / No Medication age $\overline{X}$ 66.64)

Table 13 presents the coefficient of variation as a percentage for each set. To further investigate the correlation, three different categories of graphs were used for each of the tests and each simulated impairment; box plots, linear regression, and Bland–Altman.

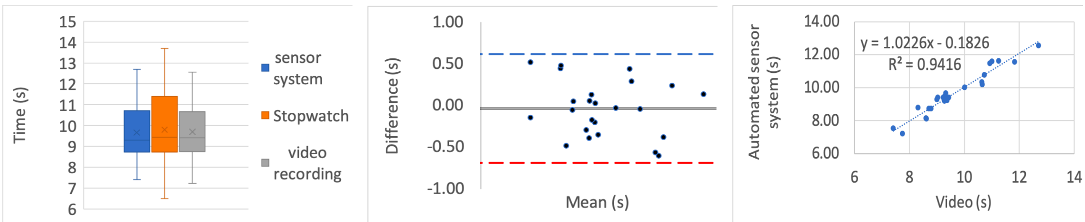
Table 12 Five Time Sit to Stand (FTSTS) Characterisation Of The Tests $\overline{X}$ (σ). The Results Are Given in Seconds

FTSTS Set	Automated Sensor System	Stopwatch	Video	Patient
Diff.	*49.71(14.58)*	*50.66(14.05)*	*49.98(14.94)*	*20.25(14.12). (Parkinson's age $\overline{X}$ 65.9)*
Fast	*17.78(5.1)*	*19.15(5.37)*	*18.87(4.93)*	*16.4 (4.4) (Vestibular Disfunction $\overline{X}$ 66.64)*

Table 13 Characterisation Of The Tests: Coefficient of Viriation Percentage (%)

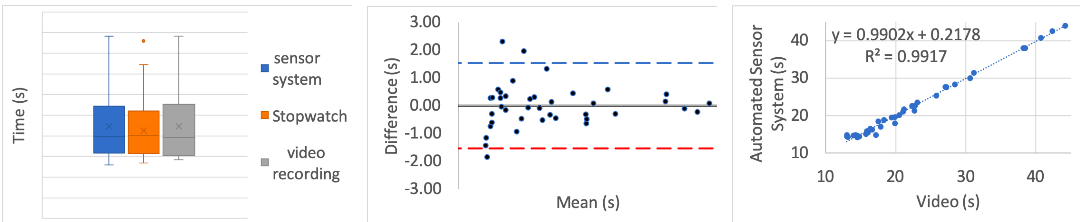
Set	Automated Sensor System	Stopwatch	Video
TUG Walk	*37.72*	*34.00*	*36.42*
TUG Turn	*31.23*	*29.75*	*30.23*
TUG Stand	*38.76*	*32.97*	*38.54*
TUG Fast	*30.46*	*24.33*	*22.45*
TUG Normal	*13.49*	*18.41*	*14.16*
FTSTS Diff	*29.33*	*27.73*	*29.90*
FTSTS Fast	*28.69*	*28.08*	*26.14*

For the box plots, the y-axis represents the time while on the x-axis the methods of recording. The whiskers which are lines anchored at the edges of the box, represent the range of measurements in seconds. The horizontal line in the box represents the mean and the x point in the box represents the median. Bland–Altman plots, in our case, show the mean and the difference between video and system measurements, respectively, with acceptance limits within 2 s. The limits of agreement were calculated at 1.96 σ (differences) per definition for the 95% bias analysis [202], [206]. The y-axis represents the difference between two measurements, i.e., video - system, while on the x-axis the mean of the measurements is shown. The horizontal black line represents the bias while the horizontal blue and red dotted lines represent the Upper and Lower level of agreement, respectively. In the linear regression graphs, the y-axis represents the measurement of the system in seconds while the x-axis represents the video measurements in seconds. The video recording was used as the "gold standard" or "ground truth" measurement for all the experiments to calculate the Lin's coefficient, as well as the regression analysis. Figure 20- Figure 26 present the results for TUG and FTSTS tests according to the simulated impairment or set of repetitions the participants were called to perform. Table 14 presents the PE, ρc and ICC results of the analysis as defined in Section 4.3 and the linear regression results of the coefficient of determination (R2) and p-value.



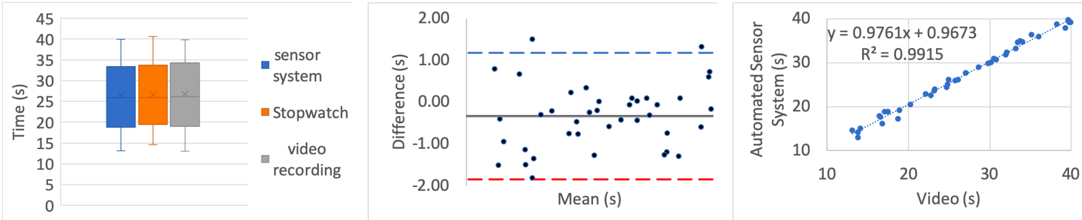
(a) (b) (c)

Figure 20 TUG Normal Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression



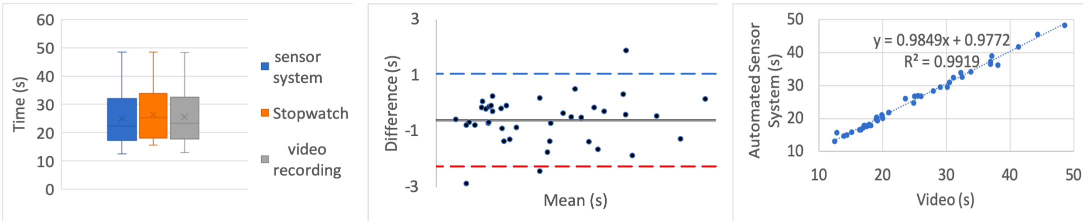
(a) (b) (c)

Figure 21 TUG Difficult to Stand Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression



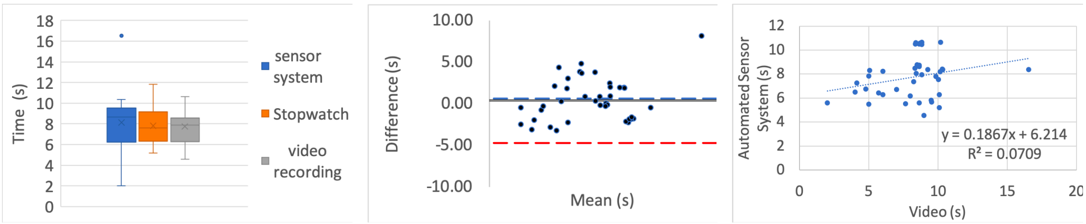
(a) *(b)* *(c)*

Figure 22 TUG Difficult to Turn Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression



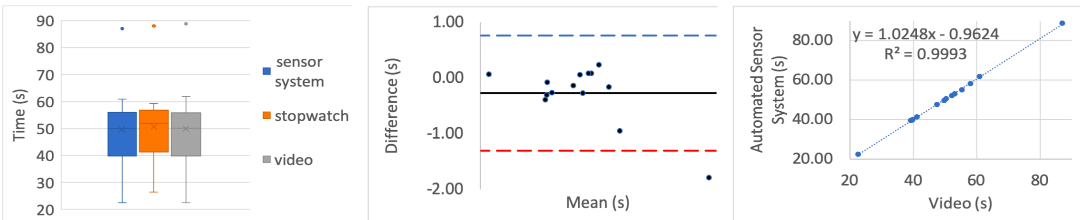
(a) *(b)* *(c)*

Figure 23 TUG Difficult to Walk Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression



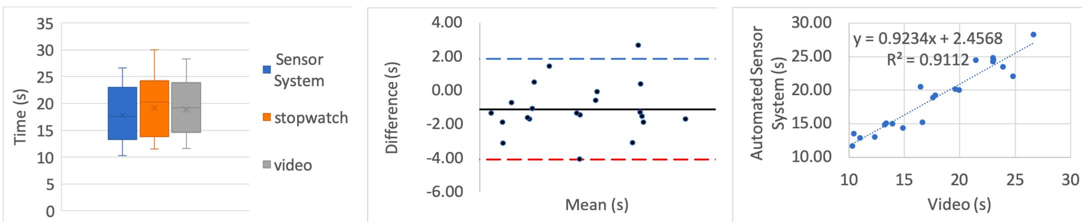
(a) *(b)* *(c)*

Figure 24 TUG Fast Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression



(a) *(b)* *(c)*

Figure 25 FTSTS Slow Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression



(a) *(b)* *(c)*

Figure 26 FTSTS Fast Simulation Aggregate: (a) box Plot (b) Bland – Altman (c) Linear Regression

Table 14 Characterisation of The Tests: Percentage Error and correlation Results

Set	PE	ρc	R^2	p-Value	ICC
TUG Walk	*13.3%*	*0.994*	*0.992*	$2.16*10^{-41}$	*0.994*
TUG Turn	*11.4%*	*0.996*	*0.992*	$5.6*10^{-41}$	*0.995*
TUG Stand	*13.8%*	*0.996*	*0.992*	$3.4*10^{-41}$	*0.996*
TUG Fast	*126.15%*	*0.246*	*0.079*	*0.097*	*0.251*
TUG Normal	*14.1%*	*0.997*	*0.942*	$4.67*10^{-15}$	*0.969*
FTSTS Diff	*4.73%*	*0.999*	*0.999*	$1.81*10^{-20}$	*0.999*
FTSTS Fast	*31.1%*	*0.931*	*0.931*	$1.93*10^{-11}$	*0.934*

4.4.1 Questionnaire

Each participant was asked to complete a questionnaire at the end of the experiment to identify the level of motivation and engagement. A total of five questions were provided with the opportunity to provide general comments at the end: (1) Was the device easy to use and set up? (2) Was the feedback sufficient? (3) Would you use the device again? (4) Was the device engaging? (5) Did the device increase your motivation for performing the task? (6) Any other thoughts, detailed responses to the above questions, recommendations, or general comments?

The participants were asked to respond on a scale from 0, meaning very poor, to 5 meaning excellent, to questions 1 to 5. The last question was open ended to allow for further feedback.The responses from the collected questionnaires were analysed by simple sum and percentage proportion analysis for each of the possible responses (Figure 27).

All of the subjects found the system easy to setup and use (Question 1: 100% ≥3) meaning that all eight responses were equal to 3, 4 or 5. 83% found the received feedback sufficient, good or excellent (Question 2: 83% ≥ 3). Additionally, all the participants found the system engaging (Question 4: 100% = 5) and that it increased their motivation to perform the task (Question 5: 100% = 5). Finally, the majority would use the system again if they ever needed home self-rehabilitation (Question 3: 88% ≥3).

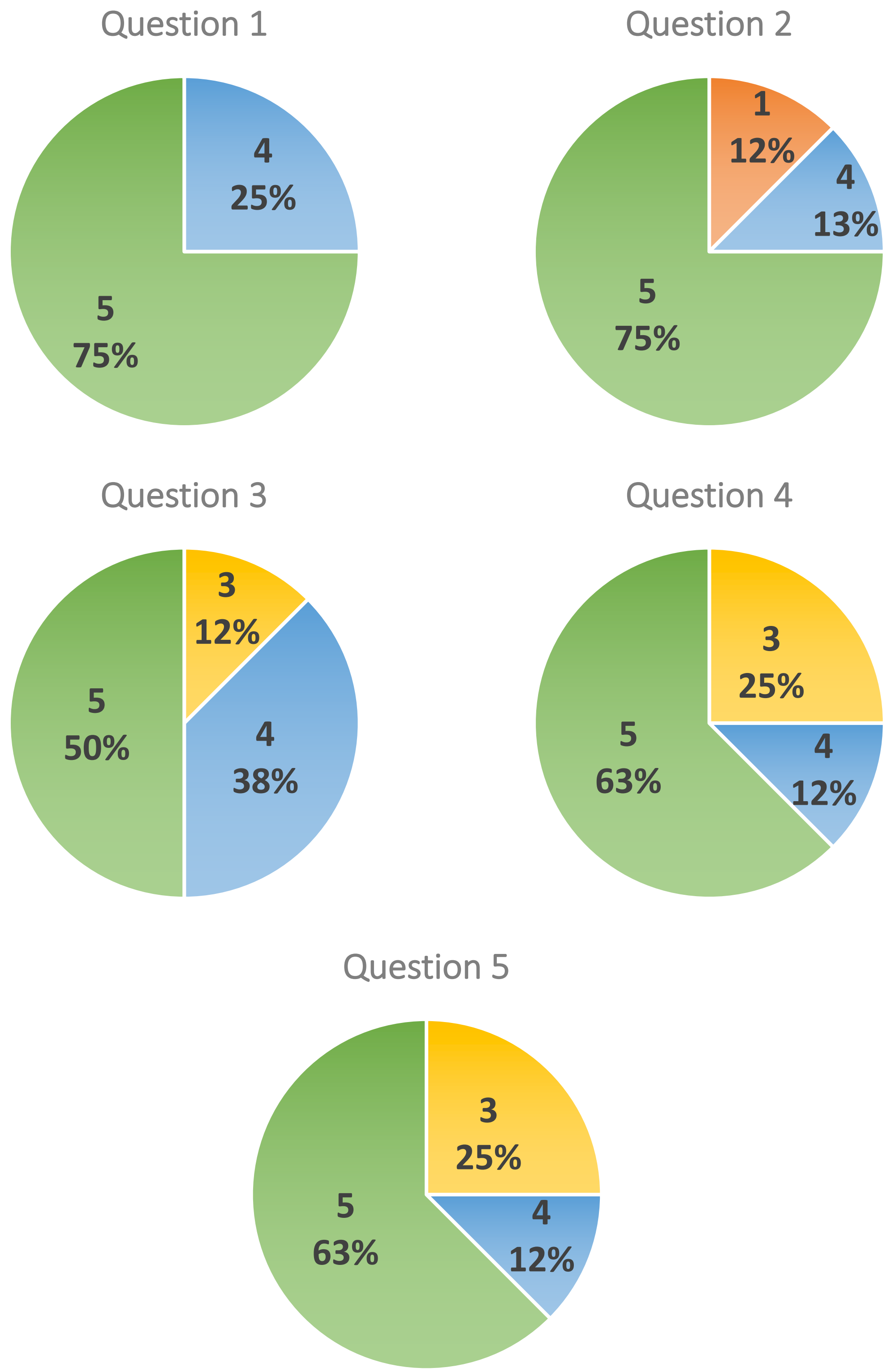


Figure 27 Questionnaire Results in terms of start-rating for each of the questions 1 to 5. Category 5 corresponds to the highest (good) rating while category 1 corresponds to the lowest (bad).

4.5 Chapter Discussion

In this chapter various evaluation tests were reviewed as well as a variety of related systems which automate clinical tests such as FTSTS and TUG. A range of difficulties were simulated using healthy subjects. There was a significant difference between participants, as evidenced through the range of results presented in the box plot analysis. The automated sensor system has consistently produced box plots that are well aligned with the video recording (gold standard ground truth). The results of the simulated difficulties were hypothesised to represent adult participants with reduced mobility. The hypothesis was tested against the international database's records for male and female patient groups. The result of the two-sample t-test hypothesis testing confirmed the validity of the assumption. As a result, the device is suitable for recording TUG and FTSTS tests for a wide range of the patient population.

It was further hypothesised that the system will demonstrate the same behaviour if used by female and elderly subjects. This hypothesis was supported by the comparative analysis to the international database where the adult groups had a strong female participant population (49.23% average of the studies that reported male/female ratios on TUG and FTSTS).

The stopwatch has lower correlation with the video as demonstrated from the box plots through displacement line. This difference is likely to be due to human error. Due to the significant variability in recording time, the stopwatch measurements were not regarded as the "ground truth" for the comparative analysis and were not used for the Bland–Altman analysis. Additionally, the box plot analysis highlights the presence of outliers in the TUG fast series and the FTSTS simulated disability case. The TUG fast series produced generally the poorest results as the participants were performing the test at a speed that exceeded the capabilities of the evaluated system. Thus, the outlier in this case can be assigned to a system fault in capturing fast completion times. An outlier of the FTSTS test, on the other hand, is present in all of the three measurement methods and thus can be the result of a particular participant taking long pauses while performing the test. For the Bland–Altman analysis all the data points (>50%) were between the limits of agreement. Points outside the limits of agreement are present in the TUG difficulty to stand simulation, turn simulation and the FTSTS fast test.

However, in all cases these points are still very close to the LOA and do not statistically significantly affect the agreement between the two measurements. The bias was in most graphs close to zero, which reflects an unbiased relation between the measurements. It is worth noticing that in the graph which represents the TUG fast simulation, the upper limit of agreement coincides with the bias and it is obvious that the automated sensor system has limitations in recording fast repetitions. By observing the linear regression analysis, can be distinguished that, even in the worst case, i.e., the tests of TUG normal and FSTS fast simulation, the correlation is statistically significant. However, the TUG test fast simulation is completely uncorrelated, demonstrating that the automated sensor system is incapable of capturing fast movements. However, these recordings would be only relevant to adults who are potentially not in need of rehabilitation. The remaining TUG and FTSTS test sets are highly correlated with perfect alignment on the linear trend line. The above observations are further supported through the ICC, ρc and R^2 results.

In the next chapter a review of various algorithmic approaches and Machine Learning methods will be presented and experiments will be carried out in order to identify the best algorithmic approach to provide motivation and engagement enhancing feedback tailored to patient needs for monitoring rehabilitation progress and at the same time be able to predict potential development of co-morbidities.

4.6 Chapter Conclusion

The low cost, non-intrusive, non-wearable, motivation and engagement enhancing system that can be individualised and support daily activities, is cost-effective, non-complex and transferable has excellent correlation and agreement with the video recordings in all the simulated conditions. The stopwatch measurements have an inherently higher PE compared to the golden standard video measurements due to the human error factor. Moreover, the transferability of the automated sensor system is presented with the FTSTS test simulation demonstrating excellent accuracy and correlation to the video recording. The relevance of this early technology to the patient population was demonstrated through comparative analysis with the international database. However, experiments with elderly subjects will be required as further evaluation steps.

Fast FTSTS (R^2=0.92) was not as accurately captured while the fast TUG test was uncorrelated between system and video (R^2=0.07). The limitation of the very low-cost motion detection sensor is apparent in these two sets of experiments as the sensor's delay in recording the event is significant and affects the recorded time. However, as the system is designed to be utilised for rehabilitation and incorporation of daily activities for increased engagement, the range of fast TUG is assumed with the scope of the study targeting less capable adult subjects. Thus, the automated sensor system is fit for purpose and has been validated for use with statistically significant accuracy (ϱc>0.99, R^2>0.94, ICC>0.96). In summary the contributions of this chapter are addressing RQ2 identified in Chapter 2 and are:

a) Deployment of a low-cost system to automatically perform the TUG and FTSTS medical tests.
b) A detailed methodology to assess a home-based rehabilitation system's accuracy against the test specifications, benchmarked against NHS standard practice and ground truth established through video recording.
c) Demonstration of transferability to other daily activities and more than one NHS test.

5 Machine Learning-Enabled Individualised Home-Based Rehabilitation

In Chapter 4 clinical tests that are widely known at global scale for patients' evaluation and diagnosis of various conditions were reviewed. Then review of systems which offer automated solutions for the selected TUG and FTSTS tests was conducted. These can offer automated evaluation of patients through a defined process. The methodology for the conducted experiments was presented along with the participants' profile, the system mapping and the presentation of experimental results. These results will provide a significant building block for the present chapter. They will be used to classify difficulties in activities relating to home-based rehabilitation. Moreover, the same tests will be used to evaluate conditions.

In this chapter the process of the FTSTS and TUG tests for home-based rehabilitation will be automated and additional early diagnosis for potential co-morbidities for stroke survivors will be provided. This will be achieved by applying Machine Learning (ML) to the system proposed in Chapter 4. However, in order to apply ML to provide interpretable and engaging information (RQ4 in Chapter 2), in Section 5.1.1, then the state of the art will be examined in terms of Accountable, Responsible, and Transparent (ART) Artificial Intelligence (AI) in terms of ethics, transparency, regulation & control, socioeconomic impact as well as design requirements and responsibility. Then, in Section 5.1.2 the organisation of the chapter, contribution, and how it is linked with the objectives of the thesis will be discussed. In Section 5.1.3 medical systems in general and in Section 5.1.4 home rehabilitation systems in particular which utilise ART will be reviewed. Discerption of the proposed methodology will be presented in Section 5.2 in terms of designing the datasets and it will be continuing by reviewing and selecting appropriate ML algorithms. An implementation and comparative analysis of algorithms will be presented based on accuracy of their predictions for the defined problem. Then a hybrid ML approach will be proposed followed by further evaluation (Section 5.3). The chapter will conclude by presenting evaluation results. Thus, this chapter will address both RQ3 and RQ4 as defined in Chapter 2.

5.1 Background

As it was discussed in Chapter 3, and presented in Table 6, one of the main factors for a successful rehabilitation system is to be able to provide an individualised rehabilitation approach. Existing approaches lack in individualised therapy. As a result, patients lose their interest quite quickly and this results in poor rehabilitation outcome. In general, individualised therapy is very challenging for an automated system to provide, given that the system should be able to understand and learn from the user, and continuous adapt based on the user's progress and preferences. Different patient conditions, daily habits and comorbidities also must be taken into account.

The basic research direction is to develop a home-based rehabilitation system based on an incremental ML approach that can learn a user profile. Such an approach should take under consideration and combine different parameters and features of the user such as BMI, weight and height. Hence, it should provide a tailored home-based rehabilitation approach by increasing gradually the level of difficulty of the tasks, constantly checking for any comorbidities that could affect the user/subject's progress, identify daily patterns and raise flags or notify appropriate care givers in case of emergency.

There are several challenges and considerations that must be addressed in terms of design, ethical implications, transparency, regulation and control, and responsibility, especially when the system is intended for medical use. In this chapter, considerations in relation to home-based rehabilitation will be examined and a novel approach to address them will be proposed.

5.1.1 State of the Art in Applied Artificial Intelligence

In order to analyse what responsible AI stands for, the impact of different factors around it which are interlinked and presented in Figure 29. In the following list each of these factors are introduced separately. The link of the proposed approach to each of these categories will be presented in the following Subsection 5.1.2 and evaluated and discussed in Subsection 5.2.6.

Ethics & Accountability: AI applied to power networks, security, smart home entertainment, and driverless cars demands that ethics should be taken under consideration in the design phase. This process should ensure that the particular AI system will not be violating human moral code [207], [208]. This links to the need of unbiased data used to train and develop ML models as well as the need to understand any unintended and possibly unethical consequences (Accountability). Instead of dealing with a “black box” AI model, the designer should pay attention to completely understand the behaviour of the particular system (Interpretability) and improve its alliance with the moral code of the intended user. Important aspects here include the beneficence or maleficence embedded into the AI model by design. Ethics play an important role of reducing bias and hence can eliminate discrimination and improve adherence to human moral code by machines.

Figure 28 Responsible AI

Transparency: To achieve better adherence to ethics there should be an ability to justify AI model decisions. An adequate level of explainability of decision making should be provided. In order to eliminate obscurity, different solutions could be applied such as a complimentary combination of simple and more advanced models, post-hoc analysis, and input modification to improve transparency [208][209].

Regulation & Control: AI intended and unintended consequences in society [208] impose the need for a regulatory framework. Hence, there is a discussion for a legal frame establishment (regulation/control) in order to impute liabilities to whoever is responsible for the model's action. Hence, according to AI application different actions have been proposed such as, issue of certificate of "safe AI", continuous monitoring and control, and limitation of autonomous decision scope [208], [210].

Socioeconomic Impact: The level of integration of AI should be evaluated as well as the environment in which such integration will take place. Moreover, the acceptability from the society should be studied in depth given that autonomous systems could possibly introduce a level of discrimination and bias. Another factor that should be considered is the behaviour of employees and employers [208], [211].

Design: AI should be designed and reviewed by interdisciplinary teams. Interdisciplinary teams could evaluate possible impacts in society, ethics and economy. Moreover, indirect impacts to animals or the environment, for example, could be studied and evaluated further. Bias could be significantly reduced as well, by sharing and combining significant knowledge and expertise of different domains [208].

Responsibility: In some applications which can cause harm to humans and can be classified as more sensitive than others, further research will be required in order to identify, who should be responsible in the case of an undesired outcomes [208], [212]. Although the margins are quite thin, and ART AI principles are interlinked, responsibility should not be confused with accountability which considers mostly the justification of the decisions of AI to the interactors of the system.

5.1.2 ART AI for Patient-Centric Home-Based Rehabilitation

As previously identified, healthcare and socioeconomic reasons in the post COVID-19 world [213][214] impose a need for home-based rehabilitation is multifaceted, Thus, there is an increased need for new approaches that combine AmI or smart monitoring as discussed in Chapter 3 and individualisation to support engagement and motivation in home-based rehabilitation. However, the technological solution must also comply to ethical principles and ART.

Existing AI solutions introduce assumptions and bias of the engineer in the decisions taken by the algorithm [215]. This issue is reflected in recent efforts for ethical AI requirements and design practices as discussed in the previous subsection, which highlights a deeper underlying consideration regarding the used values and norms, and their reinforcement through the evolution of the machine's autonomy [216]. Thus, there is a global urgent need for Accountable [217], Responsible [218], and Transparent [219] AI to enable wider use of AI [220] based on generally acceptable value systems [221]. Specifically, the following aspects of ART AI were taken under consideration:

a) Unbiased AI: A significant body of work has contributed to methods for bias-free ML models [222][218][219]. State-of-the-art bias reduction approaches was followed [225] in the proposed rehabilitation support approach based on designing an appropriate and balanced training dataset that would remove bias due to a class under-representation. This is presented in Sections 5.2.1 to 5.2.3.

b) Explainable AI: There has been strong research interest in the field of explainable AI (XAI) in recent years [226]. However, these approaches are mostly focused on the social aspect (user response to XAI) [219], or the implications of manipulating inputs to generate false negative or false positive responses [227]. Explainability will be taken under consideration in terms of user engagement combining the findings of Chapters 1 and 3 and discuss this further in Subsection 5.2.6.

c) Interpretable: Interpretable and justifiable outcomes of AI require transparency of the ML model as well as the model's decisions and behaviours [228][229]. Current EU regulation allows for individuals to enquire about AI decisions [230]. However, regulation is not well defined for the design and development of such models particularly in the medical applications domain [231]. Moreover, accountability and transparency are strongly interlinked with interpretability [224]. The approach to transparency will be discussed in terms of the design (Subsection 5.1.5) and methodology (Subsections 5.2 and 5.2.5) and in Subsection 5.2.6.

Unbiased AI is particularly needed for home-based rehabilitation systems as they are used by a variety of users including elderly and young, male and female, of varied height and weight and all need to receive the same and fair level of support. Explainability and interpretability are also key to enable user engagement as it was identified in Chapters 1 and 3 and strongly link to complexity and motivation. In this chapter, the information that the model can generate will be taken under consideration to provide further insight to the decision-making process, establish a dataset to address the bias issue, and aim for interpretability to address some of the challenges in home-based rehabilitation AI.

We review existing Artificial Ambient Intelligence approaches in Section 5.1.3 and Section 5.1.4 will be reviewed and the state-of-the-art as well as further improvements will be identified. A system to address them will be investigated and presented in Section 5.2. The evaluation of the proposed system along with the results is presented in Section 5.3, followed by the discussion in Chapter 6. Finally, conclusions and the identified future work are presented in Chapter 7.

5.1.3 Review of ART AI in Medical Systems

In medical applications, such as rehabilitation, the requirement for ART AI is heightened. In radiology AI has been already introduced in order to contribute further to diagnosis. High-quality data as well as the responsibility for curating, provenance and avoiding bias are critical. In the medical domain unintended consequences of AI can be significantly damaging or irreversible. The benefit to patient care and society on the other hand is tremendous. Thus, ART solutions become of paramount importance [231].

In [232] it has been stated that AI has a crucial role in disease detection. It is found that deep learning in particular is often difficult to comprehend. Additionally, incorporating medical knowledge in the design of ML solutions is important to generated XAI AI. Also, the outputs of XAI AI should be in human legible terms such as sentences of text, and that embedded optimisations and decisions should be explained.

Additionally, in [233] the power of prediction of ML models in comparison with statistical models has been identified. However, given that the explainability and interpretability of these models due to 'black box' conditions and complexity are limited, ML is an approach that cannot entirely be trusted by the clinicians. In order to evaluate and test the interpretability of a complex system which utilises Neural Networks, a reinforcement learning decision support system (DSS) was developed around the neural network model which would be able to learn and understand what is interpretable to different users. The developed DSS suggests that feedback in different forms must be provided and that the ART AI adapts not only the model to the user but also the form of the provided output. Thus, individualisation and layered/selectable feedback is necessary for the non-technical user of AI. However, these systems refer to the medical professional and are not targeting the home-based rehabilitation environment or the end user, which is the focus of our work.

According to [229], the target user audience is a key consideration for the XAI design. Additionally, the authors discuss the explainability of well-known models. The identified taxonomy will be used during the analysis of suitability of various models in this chapter with a focus on the end user of the home-based rehabilitation system

5.1.4 ART AI in Home-Based Rehabilitation Systems

It has been mentioned in Section 4.2.1 that AmI is a promising solution for home-based rehabilitation for a variety of reasons. Based on [234] AmI are systems are:

a) embedded to the user's environment,
b) able to recognise the user and the environment
c) able to provide individualisation
d) able to provide a good level of adaptiveness according to users' needs
e) able to be aware and anticipate a variety of users' behaviours

In Figure 29 the desired functional and non-functional properties of systems and technologies characterised as AmI are presented. According to the findings in [234] AmIs provide the appropriate framework, the right architecture and all the requirements needed in order to provide a high level of explainability as well as to operate as mediators between humans and other systems.

Furthermore, they have the opportunity to provide dialogic explanations of the system's behaviour and to observe the user's explanation needs. As it was mentioned above, due to the fact that AmI systems are by nature embedded into the environment, they are context aware, they can be individualised, and most importantly, they can change in response to the user and can anticipate the user's desires. This relates to goal setting, motivation enhancement, and engagement which are primary factors affecting home-based rehabilitation [56].

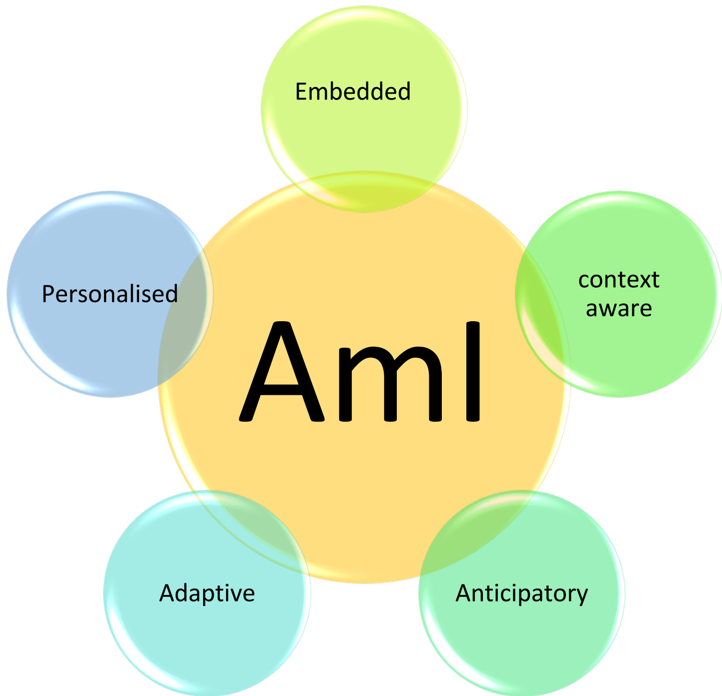


Figure 29 AmI characterisation based on[234]

It has been found in [235] that operation and philosophy of AmI is based on interaction of heterogeneous elements. It is an integration of hardware and software in order to provide further support to humans. Moreover, as it was mentioned, AmI systems can provide individualisation and hence more sophisticate and complex algorithms would need to be used. These can increase computational capacity requirements that in AmI – and Internet of Things (IoT) systems in general – is always a significant constraint [236]. Hence, our proposed home-based rehabilitation system, which is based on AmI, must also be capable of addressing the challenges of reduced computational cost and physical constrains of the utilised devices.

Recent studies have improved computational efficiency of advanced AI methods and proved applicability in AmI scenarios. For example, random decision forests in general have proven to be a very effective AI approach for medical applications at reasonable complexity [237][238]. Recently, ensemble methods but also hybrid methods that combine heuristic algorithms with heterogeneous ML models have proven to improve performance with various effects on computational footprint [239]

At slightly higher complexity but lower computation capacity requirement, in [240], eXtreme Gradient Boosting (XGboost) is used for sensor calibration as part of an ensemble method applied on an AmI system. The benefits of using this particular algorithm has been pointed out, given that it has the ability of scaling up to large volumes of data with low computational resources. XGboost is a method which develops a strong learner through an advanced ensemble learning approach that combines several weak learners through an iterative process. In this case the weak learners are individual smaller decision trees. The strong learner combines them sequentially. The goal of the algorithm is to correct the residuals of predictions of previous trees in the sequence [241]. Moreover, XGBoost offers a high level of hyperparameter tuning that makes it versatile and highly performant on a variety of applications. However, the solution presented in [240] is restricted to a single feature data stream learning problem.

Deep learning approaches are in general considered complex as they require large computational resources and large datasets, and hence often unsuitable for light home-based rehabilitation applications. However, Keras API, one of the most popular platforms to easily implement deep learning models, offers an efficient deployment of neural networks (NNs) for complex multi-dimensional problems [242][243], making NNs suitable for IoT devices.

All of the approaches mentioned above are promising in terms of accuracy and complexity and will be investigated further in Section 5.2 for suitability in our specific problem domain under the specified ART AI requirements previously discussed in Section 5.1.2.

In [244] an approach for increased security on IoT has been suggested. Due to potential growth of IoT systems which will be connected to the web, there will be an increased need for high level of security. Hence, in order to address this issue, the authors have developed an Infusion Detection System (IDS) which is able to detect anomalies at the network level. XGBoost has been used in order to create a model which is able to analyse network frames. The approach utilises ensemble learning and edge computing, in real time, demonstrating the very low computational footprint of gradient boosting algorithms.

In [241] an accurate method for predicting toluene, ethylbenzene and xylene (TEX) concentrations and corresponding enrichment factors has been developed. It has been found that although ML approaches are widely used on this particular area, due to complexity, they lack in explainabilty and interpretability. Hence, the proposed ML approach is based on regression analysis by means of XGBoost in order to estimate relationships between different parameters and improve interpretability. In [241] it is reported that XGBoost is suitable for a variety of applications and performs accurately. Most importantly, in IoT applications considered in [241] it performs better than Support Vector Machines, Random Forests, and Deep Learning NNs [241]. Note that in [241], the dataset was divided in 80/20 percent, for training and testing datasets, respectively. Hyperparameter tuning was implemented using brute-force grid search and 10 fold cross-validation to prevent overfitting the dataset. The final model was tuned to the values that reported the best accuracy. In terms of ART AI, based on [241][229] boosted trees can provide outputs to support explainability and interpretability in an IoT context. Further, boosted trees improve the bias variance trade off [245][246]. However, [229] suggests that ensembles of trees require further simplification to relay information, and make them the most appropriate approach for further investigation to improve the ART component of ensemble learning approaches.

According to [247] in order to apply successful ensemble learning two main criteria have to be taken under consideration a) accuracy and b) diversity. Ensemble learning methods were analysed in [247]. Bagging is when similar weak learners learn in parallel through an independent process with each other. Combining then the weak learners happens through deterministic averaging process. Boosting is when similar weak learners are ensembled following a sequential learning process and a combination through a deterministic approach. Stacking is when non-similar, weak learners, learn through a parallel process and they are combined together in order to produce an overall model based on weak learners' predictions. For the development of a spatiotemporal fusion method in [247], a stacking model approach has been selected. Data are inputted to three different algorithms: Back Propagation Neural Network (BPNN), K Nearest Neighbours (KNN) and XGBoost. The outputs of these are fed to a stacking model which will output the final value. However, this approach has a significant training overhead that would make individualisation impossible in an AmI scenario due to the computational capacity restrictions. On the other hand, as has been found on the above research, further improvements can be achieved by combining boosting (XGBoost) and stacking with k-nearest neighbour (KNN) on a server while at the same time improving the robustness and reducing overfitting issues.

Overfitting and underfitting are directly related to the model's performance. Simplicity of the model which utilises a small number of features, introduce high *bias*, underfits the data and decreases the flexibility of learning from the dataset. On the other hand, increased complexity leads to data overfitting, which means that the model fits the training data extremely well, but there is limited knowledge on how the model will fit the test data otherwise known as new/unseen data. In this case there is high level of *variance*. Both, high Bias or high Variance could increase significantly the error of prediction [248].

There are several ways for avoiding overfitting issues such as: (a) cross-validation [249][250], (b) training the algorithm with a bigger amount of data, (c) improving generalizability through removal of unimportant features [245], (d) regularization [251] (e) ensemble learning. Underfitting can be reduced through various approaches such as: (a) increasing the number of features in the model where this could include the extraction of features from existing features (b) increasing model complexity c) increasing the training time [252]. XGBoost algorithm decreases both variance and bias thus addressing overfitting considerations but at the cost of sacrificing explainability. This results in a need for post-hoc analysis.

Post-hoc analysis is part of explainability and transparency as discussed in Section 5.1.1. Post-hoc explainability refers to models that are not explainable or interpretable by design and in order to increase their level of interpretability they utilise different means for explanation such as visual, or local, explanations which can be made by example and/or by simplification as well as feature relevance explanations [229]. Based on [229] Linear/Logistic regression, Decision trees, KNN, Rule Based Learners, General Additive Models and Bayesian Models do not need further post-hoc analysis and hence are easily interpretable. XGboost on the other hand requires limited post-hoc analysis which can be based on feature relevance explanations and probability-based explanations.

In summary, existing approaches that ensure high accuracy, such as deep learning, do not meet the ART AI requirements and do not maintain low computation footprint suitable for individualised incremental learning on AmI systems to support home-based rehabilitation. On the other hand, XGBoost can lead to high accuracy and low computation footprint while providing some level of transparency. Hybrid methods can further improve accuracy and have a variable impact on computation footprint depending on implementation.

Motivated by the low computational cost of boosting and the potentially limited post-hoc analysis required, in this chapter, ART for AmI in the home-rehabilitation systems was investigated, a low computational footprint boosting and stacking ensemble learning method to deliver intelligent and individualised ART AI to meet the criteria presented in Chapter 3 was proposed, where, following a rigorous review of technologies for home-based rehabilitation, the requirements for engagement and motivation enhancing technologies for home-based rehabilitation were identified. In particular, XGBoost and KNN algorithms were combined without the added overhead of traditional stacking on a computational power restricted setup. Both models are categorised as transparent with little simplification and post-hoc analysis requirements for explainability [229]

5.1.5 ART Design Considerations

To design a system in accordance with the ART AI principles, design time considerations in this section were reviewed. Responsible AI ethics can be categorised in three groups [253]:

- **Ethics by Design:** technical/algorithmic amalgamation of moral thinking abilities as a major aspect of conduct of artificial self-governing systems.
- **Ethics in Design:** regulatory and engineering strategies that contribute to the examination and assessment of the moral implications of AI systems as these incorporate or supplant conventional social structures

- **Ethics for Design**: code of conduct and standards and certification processes that guarantee the honesty of responsible developers and system users as they research, plan, develop, utilise and manage AI systems.

In every case the design time considerations for the development of ethical AI models is highlighted. These categories include the ethical reasoning capabilities integrated in the model's behaviour, the analysis and evaluation of ethical implications in social structures and ensuring integrity of the developers and users. Design time considerations have recently been published by European Research and Innovation that apply to safety-critical systems with AI [208]:

1) avoid bias and prejudice in training data, or make biases clear to user population
2) ethical principles embedded into AI development.
3) interdisciplinary teams are crucial.
4) transparent data provenance (input, output).
5) lay people need to understand AI decisions.
6) decision justification.

Design time considerations were followed in the development of the proposed methodology and the design of the proposed hybrid ML approach for individualised ART-driven rehabilitation. We discuss our ART AI Design approach was discussed and the 6 design time considerations were addressed for our proposed system in Section 5.2.6.

5.2 Methodology

In combining ART principles and ML the approach presented in Chapter 4 is further enhanced and discuss a patient-centric individualised, home-based rehabilitation support system based on responsible and interpretable AI. In this section, first the experimental data set used was described, and then provide a methodology for generating a synthetic dataset to ensure unbiased decision making following the ART AI principles and design considerations reviewed in the previous section. Then, well-known ML methods conforming with our ART AI requirements discussed in Section

5.1.2 were evaluated on both datasets. Finally, a hybrid learning approach was proposed to mitigate limitations of the existing solutions in Section 5.2.5.

5.2.1 Training Dataset Design

To design a responsible ART AI driven home-based rehabilitation system an appropriate dataset needs to be generated, collected and prepared. A dataset was required which can be used both to monitor progress of activities relevant to rehabilitation and also to diagnose both individual difficulties and medical conditions or comorbidities. Advantage was taken on the small experimental dataset presented in Chapter 4 and published in [213], where a sensor-based platform was introduced for data collection and analysis of two medical tests used in subject evaluation, namely the TUG and FTSTS tests. A thorough review of patient evaluation medical tests and their relation to home-based rehabilitation is presented in Chapter 4.

The two tests were selected as they are relevant to a variety of activities of daily living, can be performed without medical supervision and in the home environment have been used to evaluate the progress and condition of post-stroke patients in line with the criteria presented in Chapter 3. Both tests evaluate lower limb strength, mobility, static and dynamic balance, functionality, and durability, all of which are relevant to rehabilitation outcomes. Improvement in performing the tests over time translates to better ability to perform daily tasks (such as sitting and standing, walking small distances), self-efficacy and engagement with rehabilitation as discussed in Chapter 3. Additionally, the use of medical tests provides the benefit of evaluating the system against medical standards and clinical specification.

However, to achieve high accuracy, large datasets representative of real patient measurements is needed, which, to the best of our knowledge, are not publicly available for the above-mentioned problem.

A poorly designed dataset used for training the models can easily lead to biased or inaccurate outcomes [220]. According to [254] a dimension is a measurable property of data quality in which some aspects of the data (e.g. precision, consistency) are reflected. Which in turn can be used to direct the quality understanding process. Hence, data can be defined as according to one or more dimensions of being high quality. In [254][255] hundreds of dimensions of dataset quality have been collected and further reduced to 15 summative dimensions which have been categorized into 4 categories. The 4 categories are the following (Table 15): (a) *intrinsic category*: dimensions indicative of the natural quality of data; (b) *contextual category*: quality of data must be regarded in a particular context; (c) *representational category:* dimensions linked with format and interpretation; (d) *accessibility category:* dimensions connected accessibility to users.

Table 15 Data Quality dimensions[254] and our approach to addressing them.

Category	***Dimensions***	***Description***	***Addressed***
Intrinsic	Accuracy	Data is correct (error-free) and reliable	No errors, outlier removal
	Believability	Degree to which data is seen as credible and true	Medical journal publications
	Objectivity	How impartial the data is	Bias reduction
	Reputation	Data contents or source are kept in high consideration	Medical journal publications & interdisciplinary team experiment design according to ethics application
Contextual	Appropriate amount	How suitable the quantity of the data	Generated synthetic data
	Completeness	Refers to the scope of the information in the data	All possible difficulties; All conditions for which results are published
	Relevancy	How usable applicable, or interesting the data is	Only publications with full feature set reported
	Value-added	Data provides a competitive advantage	Thesis contribution
	Timeliness	The age of the data	Wide range of dates in published journals
Representational	Concise representation	Data is compactly represented	No missing data, no unnecessary features
	Ease of understanding	How clear readable or understandable the data is	Simple understandable features
	Interpretability	The extent to which the data meaning is explained	All familiar measures
	Consistency	Data continuously presented in the same format	Upheld by experiment & synthetic data
Accessibility	Access Security	Access is secure or can be restricted	n/a
	Accessibility	The degree to which the data is retrievable	n/a

The dimensions for each of these categories are presented in Table 15 where the last column showcases how they were addressed in our dataset design approach. Initially goal was addressed by mitigating the intrinsic, contextual, and representational data quality dimensions as presented in [254] and Table 15 which should be considered when designing a training dataset. By addressing these dimensions, the quality of our produced dataset was ensured. An extensive discussion of the method used to produce a high quality the dataset is presented in the following Subsections and refers to these dimensions.

Specifically, to generate a synthetic dataset and address the appropriate amount and quality of data, was considered carefully the following aspects: (a) Accuracy; (b) Believability; (c) Objectivity; and (d) Reputation as presented in Table 15, through developing a method that generates data from published medical research outputs.

To generate a new synthetic dataset that will lead to high accuracy and objectivity, the experimental data from Chapter 4 was used as a starting point, and was augmented using statistical published data, collectively presented as a list of publications in [201] for TUG and [201] for FTSTS tests. further elaboration on the datasets is presented next.

5.2.2 Experimental Dataset

In Chapter 4, data on completion of TUG and FTSTS tests were collected through the presented low-cost home-rehabilitation system with 8 participants. The participants simulated difficulties in the various stages of the TUG test simulating elderly individuals performing the same test. Each *difficulty level was considered* as a *label*. *The features* recorded for each participant are:

1) Test completion time (seconds)

2) Age (years)

3) Height (meters)

4) Weight (kg)

5) BMI (kg/m2calculated using Equation 9)

6) Gender (male/female)

$$\boldsymbol{BMI} = \frac{Weight\ (Kg)}{[Height(m)]^2} \quad (9)$$

The use of *BMI*, in addition to *Height* and *Weight*, could improve the prediction accuracy. According to [256], the inclusion of features which are generated from other features (using linear or polynomial equations) can improve the performance of more complex models such as Neural Networks (NN). The influence of features such as *BMI will be investigated based* on accuracy in Section 5.3.

This dataset had no missing, malicious, erroneous, inconsistent, or irrelevant data points. Data formatting was necessary to map gender and category data entries from `String` to `Integer` values. Some outliers were present in the originally recorded data as evident in [213]. These outliers were removed following the > 3σ (σ denotes standard deviation) approach [257]. Finally, all measurements were normalised in the range $(-1,1)$.

The *categories* (interchangeably referred to *classes* or *labels*) which were recorded for TUG were: (a) Difficulty to Walk, (b) Difficulty to Turn, (c) Difficulty to Stand/sit, (d) Normal, (e) Fast. For FTSTS, the classes are: (a) difficulty to stand/sit (annotated as Slow) and (b) Fast.

There are two major issues with this experimental dataset. First, it is small since only 8 participants are included, which will lead to inaccurate outcomes. Second, the dataset is very unbalanced, potentially leading to bias and affecting objectivity requirement. Indeed, firstly, the class Normal in TUG (and Fast in FTSTS) has fewer collected data points. Class unbalance is addressed with the use of the synthetic minority over sampling technique (SMOTE) [258]. With the use of SMOTE all classes were balanced with each class having 40 datapoints both in TUG and FTSTS. In total 17 samples were generated by SMOTE for TUG of which one was for the Fast and 16 for the Normal class. Moreover, 7 were generated for FTSTS for the Difficulty class.

Besides class unbalance, the dataset suffers from data unbalance as the resulting dataset had 87% male participant entries and all participants belong to 20-40 years of age group; clearly leading to biased outcomes. To identify the effect of this bias the feature importance for the proposed method in Section 5.2.5.2 will be discussed while it is mitigated through the introduction of the synthetic dataset presented in the following Subsection.

5.2.3 Synthetic Dataset

To address requirements for dataset objectivity, accuracy and believability, and facilitate transferability to a range of conditions, such as Parkinsons and Dementia, a dataset was further synthesised based on published statistical results for TUG and FTSTS tests. Transferability and diagnosis are two desired criteria for home-based rehabilitation systems in Chapter 3, the underlying reason being the development of co-morbidity for stroke survivors and early diagnosis/warning to carers.

ML models have demonstrated success in many application fields. However, since advanced ML models are data driven, to translate these successes to medical fields, it is necessary to provide large datasets to develop and train the models. Generating an open access dataset based on collected data is a significant obstacle due patient data privacy and cost [259]. Hence, the rehabilitation field suffers from a lack of appropriate datasets to develop and test advanced ML models. In many other fields, where real data collection is expensive or impractical, synthetic datasets are generated and employed. However, the medical domain has been very reluctant to embrace synthetic datasets. Most recently, the community has recognised this adverse effect [259][260]. In [260], [261] the main arguments against synthetic datasets are rebutted through a comparative analysis of actual and synthetic datasets used to train and test the model. It is demonstrated that the synthetic dataset can produce equal performance if synthesised from statistical representative of the actual cohort of patients. In [259] it is discussed that if the model performs accurately when tested with unseen real patient data then the synthetic dataset can be accepted as representative of the reality. In this chapter this line of work was followed, a dataset ensuring statistical agreement with real measurements and test the trained model using experimental dataset in Section 5.3 was synthesised.

The reasons for synthesising a dataset are:

1. To meet the requirements of Chapter 3 for transferability and co-morbidity diagnosis, the underlying reason being the development of co-morbidity for stroke survivors and early diagnosis/warning to carers.
2. To improve variance and bias of our experimental dataset.
3. To improve accuracy of our model.

4. To include a wider range of height, weight, age and BMI that is representative of the wider population that is discharged to home-based rehabilitation directly impacting bias.
5. To improve believability by sourcing information from medical journals, where larger cohorts of geriatric subjects and patients have participated in experiments.
6. To improve objectivity by including information from experiments with participants diagnosed to have the medical conditions that can be developed as co-morbidities, such as Parkinson's and Dementia.

It is important to highlight here that the synthetic dataset will be used for initial training of the model. However, as one of the criteria in Chapter 3 is to provide individualised solutions, our model would be constantly re-trained and corrected using real data acquired by the user of the device presented in [213]. Thus, throughout the system lifetime the synthetic data will eventually be proportionally a small contributor to the model. Furthermore, the synthetic dataset could not be generated from the experimental dataset using an oversampling method.

SMOTE could not be used in this case as: (a) experimental dataset is gender and age unbalanced (b) SMOTE requires an input dataset and the original dataset, described in the previous Subsection, does not link the 6 features with specific conditions. Thus, an alternative approach for data synthesis is proposed next.

A database of medical/clinical research publications that present statistical results of experiments with patient cohorts for TUG and FTSTS was consulted. Each publication states the medical condition of the cohort, the number of subjects and the statistical characteristics of the cohort. To the best of our knowledge this database presents a comprehensive review of all medical conditions for which TUG and FTSTS are used to evaluate patients.

Every condition reviewed in [201] and [262] was included for TUG and FTSTS, respectively. Publications cited in [201][262] were included as inputs in the synthetic dataset algorithm, if they presented statistical descriptors for all of the features used in the Experimental dataset above, or if the features could be extrapolated or calculated (e.g., BMI). Publications were in turn excluded if any of the features was not reported or could not be extrapolated from the information presented. Because the selection is not made using a specific feature as the criterion, this selection method does not introduce a particular bias to the pool of included publications. Additionally, publications were not excluded based on race, sex, or ethnicity of the participants. In addition to the difficulty classes mentioned in the experiment dataset for TUG, the following condition classes were finally considered: *Healthy, Geriatric, Parkinsons, Parkinsons non fallers - medication, Parkinsons non fallers – no medication, Parkinsons fallers, Dementia mild/moderate, Dementia severe, Arthritis improvement, Arthritis knee arthroplasty, Arthritis, Stroke, Brain Injury, Bilateral vestibular hypofunction, Unilateral vestibular hypofunction, Spinal injury, Paraplegia, Tetraplegia.* In total 16 publications remained for TUG based on the inclusion criteria, namely [263][264][265][266][267][268][269][270][271][272][273][274][275][276] [277][278][279]. A total cohort of $n = 937$ subjects is represented by these publications with age in the range [5, 112] years, height in the range [0.81, 2.20) m, and weight in range (30, 136) kg. Each condition class has 280 datapoints and the overall set has a total of 5040 datapoints.

Similarly, for FTSTS the condition classes are: *Healthy, Geriatric, Geriatric fallers, Parkinsons stage 1, Parkinsons stage 2, Parkinsons stage 2.5, Parkinsons stage 3, Parkinsonsstage 4, Parkinsons, Arthritis, Arthritis knee arthroplasty, Stroke, Vestibular disorder.* In total 12 publications remained for FTSTS based on the inclusion criteria, namely [266][280][281][282][283][284][285][164][286][287][288] [289]; while 3 were excluded. A total cohort of $n = 2381$ subjects is represented by these publications with age range [11, 93] years, height in the range [0.94, 2.35) m, and weight in range (22, 120) kg. Each class has 240 datapoints and the overall set has a total of 3120 datapoints. In summary, Table 16 presents the final datasets, each with its own features and labels presented for clarity.

According to [283], BMI has low correlation to the FTSTS test completion time (p-value= 0.4) so higher variability was introduced compared to the TUG synthetic dataset as reflected in Figure *30*. The same is true for height [285], [289].

On the contrary, hand positioning has a significant correlation to completion time. But all participants in the experiments and included publications followed the same hand positioning; hands crossed over chest. So, this parameter was ignored [290] given that it would be the same value for all datapoints and hence provided no difference in completion time. Algorithm 1 is proposed to generate synthetic data points from the statistical properties presented in considered publications, where μ and σ denotes class mean and standard deviation, respectively.

Table 16 Summary of datasets with their respective features and labels

Dataset	Features	Labels
TUG EXP	*Test completion time*	*Difficulty to walk*
	Age	*Difficulty to stand*
	Height	*Difficulty to Turn*
	Weight	*Normal*
	BMI	*Fast*
	Sex	
TUG SYNTH	*Test completion time* *Age* *Height* *Weight* *BMI* *Sex*	*Healthy, Geriatric, Parkinsons, Parkinsons non fallers - medication, Parkinsons non fallers – no medication, Parkinsons fallers, Dementia mild/moderate, Dementia severe, Arthritis improvement, Arthritis knee arthroplasty, Arthritis, Stroke, Brain Injury, Bilateral vestibular hypofunction, Unilateral vestibular hypofunction, Spinal injury, Paraplegia, Tetraplegia*
FTSTS EXP	*Test completion time*	*Difficulty to stand/sit*
	Age	*Fast*
	Height	
	Weight	
	BMI	
	Sex	
FTSTS SYNTH	*Test completion time* *Age* *Height* *Weight* *BMI* *Sex*	*Healthy, Geriatric, Geriatric fallers, Parkinsons stage 1, Parkinsons stage 2, Parkinsons stage 2.5, Parkinsons stage 3, Parkinsonsstage 4, Parkinsons, Arthritis, Arthritis knee arthroplasty, Stroke, Vestibular disorder*

The algorithm is based on the following observations:

- There is a linear relationship between time of completion and age in both TUG and FTSTS tests [278][281] but there are several exceptions to this rule so variability is required to represent a more realistic relationship.
- In [291] an almost linear relationship is presented between age and BMI for adults over 40 years of age. This is further supported by the widely accepted BMI charts [72]. But it is not absolute, so variability is again required.
- BMI, by definition, has a linear relationship to Weight (Equation 9)
- According to [268] female subjects perform faster than male subjects in TUG which is a superset of activities in relation to FTSTS.
- Sex and Weight also have a relationship that is linear in regard to the mean weight value of the population with higher variance (wide standard deviation causing overlap between the two populations) [292].
- Height is calculated using Equation 9 after the pair {BMI, Weight} has been established.
- Sex and Age have no correlation (observation supported by our experimental dataset and the mean age value reported in each of the included publications).

Algorithm 1 Generate features for *n* datapoints of Class *x*

Require: $n,\ X,\ \mu_{\text{time}},\ \sigma_{\text{time}},\ \mu_{\text{age}},\ \sigma_{\text{age}},\ \mu_{\text{bmi}},\ \sigma_{\text{bmi}}$

Require: $\mu_{\text{weight}},\ \sigma_{\text{weight}},\ \mu_{\text{height}},\ \sigma_{\text{height}},\ \%_{\text{female}}$

$time \leftarrow sort(X \sim N(\mu_{time},\ \sigma^2_{time}))$

$age \leftarrow sort(Y \sim N(\mu_{age},\ \sigma^2_{age}))$

$age \leftarrow insertVariability(Y)$

$bmi \leftarrow sort(Z \sim N(\mu_{bmi},\ \sigma^2_{bmi}))$

$bmi \leftarrow InsertVariability(Z)$

$weight \leftarrow sort(W \sim N(\mu_{weight},\ \sigma^2_{weight}))$

$gender \leftarrow S_i = 0\ \forall i\ \in \{0, n * \%_{female}\} \wedge S_i = 1\ \forall i\ \in \{n * \%_{female} + 1, n\}$

$\{bmi, weight, gender\} \leftarrow insertVariability(\{Z, W, S\})$

$height \leftarrow H = \sqrt{\frac{W}{Z}}$

Ensure: $\mu_{\text{height}} = \mu H$ **and** $\sigma_{\text{height}} = \sigma_H$

These observations are further supported by the Experimental dataset observations both in terms of linearity and variability.To introduce linearity, e.g., to ensure a linear relationship between time of completion and age, pseudorandom numbers from the

Gaussian distribution were generated for each feature and generate a sorted sequence (function sort in Algorithm 1).

Since lists of all features are sorted from lower to higher, the linear relationship is generated, as for example the lowest completion time will be aligned to the lowest age. Then to introduce variability the order of several data points was swapped (function `insertVariability` in Algorithm 1) in one of the two feature columns, for example, keeping completion time sorted and swapping values in the age column. If variability is required between one feature and a set of other features, then the same swaps are applied to the full set. For example, if data point 3 would be swapped with data point 7 in BMI to introduce further variability compared to age, the same swap would be applied to Height and Sex as well, thus maintaining the correlation between the BMI, Weight and Gender sequences. Since the Height is generated using the BMI Equation 9, in order to ensure that the generated heights correspond to the reported height mean and standard deviation of the publication. Thus, the final step of the algorithm validates this through comparing the generated list mean and standard deviation to the parameter retrieved from the literature.Finally, to verify the validity of the generated datasets for FTSTS and TUG, the correlation matrix for comparison of the features of the synthetic datasets to the original recorded experimental datasets was used. The correlation matrix was used as well, as a guide for the amount of variability to be introduced. The final correlation matrices are presented in Figure 30. It is evident that a similar correlation matrix emerges in both the experimental and the synthetic datasets. However, the gender bias is the most evident bias in the case of the experiment.

This bias is demonstrated through the apparent high correlation of sex to height and weight. As the experiment recruited only 1 female subject this is far from realistic. The synthetic dataset, on the other hand, demonstrates a weaker relationship between sex and height as well as sex and weight which is closer to reality and thus compensating for the bias in the experiment dataset. Figure 31 presents both the datasets after projecting all the features to 2 dimensions using PCA. Visualising the data also demonstrated a potential polynomial function as displayed in Figure 31a, Figure 31c where all the points seam to follow a polynomial curve with different offset for each of the classes.

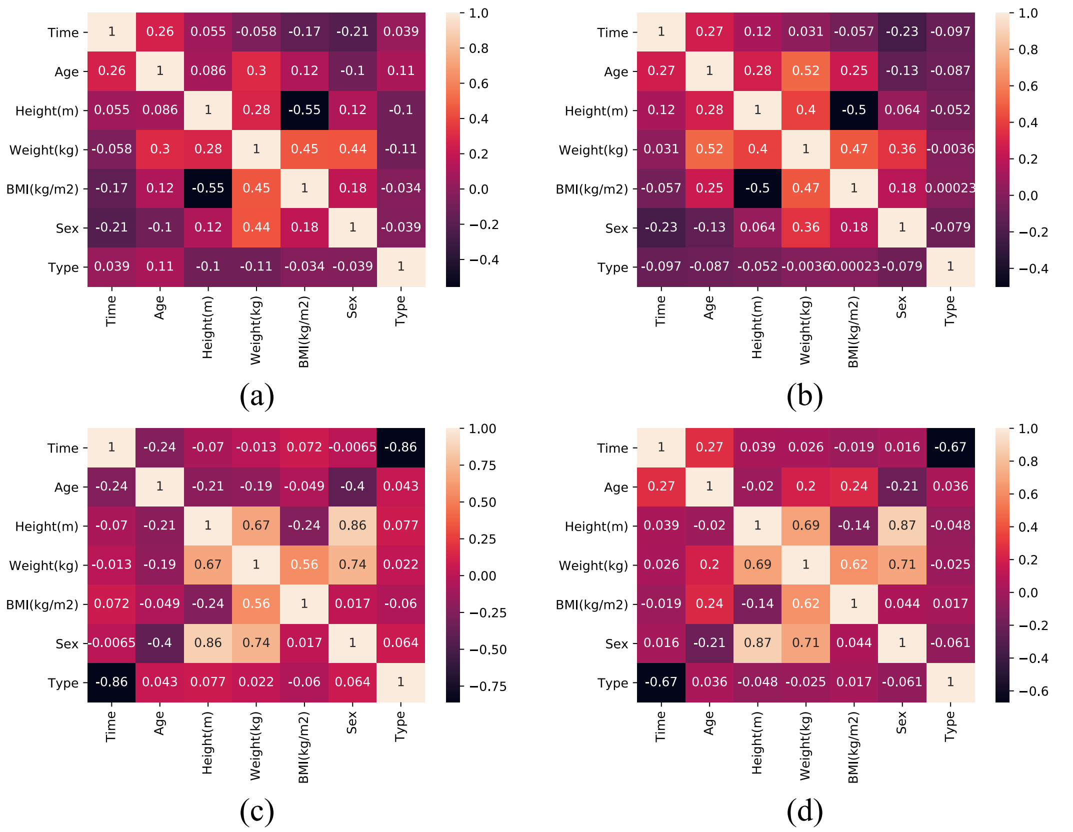


Figure 30. (a) Correlation matrices for TUG experiment, (b) TUG synthetic data, (c), FTSTS experiment and (d) FTSTS synthetic data

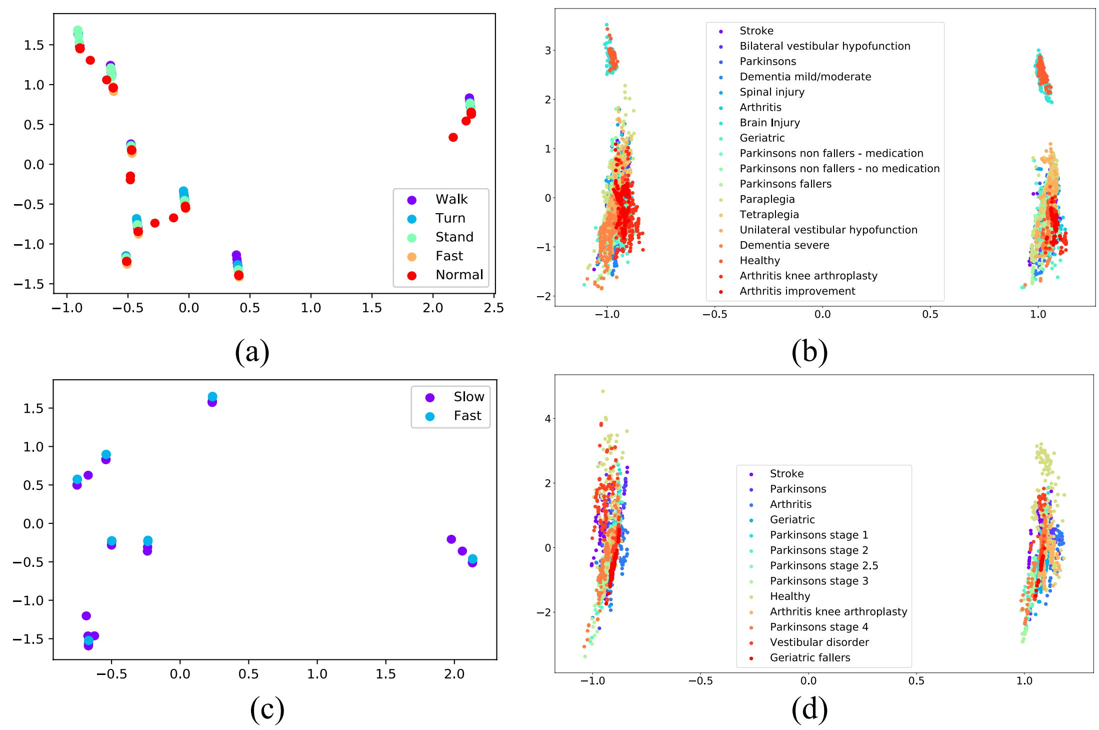


Figure 31 (a) TUG exp. data after PCA analysis reducing the features to 2, (b) TUG synth, (c) FTSTS exp., and (d) FTSTS synth. respectively.

5.2.4 Evaluation of ML Methods for Supervised Classification

Several ML models exist with widespread use in a variety of medical applications. Here, the most relevant to the home-based rehabilitation problem will be reviewed. Then an implementation of those and a comparison based on their accuracy will be carried out in order to identify the most promising models for further investigation and development. As the data were labelled both for the experiment and the synthetic datasets, supervised learning methods were selected. The design requirements-imposed restrictions: Firstly, through the computational capabilities/capacity of the ambient intelligence system; secondly, through the available programming language to collaborate with the sensors and components of the system. As a result, pandas and sklearn libraries within Python platform were selected. In the following subsections the available algorithms will be examined.

5.2.4.1 Support Vector Machines

SVMs[293] are a supervised learning algorithm, and they can be utilised in order to address regression (SVR) as well as classification (SVC) problems. However, SVMs perform well with smaller datasets and given that the computation time will be less in comparison with big data sets. SVMs utilise Kernel Functions in order to identify the appropriate Support Vector Classifiers in higher dimensions. There are different types of Kernel functions and most popular have been illustrated on Figure 32.

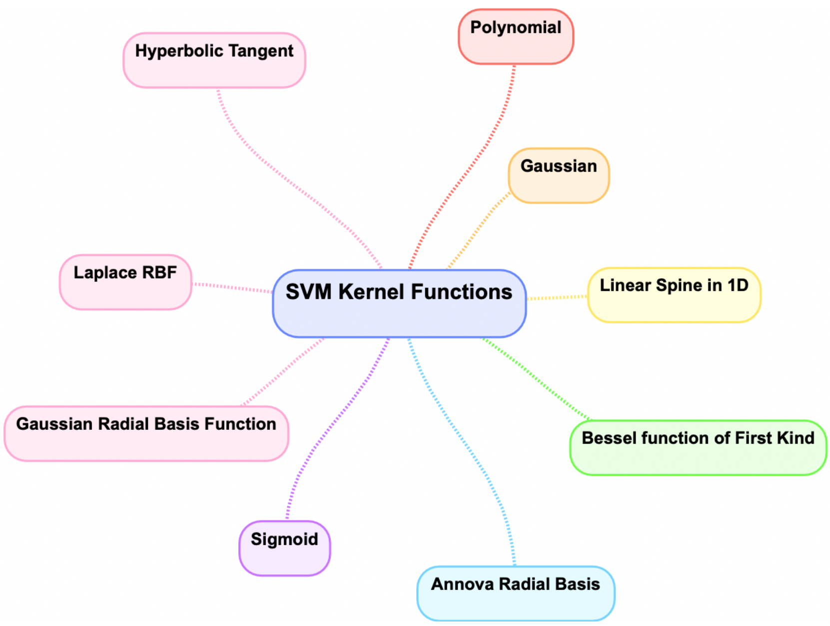


Figure 32 Different SVM Kernels where RBF is Radial Basis Function.

Using these mathematical functions, kernels are able to take data as an input and transform them to the analogous form. However, kernel functions calculate the relationship of every set of two data points as if they are located in the higher dimension. They do not perform the transformation. This is called the Kernel trick. The Kernel trick is responsible of reducing the requirements of computational capacity on SVM's by avoiding the mathematical calculations which is the tool in order to transform the data from low to high dimensions.

Computational requirements are directly linked to the selected kernel [294] [295][296][297]. Overall, SVMs are powerful yet light weight in terms of computational requirements (subject to kernel selection).

All possible kernels used by the Python library were tested with the SVM in our implementation including: linear, polynomial of degree 4, radial basis function (RBF), and sigmoid. In all cases the inverse of strength regularization, which is one of the parameters, denotated with letter C was set equal to 1. There is a relation between parameters and behaviour of the model. It was noticed experimentally that, as the C increases the model overfits and model underfits when the value of C decreases.

5.2.4.2 XGBoost

As it was aforementioned XGboost is a method which develops a strong learner through an advanced ensemble learning approach that combines several weak learners through an iterative process. In this case the weak learners are individual smaller decision trees. The strong learner combines them sequentially. The goal of the algorithm is to correct the residuals of predictions of previous trees in the sequence. XGBoost algorithm can be utilised for regression and classification.

The algorithm uses regularization parameter λ to reduce the similarity scores given that λ is ın the denominator of the fraction. Hence, lower similarity results in lower gain which in return will mean that a lower value is needed for γ to prune more branches. Thus, λ has a great impact on tree's sensitivity. At the same time λ reduces the contribution/sensitivity of a single observation to the new prediction.

XGBoost can mitigate variance and bias in datasets while automatically correcting model biases through the random forest underlying approach. This also results in reduced risk of overfitting. [298]

XGBoost concentrates a high level of many advantages in comparison with other ML algorithms. For example, it allows parallel processing, it is highly flexible, it can handle missing values, it can be efficient on tree pruning, there is a built-in Cross-validation and it can built up on existing models [299][300][301][302].

Parameter and hyper-parameter tuning are quite important because they influence the behaviour of the training algorithm directly and have a major impact on the model's output. In Figure 33 the XGBoost parameters are presented. Of those, the parameters that were tuned in our proposed method have been highlighted. Those that are not highlighted are superseded by those used or present alternative possible settings. In any case our approach completely covered any possible optimisation of the XGBoost model.

The Grid search hyperparameter optimization technique has been utilized. The process started by defining the objective as multisoftmax because of our multiclassification problem definition presented in Subsections 5.2.1 and 5.2.2. Also `num_class` is defined as the number of classes in each of our classification problems. For all the following steps the `seed = 27` was used to maintain comparability between grid search cross validation results.

Leaving all other values at default the effect of `booster` changing between the two possible values is evaluated. It was found that `gb_linear` reduced accuracy so the `gb_tree` default setting is maintained. With the above set up, the `learning_rate` using grid search is evaluated between the values 0.05 and 0.3; the optimal value was evaluated to be 0.1 for all cases at this point. Then, the value of `n_estimators` (i.e. maximum number of trees) is evaluated at a range between 0-5000. Optimal values for `n_estimators` was 11 for the TUG experiment dataset, 238 for TUG conditions dataset, 1 for FSTS experiment dataset and 318 for FSTFTS conditions at this point.

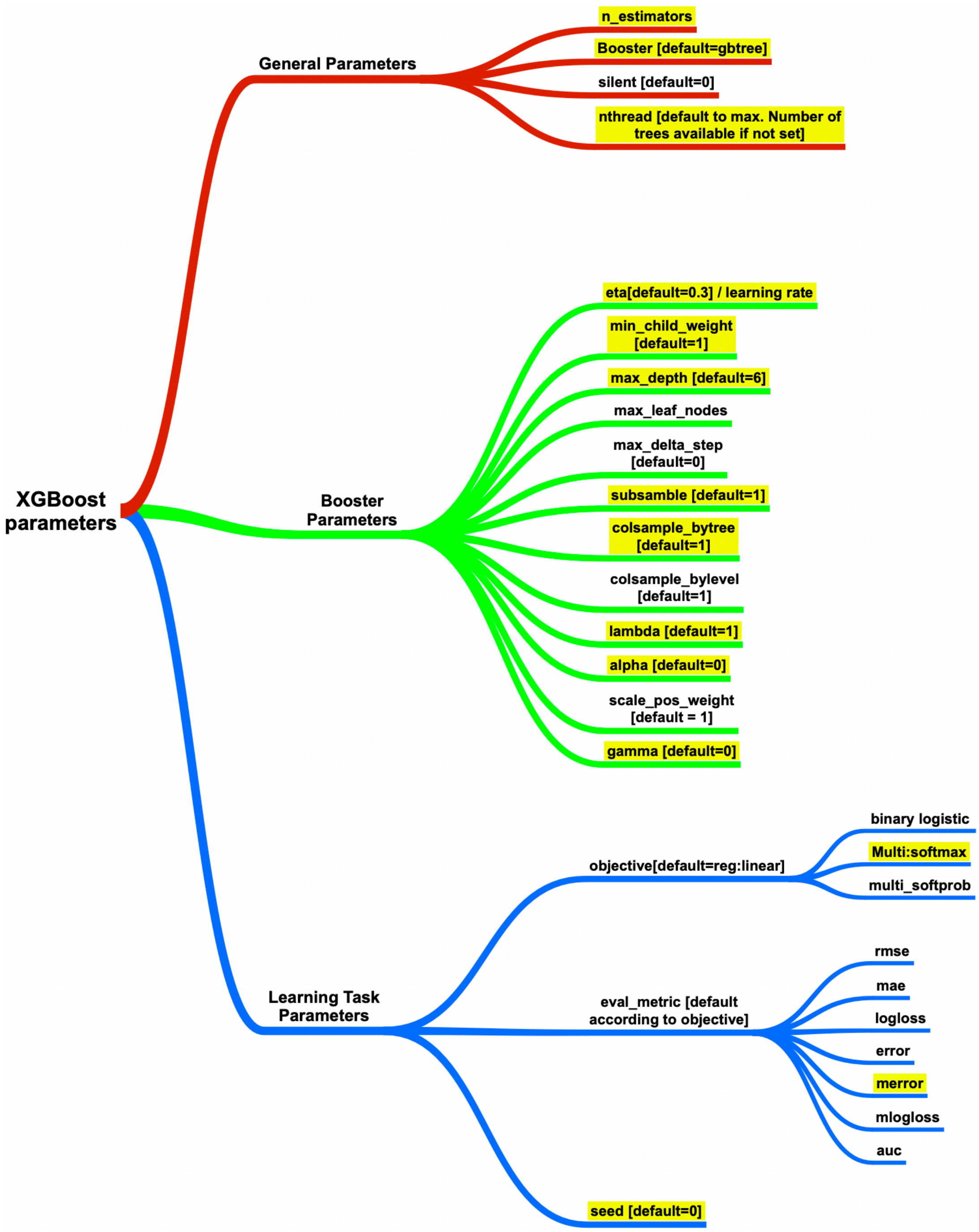


Figure 33 XGBoost parameters. Parameters that were tuned using grid search are highlighted in yellow

The next step was to tune `max_depth` and `mean_child_weight`. The ranges tested were 3-20 with step1, and 1-6 with step 1, respectively. These two variables were cross validated with grid search simultaneously with a scoring parameter set to accuracy. The results of grid search tuning are presented in Table 17.

Table 17 Optimal hyperparameter tunined values for XGboost for all TUG and FTSTS datasets.

Parameter	***TUG exp.***	***TUG synth.***	***FTSTS exp.***	***FTSTS synth.***
objective	'multi:softmax'	'multi:softmax'	'multi:softmax'	'multi:softmax'
Num_class	5	18	2	13
nthread	4	4	4	4
n_estimators	11	786	1	162
Learning_rate	0.1	0.01	0.01	0.1
Eta	0.2	0	1	0
Max_depth	7	3	20	9
min_child_weight	1	2	1	1
scale_pos_weight	1	1	1	1
subsample	0.85	0.7	0.8	0.85
colsample_bytree	0.85	0.85	0.8	0.85
gamma	0.1	0	0	0
reg_alpha	[default]	[default]	[default]	1e-5
Reg_lamda	[default]	[default]	[default]	[default]

Setting the previously optimised parameters to their respective optimal values and keeping all untuned parameters to their default values `gamma` next is tuned. The γ values of 0 to 0.5 with step 0.1 were tested with scoring set to 'accuracy'. The optimal results were `gamma = 0.1` for TUG experiment and `gamma = 0` for all other cases.

At the end of this step the `n_estimators` were re-evaluated and the only case that reported a new optimal value was synth-FTSTS with `n_estimators = 162`.

The next set of parameters to be tuned were `subsample` and `colsample_bytree` the ranges tested were 0.6 to 1 for each of the variables. First, a 0.1 step is used and then evaluated a second time with a range that focused around the first result with a 0.05 step for further fine tuning. The optimal results are presented in Table 17.

Then, the `reg_alpha` and `reg_lamda` parameters were tuned. The values tested with `gridsearchcv` were [0, 1e-5, 1e-2, 0.1, 1, 100] for both parameters. No benefits were observed by changing `reg_alpha` in terms of accuracy apart from the FTSTS conditions case (synthetic dataset) case were `reg_alpha = 1e-5` was the optimal. For all other cases the optimal `reg_alpha` was 0 which is the default setting. `reg_lamda` did not improve accuracy at all. So, it remained at the default value for all the models. Finally, the `learning_rate` (or eta) values were further tuned after all the previous parameters were set to the optimal values above. The final optimal results are presented in Table 17.

5.2.4.3 Neural Networks

There is a lot of discussion nowadays regarding NNs and hence it is worth investigating their applicability to home-based rehabilitation systems. Although NNs were discovered several decades ago the computational limitations back then were forbidding their application. However, given that the technology has progressed rapidly, and particularly computing hardware, more and more scientists apply NN's on various problems and datasets. From healthcare and diagnosis to driverless cars, NNs seem to be quite a promising approach. The NN's structure, consists of neurons connected between each other on different layers with a variety of architectures. Although this sounds like a simple concept, it quickly becomes quite complicated and it seems to be the leading black-box model given that it is hard to analyse and interpret how the predictions/decisions were made.

Because of the widespread adoption of NNs, general purpose libraries have been developed such as TensorFlow. TensorFlow a library which has been developed by Google, seems quite a promising approach for building a variety of applications using Python. It helps to simplify and deploy complex and large-scale NN models onto a variety of different hardware setups. TensorFlow is based on the concept of a computational graph. In a computational graph, nodes represent either persistent data or math operation and edges represent the flow of data between nodes. The data that flow through these edges is a multi-dimensional array known as a tensor. The output from one operation or group of operations is fed into the next input. While TensorFlow was designed to support NNs, any domain where computation can be modelled as a data flow graph can be supported. In order to allow flexibility and easy of programming for the NN designer, TensorFlow typically is encapsulated within an additional library called Keras. Keras provides efficient NN models that are proven to perform well when trained on sufficient, balanced data.

As discussed in the previous Section Keras is able to provide models that perform well on IoT devices and respect the computation capacity constraints. However, in Chapter 3 it is identified individualisation as a main requirement. In that respect, incrementally training NNs can become very computationally intensive and thus impossible for IoT or AmI solutions. Furthermore, as discussed at the beginning of this Section NNs have high levels of obscurity.

As discussed in Section 5.1.1 and 5.1.3, computational capacity requirements for training and thus re-training would be prohibitive for deployment on AmI [303]. Furthermore, Convolutional NNs have proven to perform similarly to XGboost in terms of accuracy in non-imaging applications [304][305]. XGboost has proved to produce models of very high accuracy in a variety of problems when compared to NNs [306] [77]. Thus, accuracy is not being sacrificed in favour of explainability and computation constraints.

Hence, this algorithmic approach does not meet the ART AI requirements and NNs were not evaluated further for this thesis as they did not meet the ART AI requirements and also the individualisation requirements on a resource constrained AmI system.

5.2.4.4 K-Nearest Neighbour

K-Nearest Neighbour, often referred as KNN, is a well-known algorithmic method for classification. In order to identify the optimal value for K the method of grid search is followed given that there is no literature based optimum value for this parameter and it is dependent on the dataset. Low values of K emerge as optimum when the data have a high level of noise and/or outliers. Another possibility is that clusters overlap in some of the dimensions [307]. In our case for every dataset the optimal value was K=1. This was identified through grid search cross validation. For our datasets this is a result of the overlap between classes as evidenced through the PCA analysis presented in Figure 31.

5.2.4.5 OLS Multiple Regression

Multiple regression in comparison with linear regression predicts the required output (value y) by adding multiple parameters in a polynomial equation. Hence, it allows us to add more data into the system and take multiple parameters under consideration in order to predict the output.

Given that different parameters can be included into the equation, multiple regression allows us to understand the importance of each parameter to the final prediction. Equations are very similar to the equations of linear regression with some differentiations. It is worth mentioning that a parameter can be ignored if including its contribution does not affect the result.

For the regression approach the Python OLS multiple regression algorithm was used [308]. No parameter tuning is required for this algorithm. The algorithm is characterised by simplicity, but in many applications does not provide high enough accuracy.

5.2.4.6 Hyper-Parameter Tunning and Comparison of Algorithms Based on Accuracy

In the previous Subsections various algorithms were presented and discussed in relation to their methods and their complexity. In [309] a review of all classification algorithms available in Python libraries is presented along with their computational requirements and effect on bias and variance. Based on this review, the most suitable classifiers to our problem specification and multiclass classification in the sklearn library are Support Vector Machine (SVM) and XGBoost for the following reasons. SVMs are powerful yet light weight in terms of computational requirements. The default one-vs-all multiclass classification approach is used. On the other hand, XGBoost can mitigate variance and bias in datasets while automatically correcting model biases through the random forest underlying approach. This also results in reduced risk of over-fitting. XGBoost operates on a one-vs-all principle but optimises the models based on SoftMax probabilities.

Although NNs will not be included in this thesis, initial experiments were conducted, and early findings included in the Future Work Section 7.2. The details regarding the hyperparameter optimization were aforementioned in the previous Sections for each algorithm. In this Subsection comparative experiments were presented as well as evaluation of the results based on accuracy.

With a relatively small number of features (seven), initial implementation of simpler approaches such as SVM and Polynomial Regression was carried out. All possible kernels were tested with the SVM including linear, polynomial of degree 4, rbf, and sigmoid. In all cases the regularisation parameter C=1which is the default value in sklearn Python library.

The best performing kernel of the SVM approach was the linear and only the accuracy of this kernel is discussed in this Section. However, the subtle correlations between the features as well as the sheer number of classes in the synthetic dataset proved challenging for these simpler approaches as demonstrated by the very low accuracy in Table 18. For the regression approach, the Python OLS multiple regression algorithm was used [308]. Both linear SVM and the regression approach are excluded from further hyper-parameter tuning because of their accuracy being <40% in all multiclass cases.

Higher complexity models were then tested. They were selected because they have low computation requirements for the prediction phase and were appropriate for the ambient intelligence hardware developed in our experiments [213] as previously discussed in the earlier Subsections.

XGboost was initially set up with maximum depth equal to double the number of classes and the `multi:softmax` objective function which are required settings for multiclass problem with all other hyper parameters set to default. This is discussed as XGBoost before hyperparameter tuning in in Table 18.

Table 18 5-fold cross validation accuracy results of ML models. Accuracy is calculated as the fraction of correct predictions using `sklearn.metrics.accuracy_score`

Model	***TUG exp***	***TUG synth***	***FTSTS exp***	***FTSTS synth***
SVM	0.3099	0.3921	0.9778	0.3625
Regression	0.35	0.0605	1	0.0705
KNN(k=5) before tunning	0.475	0.7006	0.6917	0.6791
XGBoost ***before*** *hyperparameter tunning*	0.4600	0.7667	0.9778	0.7827
XGBoost ***after*** *Hyperparam*eter *Tunning*	0.65	0.8006	1	0.7965

Moreover, the results after hyperparameter tuning are presented for comparison and to demonstrate both the heightened capacity of the earlier version and the impact of tuning. KNN was also parametrised with k equal to 5 prior to tuning. The number 5 for nearest neighbours was used as an even number that respects the low computation requirements and is a starting point to be further optimised. As these methods reported higher accuracy, further hyper-parameter tuning is undertaken in the next Section.

Again, these results are provided as a baseline for comparison for the hyperparameter tuned version of the models. The accuracy of all the models were evaluated using an 80% training and 20% test dataset split and 5-fold cross validation [310]. Note that it is shown in [311] that a number between 5 and 10 folds provides similar results.

Table 18 presents the accuracy score of the model (all classes) averaged across the 5-fold cross validation. This table demonstrates the accuracy of each model for the particular dataset with the purpose of identifying which model would be most suitable for further tuning and evaluation. XGBoost and KNN have a significantly better accuracy in all datasets compared to other models. Additionally, XGBoost which is based on decision trees and KNN are easier for humans to understand when many features are involved [241] with little or no post-hoc analysis [229] as discussed in earlier Sections. Moreover, the XGBoost Python implementation can provide results such as the probability of each cluster, and the decision weights used for each tree to improve transparency and support interpretability and explainability and will be further discussed in Section 5.2.5.

Differences in the results between the experiment (exp.) and synthetic (synth.) datasets presented in Table 18 can be explained by either the increased bias of the exp. dataset in the TUG case or the larger number of classes (5and 2 in exp., vs. 18 and 13 in TUG and FTSTS datasets, respectively) and higher number of data points in the syth. dataset in the FTSTS case. Indeed, as a result of bias (see Figure *30*) the classification error on exp. dataset is generally higher for TUG (e.g., TUG exp. KNN vs TUG synth. KNN results). For FTSTS, on the other hand, the increased number of classes (e.g., 13 for FTSTS synth. vs 2 for FTSTS exp.) can lead to higher probability of misclassification affecting accuracy of classifiers when applied to synth. datasets (e.g., FTSTS exp. XGBoost vs FTSTS synth. XGboost results). Finally, for TUG and FTSTS the difference between exp. and synth. must be discussed in terms of dataset size. In the case of exp., smaller datasets had a profound effect on accuracy.

In summary, exp. dataset results are negatively affected by bias and small number of training samples. On the other hand, the results on the synth. dataset demonstrate that similar or better accuracy can be achieved with well-designed dataset, even with significantly higher number of classes.

Finally, using grid search cross validation [312] several sessions were performed of hyper-parameter tuning to identify the most optimal parameters and develop separate XGBoost models for each dataset. The hyper parameters tuned for XGboost were: `booster, eta, n_estimators, max_depth, min_child_weight, gamma, learning_rate, subsample, colsample_bytree, reg_alpha, reg_lamda`. Similarly, the k parameter of KNN was tuned using grid search cross validation.

The feature importance obtained by XGboost classifier after hyper-parameter tuning, is presented in Figure 34. Features have clearly different contribution to the prediction in the case of the experimental dataset models and the synthetic models. This is predominantly a result of reducing bias between male and female participants and removing the original experimental data biases in the correlations of height, weight, and BMI with sex. Additionally, as the synthetic datasets cover a wider variety of ages and conditions, the correlation between the condition and the completion time is reduced compared to the clear correlation between stages of difficulty and completion time in the experimental dataset. Thus, the synthetic dataset led to better balancing the features.

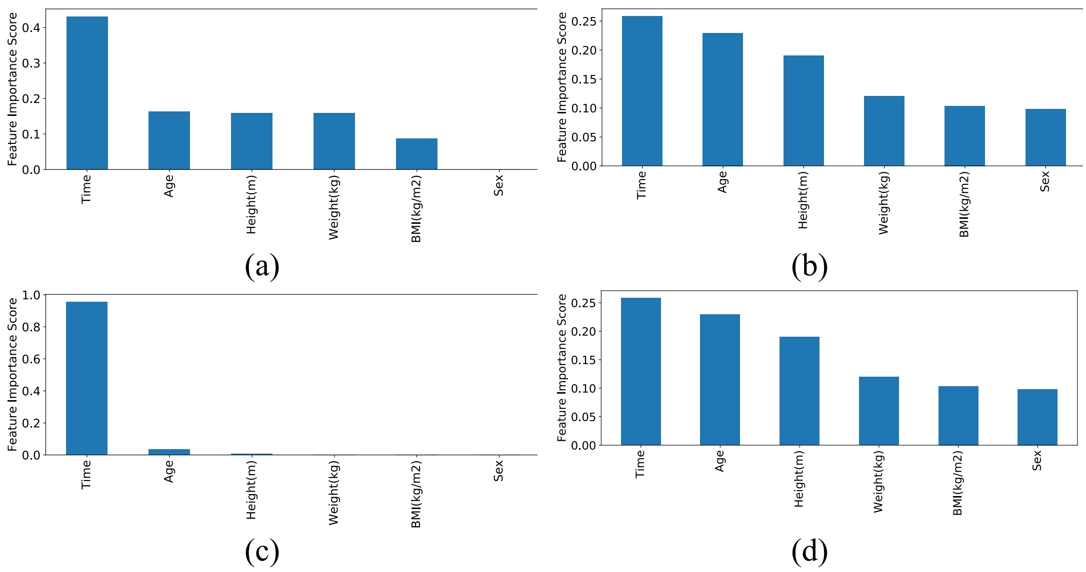


Figure 34 Impurity-based feature importance of the xgboost forests based on mean decrease in impurity as calculated by XGB Classifier.feature_importances_for (a) TUG exp., (b) TUG synth., (c) FTSTS exp., and (d) FTSTS synth.

5.2.5 Hybrid ML Approach for Individualised ART-Driven Rehabilitation

According to the results of the ML method evaluation in the previous Section, it is evident that the XGBoost approach demonstrates the highest prediction accuracy among all tested approaches. However, using this model alone does not satisfy the requirements for our proposed system as presented in Chapters 3 and 4, since the scope of the sensory system requires both the monitoring of rehabilitation progress over long periods of time against goals, and the ability to adapt/diagnose a variety of conditions related to stroke survivors such as co-morbidity developed after discharge. Moreover, identifying the areas of difficulty relates to identifying challenges in the user's daily activities. Thus, combining the models developed for the experiment and/or the conditions is required for each of the two medical tests. Additionally, the combination is necessary to reduce bias and improve variance in the training data.

Figure 35 presents the method used to combine the algorithms through the proposed hybrid ensemble learning approach inspired by the stacking method.

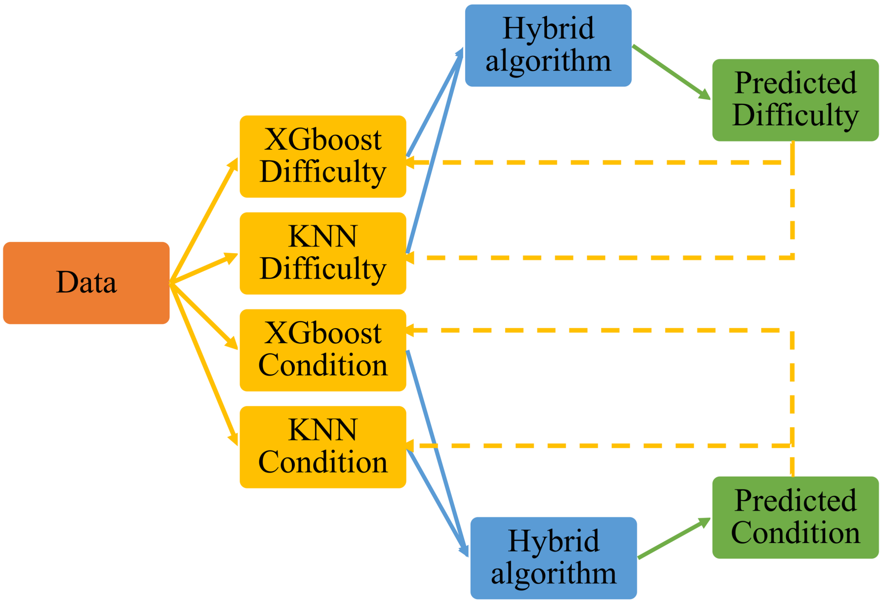


Figure 35 Graphical representation of the hybrid stacking model combining Difficulty prediction and Condition prediction for a medical test. The data refers to experimental plus synthetic data.

The algorithm uses the rate of improvement for the specific user as an additional feature and combines the XGBoost models (Experimental & Synthetic data trained) along with a further supervised model (KNN-based) to establish a final prediction of both the condition and the difficulty faced by the user (Algorithm 2). The model is incrementally re-trained according to individual performance. The XGBoost prediction returns both the condition class (e.g., Stroke, as defined in Section 5.2.3), and the difficulty class (e.g., difficulty to turn, as defined in Section 5.2.2).

The system is initialised with the condition the user is diagnosed by a medical professional. If the condition predicted is consistently different to the initialised one, then a co-morbidity may be developing. The algorithm monitors this against the baseline prediction set (baseline) and produces outputs to alert the user to this event.

Algorithm 2 Hybrid Learning Method

```
 Time_completion ← ReadSensors()
 improvement ← PolinomialFit(history).get_rate()
 baseline ← MostFrequentPrediction()
 xgb_prediction ← XGBoostModel.predict(time_completion)
 knn_prediction← KNNModel().predict(time_completion)
Closest, farthest ← CalculateEucledianDistance(xgb_prediction,
knn_prediction,healthy)
If improvement = true then
  Retrain(closest) {Closer to Healthy}
 else if improvement = false then
  Retrain(farthest)
else
  Retrain(baseline) {Steady}
end if
userState ←
stateCalculation(sensorData,improvement,finalprediction)
```

Additionally, the raw sensor data from the system, the improvement, baseline and final prediction are used to generate a set of user state flags (`stateCalculation`). These flags are generated to improve the safety of the user of the system and inform carers if necessary. Those are:

- The test starts and does not complete, possible indication of fall.
- Noise and unexpected sensor inputs, possible miss-use, sensor faults or multiple users interfering with the test.
- Improvement rate is too high, and prediction deviates significantly from baseline, possible indication of user forcing themselves to achieve the goal faster.
- Negative improvement is consistently reported, possible deterioration of user condition.

The KNN models are trained with exactly the same training set as the XGBoost models and incrementally retrained as in `Algorithm 2`. The grid search tuned k-value was used here for each of the four cases. As the KNN is trained in a supervised fashion, it was used in combination with XGBoost to provide a corrective mechanism. Moreover, the KNN method had a better accuracy compared to other tested models which made it the only alternative candidate (see Section 5.3) and was also used with the SMOTE algorithm for dataset balancing in earlier section of this chapter. The goal set in the system is to gradually decrease the completion time that the patient takes to achieve the exercise (goal-oriented rehabilitation). The algorithm is implemented with the parameterisation of the goal in mind so that different levels of improvement rate can be defined to make goals realistic for each individual. Finally, both the XGBoost and KNN methods have previously been used in IoT applications with real-time design requirements, and the Hybrid algorithm does not add statistically significant computational overhead as it amounts to a simple Euclidean distance calculation between the new point and the centroid of the healthy cluster and an `if-then-else` statement. This is significantly smaller than the third model required for stacking. Thus, the algorithm is appropriate for use in AmI in terms of computational requirements. Additionally, XGBoost has a light computational footprint incremental training method in its Python implementation.

5.2.6 ART AI Design Approach

This Section discusses the design considerations and ART AI concepts of the proposed method. As identified by [232], medical knowledge was embedded in the design of the system both through the medical journal published information used to generate the condition datapoints in the synthetic dataset and the medically approved tests implemented. Additionally to the outputs discussed in the earlier Sections, also results were generated demonstrating the inner workings of the algorithm (which model's output was selected by the hybrid approach, what the KNN and XGboost models predicted individually, which were the other options and with what probability).This addresses the layered feedback requirement identified in [233] and the exposure of the embedded optimisations of the model as discussed in [232].

All these results can be used as future work by the interface to enable the user to a) understand and b) overwrite the algorithm decisions as suggested in [232], [233]. The benefits address both the interpretability and explainability requirements set out in this chapter. The previous Section discussed the approach for eliminating any bias generated by the system designer leading to fair and non-discriminatory decisions

The 6 design time considerations were followed [208] in the development of the proposed method as follows:

1) Bias of the experimental dataset, in terms of the female/male balance, is mitigated using SMOTE and balanced data between female/male participants through the synthetic dataset. Biases in the remaining features (e.g., Age, Height, Weight) were also addressed as a result of the SMOTE method and the wide variety of sources used in the synthetic dataset;
2) ethical principles followed as presented in [253], [313] when designing the datasets and developing the inference algorithms;
3) interdisciplinary team (electronic & electrical engineering for the design, manufacturing and development of the system and Biomedical Engineering for guidance and design of the experiment and testing) and literature review were significant contributors in the identification of design requirements [56], [213] as was also the clear approach to inclusion/exclusion of each paper;
4) data provenance both in terms of input and output is transparent to the user and published at the edge node using open protocols. Both KNN and XGBoost are categorised as transparent by [229] and all the generated model information is also included in the model output. However, for security reasons, only authorised users can view and access both input and output information (system login functionality [56]);
5) the feedback is simplified so that lay people can understand it as presented in [56] and users can inquire further into the factors affecting the decision following the model presented in [234]. This follows the ART AI principles of [232], [233];
6) decisions are justified both through the probability results, the persons rate of improvement and the model's selection process. Results are available to the user.

In the case of health applications, the value system to be used is one that does not in any case worsen the quality of life or the health of the patient. As a result, the manner in which feedback is presented had to be adjusted to ensure that the user is not urged to always perform a faster exercise but rather focus on long-term goals. Our proposed method is focused on beneficence by its application specifications (rehabilitation support) and through the goal setting theory approach discussed in Chapter 1.

Furthermore, the alerts generated by the system support both the identification of system failures but also the Responsible AI principles. The system generated useful information for the state of the user that can identify user safety and critical events.

5.3 Evaluation and Results

The evaluation methodology is summarised in Figure 36.

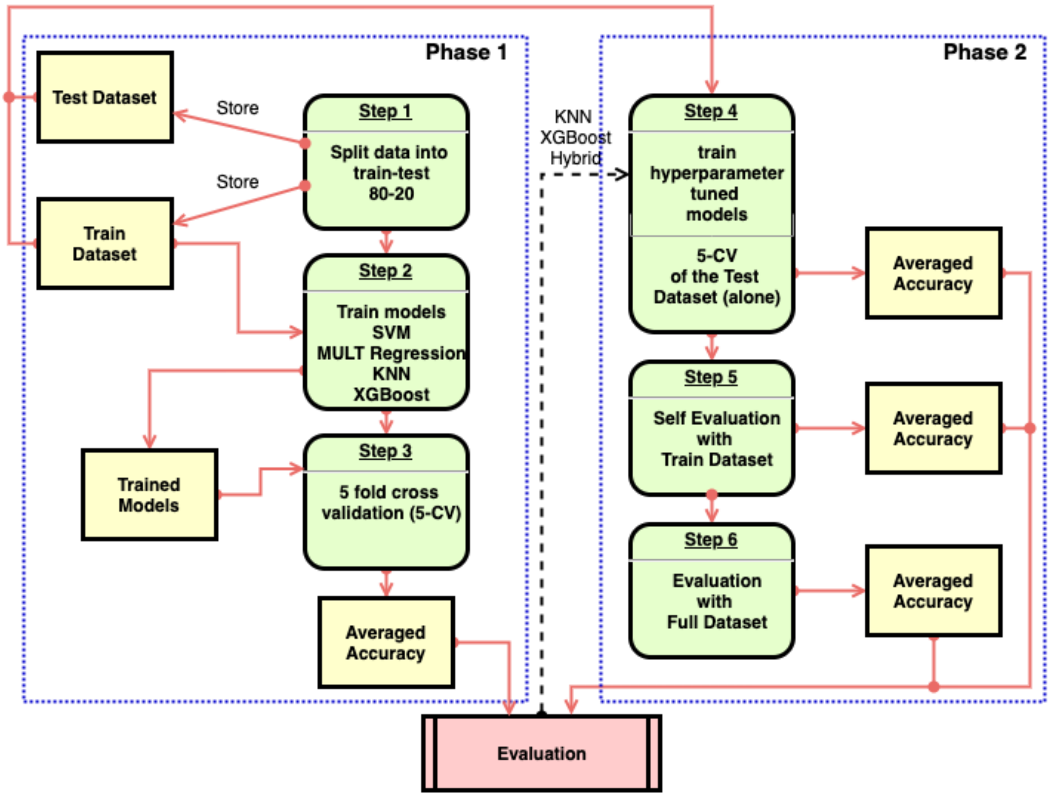


Figure 36 ML Model Evaluation Methodology Flowchart where "Evaluation" stands for the critical evaluation of the results presented in this thesis. Phase 1 Results were presented in Section 5.2.4.6. Phase 2 is discussed in this Section.

The evaluation methodology was used for evaluation of the model and comprised of the following steps:

Phase 1:

1) Split the four datasets using an 80%-20% split. This generates one train-test split for each of the TUG exp, TUG synth, FTSTS exp, FTSTS synth datasets. These splits are saved in static variables, 4 for training and 4 for testing, so they can be reused in Step 4 (a total of 8 variables).
2) Then each model is trained with the 80% train dataset and tested with the remaining 20% test dataset.
3) For verification of the results in Step 2 a 5-time cross validation (CV) over the full dataset (without using the Step 1 train-test split) is executed for each of the TUG exp, TUG synth, FTSTS exp and FTSTS synth, respectively.
    a. The outcome is the generation of 5 random train-test splits (of 80% train and 20% test) for each dataset.
    b. The accuracy of prediction of the test data is calculated by comparing the true labels (already known) against the predicted labels (predicted by the trained model).
    c. The result is 5 different accuracy figures (one for each random split of the 5 cross validation tests) for each dataset individually.
    d. The 5 accuracy figures of each dataset are averaged to get the overall accuracy of the dataset.
    e. The outcome is 4 final accuracy results one for each dataset. These average values are presented in Table 18.

Phase 2:

4) After the models are selected based on accuracy results as presented in Section 5.2.4.6 and the hybrid model is developed to improve the overall accuracy as presented in Section 5.2.5, the 8 variables from Step 1 (initial train-test split) are used to evaluate the performance of each model.
    a. Each model is trained with 80% then tested with the remaining 20%.
    b. 5-time cross validation using the test set alone is performed to see repeatability of the results. Repeatability is important to prove that the model will have consistently the same accuracy over time.
    c. Then, the accuracy outcome of the prediction of the cross validation for each dataset is averaged, the results are presented in Table 19.
5) A self-evaluation was run using the train dataset (80%) and test with the same train dataset without the assigned labels (80%) to understand self-prediction

accuracy of each model. Self-prediction is defined as the ability of the model to provide correct predictions for an already known datapoint used for training. This enables us to see if there is overfitting, or underfitting tendencies in the models. The results are presented in Table 20.

6) Finally, 5-time cross validation is performed of the full set of each data set (random 80% and 20% splits) as presented and discussed in the following paragraphs. This allows us to evaluate the robustness of the model. Robustness refers to the ability of the model to avoid overfitting or underfitting. The average accuracy of this final full set cross validation is discussed in the following paragraphs.

The accuracy, comparison between the proposed hybrid approach and benchmarks, averaged over all classes, is presented in Table 19. These results justify the combined use of the two models based on accuracy.

Table 19 Accuracy results of XGBoost, KNN, and hybrid models using the test dataset

Model	TUG exp	TUG synth	FTSTS exp	FTSTS synth
XGB 5-f	0.625	0.6052	0.7	0.577
KNN 5-f	0.45	0.5932	0.7	0.5930
Hybrid 5-f	0.5789	1	1	1

Table 20 Accuracy results of XGBoost, & KNN models using the train dataset (self-evaluation)

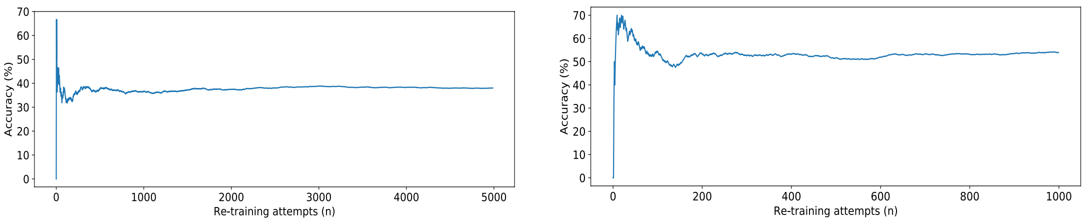
Model	TUG exp	TUG synth	FTSTS exp	FTSTS synth
XGB 5-f	0.8313	1	0.9697	1
KNN 5-f	0.5938	0.7927	0.9417	0.7776

According to Figure 36 Step 6, for TUG, the accuracy of the hybrid approach on the full dataset is 57.89% (matching the one for the test dataset in Table 19) which is an improvement from the original 42.5% for the full dataset using the XGBoost model alone. Similarly for FTSTS, the accuracy of the hybrid approach was 100% (again matching the one for the test dataset in Table 19) which is an improvement over the original value of 70% for XGboost over the full dataset (also matching the one for the test dataset in Table 19).

To further investigate the Hybrid approach's reduced performance in the TUG experiment dataset trained model (Table 19) the prediction accuracy is examined in each of the considered classes. Difficulty to walk is most often misclassified by XGBoost (60% miss classification as difficulty to turn or difficulty to stand). This is corrected by KNN in only 17.89% of the misclassified cases. At the same time however, difficulty to turn (28.57% misclassification), difficulty to stand and fast (0% misclassification for both) have a far better prediction accuracy. Moreover, normal is occasionally misclassified as fast (57%), but as they are both healthy conditions, this misclassification is not considered further. Similarly, for FTSTS, the misclassification was 0% for both difficulty/slow and fast classes.

Additionally, it can be inferred from Table 19 that the number of data points in TUG synthetic versus FTSTS sythn. do not have a significant effect on the accuracy of the hybrid model. It is rather the misclassification of some classes in TUG experiment that have a more profound effect. The reasons behind this misclassification are further discussed in the following section.

The accuracy of the proposed combined approach not only demonstrated improved outcomes (Table 19) but it improves over time as presented in Figure 37. The graphs display the prediction accuracy cumulatively calculated over the full history of use.



(a) (b)

Figure 37 (a) Prediction accuracy improvement over time for TUG with difficulty to walk (one of the difficulties that are miss classified in 30% of the cases) and Geriatric classification used 5000 times by individual 1, (b) Prediction accuracy improvement over time for TUG performed Normally (anon-difficulty that is miss classified as Fast in 71% of the cases) 1000 times by individual 3 classified as Dementia Severe.

As demonstrated in Figure 37 the number of correct predictions increases over time resulting in a gradual increase in prediction accuracy. The graphs demonstrate that the system’s behaviour is similar to that of a positive feedback loop in control theory. This is an expected result as the model is incrementally re-trained in a similar manner. This will continuously improve the sensitivity of the model to the specific individual. Moreover, as the user continues to use the system, a possible false negative in a single use will be counterbalanced by true positives/negatives over the lifetime of the system. This is examined further through Figure 38.

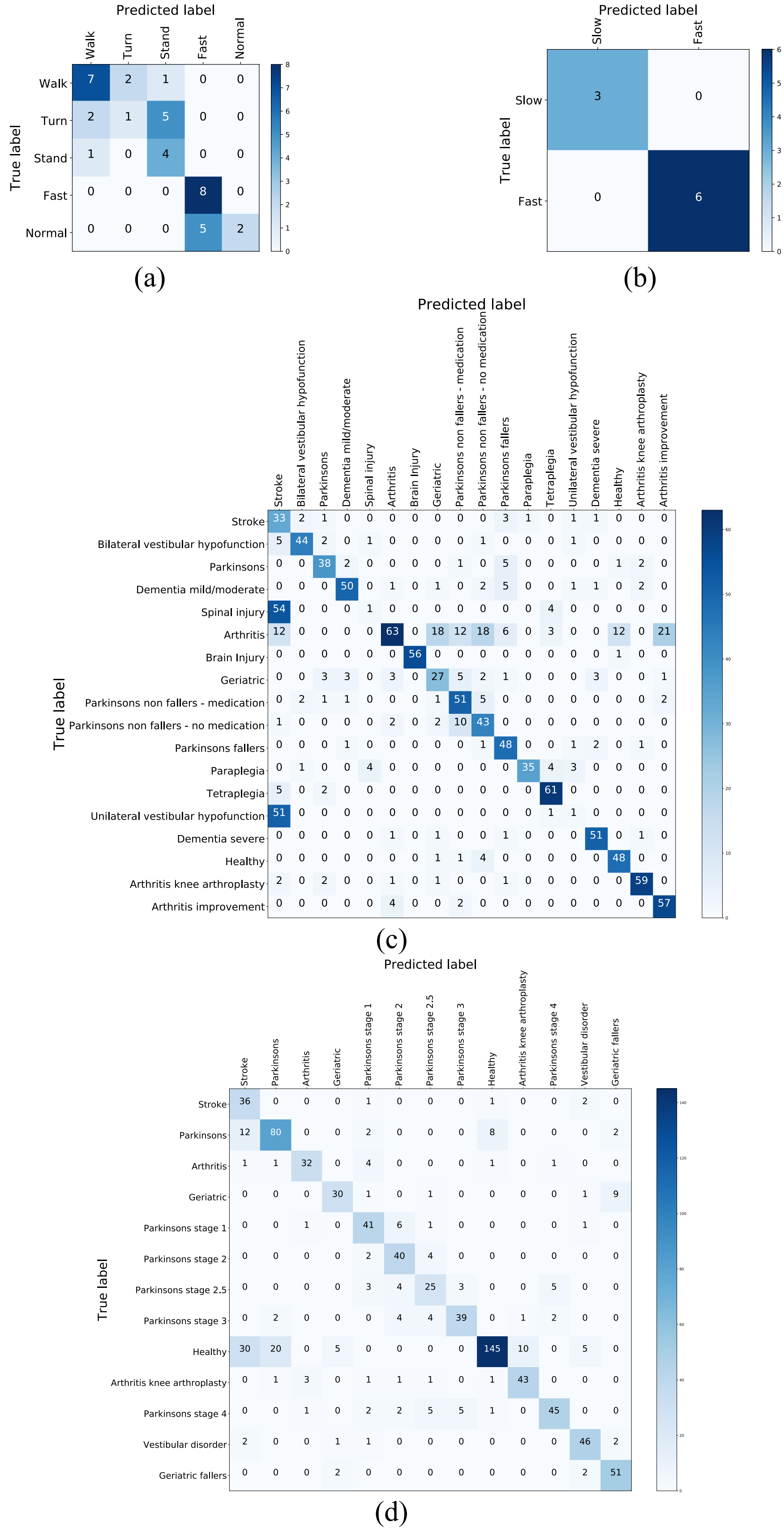


Figure 38 Confusion matrices for the hybrid model presented separately for (a) TUG difficulty, (b)FTSTS difficulty, (c) TUG conditions, and (d) FTSTS conditions.

To further investigate the low accuracy of the TUG test, Figure 38(a) Figure 38(c) demonstrate the confusion matrices of the hybrid model for the difficulties and conditions, respectively. Additionally, the confusion matrices for FTSTS are presented Figure 38(b), Figure 38(d). It is evident that Difficulty to Turn is often confused with Difficulty to Stand or Difficulty to Walk. Moreover, Spinal Injury and Unilateral Vestibular hypofunction are often confused with Stroke.

This explains the relatively low score in TUG accuracy. Furthermore, Arthritis is most often miss classified as a range of different conditions. These observations will be discussed in Section 5.4. Overall, the sensitivity (Equation 28), specificity (Equation 29), precision (Equation 30) and f1 score (Equation 31) for four confusion matrices (corresponding to the four datasets) are presented in Table 21, where:

$$Sensitivity = \frac{TP}{TP+FN} \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

$$f1\ Score = 2\ \frac{Precision * Sensitivity}{Precision + Sensitivity} \quad (13)$$

and where TP is the true positives, TN the true negatives, FP the false positives and FN the false negatives as dictated by the confusion matrices for each model.

Table 21 Hybrid Model Confusion Matrix Metrics

Metric	**TUG exp.**	**TUG synth.**	**FTSTS exp.**	**FTSTS synth.**
Sensitivity	0.33	0.9	1	0.87
Specificity	0.78	0.94	1	1
Precision	0.78	0.94	1	1
f1 Score	0.47	0.92	1	0.93

Further, the most frequently predicted condition for each of the TUG and FTSTS experimental dataset classes was investigated. The results are presented in Table 22. These results demonstrate the validity of the claims in Chapter 4, specifically in Table 11 and Table 12, where the simulated experimental conditions were related to patient and elderly subject completion times.

Table 22 Hybrid model most frequent association of difficulty and condition for the experimental test dataset of the 8 participants

Test	Cluster	Condition
TUG	Diff. Walk	Geriatric
	Diff. Turn	Paraplegia
	Diff. Stand	Dementia Severe
	Normal	Arthritis
	Fast	Unilateral Vestibular Hypofunction
FTSTS	Diff.	Vestibular disorder
	Fast	Parkinson’s stage 4

In all cases the hybrid approach has excellent precision and specificity. This is important as false alerts are avoided and do not cause unnecessary alarm to carers. As expected, TUG experiment performs lower than other models in sensitivity due to the misclassification between the aforementioned difficulty classes. We consider this to be less problematic as the user is not misclassified as healthy.

These results demonstrate the validity of the claims in [213] where the simulated experimental conditions were related to patient and elderly subject completion times.

Similar tests have been performed for each participant of the experiment and for each simulated difficulty. The prediction accuracy improved or stayed the same over time in 81.82% of the tested cases for the test set. Accuracy improved over time in 100.0% of all the tested combinations of {individual, difficulty} in the full set combining test and train data.

5.4 Chapter Conclusion

Based on the results, the proposed method is capable of providing accurate home-based rehabilitation support, with improvement of individualisation in terms of accuracy overtime based on observations of a specific individual. The method is suitable for both rehabilitation goal setting as well as diagnosis of comorbidities.

In the case of difficulties in the experimental dataset for both TUG and FTSTS tests, time of completion has clearly a higher importance compared to other features. However, in terms of conditions (synthetic dataset) all features are contributing in some cases almost equally (refer to Figure 34). The high importance of the time of completion feature in the experimental dataset is expected because of the bias presented in the dataset due low variability in terms of other features (e.g., all participants belong to the same age group and only one was female).

In the confusion matrices the most often miss-classifications of the hybrid model are identified. These are attributed to the use of a very small number of sensors described in Chapter 4. Given that a small number of sensors were utilised, the completion time of the different stages of the TUG was not extracted from the overall completion time and thus not supplied as separate features. Instead, the full time of completion of the TUG is the used feature. This results in the model's inability to clearly differentiate between the two difficulties of Turn and Stand.

Furthermore, it was identified that two conditions in TUG are always miss-classified as Stroke. This is believed to be due to the similarity of the effect of the condition on the walking pattern. For example, a severe stroke could result in severe vestibular hypofunction or have similarities to spinal injury due to the effect on the motor control system. Moreover, there was no differentiation between mild, moderate or severe stroke symptoms. It is possible that a refinement of the dataset to differentiate the severity of stroke would improve performance of the hybrid model. A similar effect is seen in the case of Arthritis where various stages of arthritis are not differentiated. On the contrary, the differentiation between stages of dementia and Parkinson's have proven to have better performance as evident from the confusion matrices in both TUG and FTSTS. In terms of FTSTS the Difficulty/Slow class is never confused with the Fast class. However, the Healthy class is very rarely miss-classified as Stroke (Figure 38(b), Figure 38(d)). As before it was hypothesised that a fine-grained presentation of stroke stages in the dataset would address this issue. It is possible that very mild cases or early stroke signs are miss-classified as Healthy which is reasonable as mild stroke patients might not have significant disability developed. Addressing this issue would be future work. In summary, the contributions of this chapter are presented in Figure 39.

These contributions are addressing RQ3 and RQ4 as presented in Chapter 2:

a) Methodological steps to produce a new synthetic dataset based on statistical results reported in the literature for training ML algorithms to avoid bias in autonomous system outcomes, (Section 5.2.3);

b) A novel hybrid ML algorithm to meet the individualisation, interpretability, and ART design considerations while maintaining a low computational footprint, (Section 5.2.5);

c) Interpretability of the designed solution, including feature importance for a patient-centric individualised, responsible home-based rehabilitation support (Sections 5.2.5, 5.2.6).

d) A detailed simulation performance comparison and analysis demonstrating that the proposed approach outperforms existing work, used as benchmark, by 5% for FTSTS and 15% percent for TUG test (Sections 5.3).

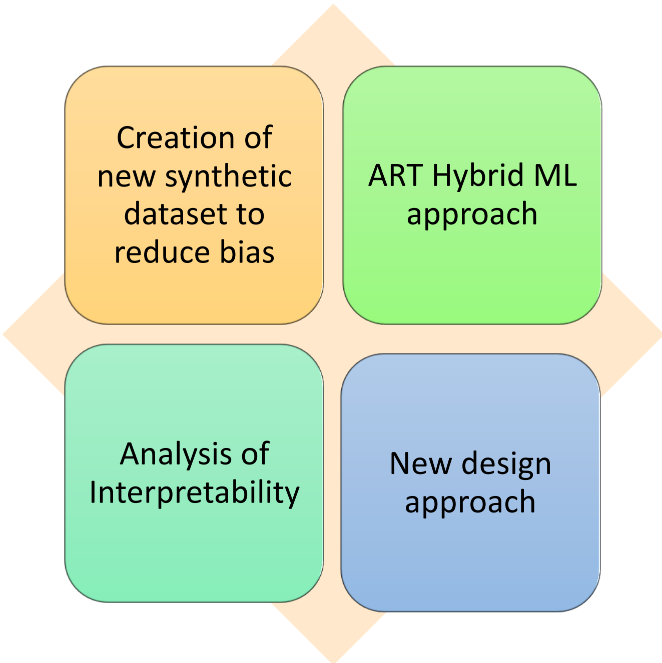


Figure 39 Chapter 5's contributions

6 Discussion

This chapter discusses initially the main findings regarding the successful criteria of systems for rehabilitation and how they can motivate patients in order to engage with their rehabilitation goals. Later the proposed system and the experiments that were carried out are discussed and the chapter conclude with the discussion of the proposed ART AI hybrid approach for individualised rehabilitation and comorbidities detection. The discussion in this chapter will cover the thesis as a whole.

Based on the analysis in Chapters 1 and 3, an ideal home rehabilitation device should meet all the criteria and requirements identified in Chapter 3 as follows. The system needs to avoid wearable or intrusive components. It needs to support enhanced motivation and engagement by being incorporated into the daily activity routine. It must be cost-effective and not complex to install, maintain, and use. It needs to support the needs of all patients, regardless of age and background. Moreover, it needs to be portable and transferable to other domains such as diagnosis of co-morbidities. Data and patterns from online databases are quite important to tailor rehabilitation, as the device can learn the patients' requirements and goals, adapt to their individual needs, and provide suitable challenges, for example, through ML Individual choice and personal control are mandatory for success.

This thesis suggests an AmI system which potentially could assist patients with their rehabilitation goals. The successful system focused on supporting specific daily activities that have measurable outcomes specified in recognised health care tests. The proposed system in this thesis initially was suggested in order to increase patient's rehabilitation goals by increasing the level of difficulties through daily activities by providing sufficient feedback which was easy to interpret.

However, given that there are no data available for all daily activities in a home environment further process, comparison and evaluation of such a system would introduce a significant challenge. In order to evaluate the system, the state of the art was followed and two certified NHS tests was selected, the TUG and the FTSTS. Additionally, information regarding the relation between time of completion and condition diagnosis exist. Hence, by designing, manufacturing and testing the system with human subjects as presented in Chapter 4, experimental data could be used in order to carry out comparisons against data from a synthetic dataset as presented in Chapter 5. In order to evaluate such a system, experiments had to be carried out with patients. According to the results, the automated sensor system was fit for purpose and has been validated for use with statistically significant accuracy ($\varrho c>0.99$, $R^2>0.94$, ICC>0.96) while the participants identified the system as engaging and motivation enhancing.

Training and testing were carried out on various ML model and initial steps were made in order to automate a procedure which could be beneficial for the user in order to: (a) provide sufficient monitoring, (b) carry out early diagnosis or raise flags in case of co-morbidity, (c) provide a higher level of engagement with rehabilitation goals by increasing self-efficacy, and (d) reward patients progress by providing sufficient, simple and interpretable feedback.

Based on the results of Chapter 5, the proposed method is capable of providing accurate home-based rehabilitation support, with improvement of individualisation in terms of accuracy overtime based on observations of a specific individual. The method is suitable for both rehabilitation goal setting as well as diagnosis of comorbidities. An important observation is the feature importance of the XGBoost model which is evidently different for each of four datasets. The final model reaches up to 100% accuracy for FTSTS in both the prediction of difficulty and the prediction of associated patient condition. In the case of TUG, the performance reaches up to 100% in the prediction of patient condition and 83.13% in the prediction of area of difficulty.

Though a direct comparison with other ML-based approaches used for FTSTS and TUG tests cannot be made due to unavailability of common datasets and minor differences in classification of test stages, by analysing the accuracy results reported in the literature, it is concluded that the proposed hybrid model achieves acceptable accuracy that surpasses the results reported in the literature. Indeed, the hybrid model reports higher accuracy than state-of-the-art AI methods that use intrusive means of monitoring such as cameras or the Kinect sensor. For example in [314] FTSTS decision tree models and k-NN have demonstrated 92% and 91% accuracy, respectively. In [315], the accuracy of the proposed classifier was 94.68% for FTSTS, and a review of similar publications demonstrates that this is the state of the art for

FTSTS AI models. Similarly, in [316], accuracy of all assessed classifiers including NNs did not exceed 85% for TUG. As a result, our proposed hybrid approach improves accuracy over and above the state of the art for both tests while also addressing a series of constrains such as computational requirements, incremental re-training, and ART AI, thus, addressing all the criteria in Chapter 3 to enhance engagement and motivation.

7 Conclusion

This thesis demonstrated that the existing approaches in home-based rehabilitation do not meet all the criteria in the motivation, acceptance, and technological categories required for engagement and motivation enhancement. The thesis identified the criteria for a system that will provide the required level of self-rehabilitation commitment as nonintrusive, nonwearable, motivation and engagement enhancing through a list of motivation methods, individualized, supporting daily activities, suitable for the elderly, cost-effective, simple, transferable, and intended for use in rehabilitation.

A low-cost system to automatically perform the TUG and FTSTS medical tests was designed and deployment of. A detailed methodology was presented to assess a home-based rehabilitation system's accuracy against the test specifications, benchmarked against NHS standard practice and ground truth established through video recording. The system's transferability was demonstrated to other daily activities and more than one NHS test. The results demonstrated that the stopwatch measurements have an inherently higher PE compared to the golden standard video measurements due to the human error factor. The automated sensor system is fit for purpose and has been validated for use with statistically significant accuracy ($\rho c > 0.99$, $R^2 > 0.94$, ICC > 0.96).

To achieve the requirement of individualisation this thesis presented a ML approach. Methodological steps to produce a new synthetic dataset based on statistical results reported in the literature for training ML algorithms to avoid bias in autonomous system outcomes was presented. A novel hybrid ML algorithm based on stacking, XGBoost and KNN was developed to meet the individualisation, interpretability, and ART design considerations while maintaining a low computational footprint. Interpretability of the designed solution was analysed, including feature importance for a patient-centric individualised, responsible home-based rehabilitation support. The re-training element proves to be incrementally improving overall model accuracy which means that as the user continues rehabilitation at home the device better adapts to the user's specific difficulty areas and conditions. A detailed performance comparison and analysis demonstrating that the proposed approach outperforms existing work, used as benchmark, by 5% for FTSTS and 15% percent for TUG test.

7.1 Limitations

The system was not tested with elderly or patients. The relevance of this early technology to the patient population was demonstrated through comparative analysis with the international database. However, experiments with elderly subjects will be required as further evaluation steps.

The system could potentially support rehabilitation but further clinical trials would be required to measure this benefit as well as the motivation enhancement and engagement in the patient population.

Fast FTSTS ($R^2 = 0.92$) was not as accurately captured while the fast TUG test was uncorrelated between system and video ($R^2 = 0.07$). The limitation of the very low-cost motion detection sensor is apparent in these two sets of experiments as the sensor's delay in recording the event is significant and affects the recorded time. However, as the system is designed to be utilised for rehabilitation and incorporation of daily activities for increased engagement, the range of fast TUG is assumed with the scope of the study targeting less capable adult subjects.

Miss-classification occurs between the difficulty to turn and difficulty to stand cases. Also, normal TUG tests are confused with fast in 57% of the cases. The confusion matrices imply that some of the patient conditions such as Stroke and Arthritis are largely generalised classes and if broken down to severity the model could improve. Furthermore, to better differentiate between difficulty to turn and difficulty to stand the TUG completion time feature could be split into two additional features for individual stages of the test.

7.2 Future Work

Additionally to addressing the above limitations, future work could also follow additional directions to expand the research. In the proposed system the architecture as a whole can be altered, in hardware and software. Given that the system should maintain a low cost, the sensor selection for further improvement should be carefully addressed. However, sensors could be further optimised or increased in numbers in order to gain the ability to extract more features and break down the whole medical test into smaller segments. This can provide additional information for the test and the user and can increase the prediction and evaluation accuracy. For example, the same optimised sensors could be located in the middle of the 3 meters distance for the TUG test. A detailed mapping regarding the location of the new sensors should be taken under consideration. However, given that the sensors will not require continuous adjustments re-training requirements could be limited to the optimum minima.

The rapid development of the hardware in order to support Neural network ML by using Tensor Flow can offer an attractive approach especially with Microsoft's recent light-weight algorithms [317]. Although given that the model will be continuously retrained in order to offer an individualised approach for rehabilitation for the particular user the carbon footprint due to the power consumption will be increased in comparison with other lighter algorithmic approaches. The continuous power consumption through voltage transformation from domestic supply to 5V increases the carbon footprint and should be taken under consideration in future design approaches.

As part of the work on this thesis and although these have not been presented, initial experiments with NNs have been carried out. Keras was used to create a CNN with 6 hidden layer neurons in a Dense layer with activation function *relu* and followed by a second Dense layer with a *softmax* activation function. The *RMSProp* optimiser was used for the CNN and the models were optimised over 100 epochs. The model was not hyper-tuned and given that the initial results were significantly lower in terms of accuracy in comparison with other algorithmic approaches (namely KNN and XGboost), this approach was temporarily abandoned. Although the NN is a promising approach for the future, the level of interpretability due to a "black box" limitation is always questioned. However, as we mentioned above the rapid development of various hardware options with increased computational capacity and acceleration of NN algorithms in combination with new lighter algorithmic approaches could be an option for testing in the near future. However, there will still persist a question on how NNs can address the black box issue in terms of ART.

Although the proposed system in this thesis was used for stroke rehabilitation and detection of co-morbidities, it might offer a good approach on a variety of conditions and not only for unhealthy but for healthy subjects as well. The tests were selected given that there are data available in order to help with comparison and evaluation of the system itself and co-morbidities. However, the operation could be focused on particular daily activities in which the subject shows lack of ability or confidence in order to carry them out. Moreover, for healthy elder subjects the system will be able to extract patterns and help in terms of monitoring while at the same time can raise flags or warning to the rest of the family when the daily patterns have changed and this could be an indication of an early stage of an illness such as dementia, for example. Hence, the system could be optimised and tested in many ways and in various scenarios. Given that there is a good level of transferability it could be used on a variety of healthy and unhealthy subjects.

This thesis concluded by presenting the Hybrid ML model approach as well as the input and output data generated to support ART AI. In the future the user interface could be further developed. We aim to follow the dialogic XAI model presented in [234] in combination with state-of-the-art Human Computer Interaction (HCI) and cognitive theory approaches for ART AI as discussed in [318]. Moreover, we aim to address the miss-classification issues discussed in Chapter 5 by improving the presentation of both the TUG features and the generalised classes.

References

[1] "State of the Nation: stroke statistics," Feb. 2017. Accessed: Dec. 11, 2017. [Online]. Available: https://www.stroke.org.uk/resources/state-nation-stroke-statistics.

[2] G.-Miss. B. M. G. Dr.R.Y.Patil, Mrs.P.N.Jagtap, and Mr.V.B.Ingale, "Stroke a brain attack," *Stroke Brain Attack*, vol. 3, no. 8, pp. 01–23, Sep. 2013.

[3] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *The Lancet*, vol. 377, no. 9778, pp. 1693–1702, May 2011, doi: 10.1016/S0140-6736(11)60325-5.

[4] N. Kostanjsek, "Use of The International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and health information systems," *BMC Public Health*, vol. 11, no. 4, p. S3, May 2011, doi: 10.1186/1471-2458-11-S4-S3.

[5] "International Classification of Functioning, Disability and Health (ICF)." https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health (accessed Nov. 26, 2020).

[6] "WHO | International Classification of Functioning, Disability and Health (ICF)," *WHO*. http://www.who.int/classifications/icf/en/ (accessed Dec. 07, 2017).

[7] "SSNAP-Acute-Organisational-Audit-2014-Public-Repor.pdf." Accessed: Nov. 30, 2017. [Online]. Available: https://www.strokeaudit.org/Newspress/SSNAP-Acute-Organisational-Audit-2014-Public-Repor.pdf.

[8] Y. Wang, A. G. Rudd, and C. D. A. Wolfe, "Age and ethnic disparities in incidence of stroke over time: the South London Stroke Register," *Stroke*, vol. 44, no. 12, pp. 3298–3304, Dec. 2013, doi: 10.1161/STROKEAHA.113.002604.

[9] K. Ohene-Frempong *et al.*, "Cerebrovascular accidents in sickle cell disease: rates and risk factors," *Blood*, vol. 91, no. 1, pp. 288–294, Jan. 1998.

[10] S. Banerjee, R. Biram, J. Chataway, and D. Ames, "South Asian strokes: Lessons from the St Mary's stroke database," *QJM Mon. J. Assoc. Physicians*, vol. 103, pp. 17–21, Oct. 2009, doi: 10.1093/qjmed/hcp148.

[11] U. K. (mr) Web Master, "Data Access Request Service (DARS)," Jan. 22, 2015. http://content.digital.nhs.uk/DARS (accessed Nov. 30, 2017).

[12] W. Lo, "Childhood Hemorrhagic Stroke: An Important but Understudied Problem," *J. Child Neurol.*, vol. 26, no. 9, pp. 1174–1185, Sep. 2011, doi: 10.1177/0883073811408424.

[13] W. Lo, J. Stephens, and S. Fernandez, "Pediatric Stroke in the United States and the Impact of Risk Factors," *J. Child Neurol.*, vol. 24, no. 2, pp. 194–203, Feb. 2009, doi: 10.1177/0883073808322665.

[14] E. S. Roach *et al.*, "Management of Stroke in Infants and Children: A Scientific Statement From a Special Writing Group of the American Heart Association Stroke Council and the Council on Cardiovascular Disease in the Young," *Stroke*, vol. 39, no. 9, pp. 2644–2691, Sep. 2008, doi: 10.1161/STROKEAHA.108.189696.

[15] H. J. Fullerton *et al.*, "Infection, vaccination, and childhood arterial ischemic stroke," *Neurology*, vol. 85, no. 17, pp. 1459–1466, Oct. 2015, doi: 10.1212/WNL.0000000000002065.

[16] C. K. Fox, S. Sidney, and H. J. Fullerton, "A community-based, case-control study of childhood stroke risk associated with congenital heart disease," *Stroke J. Cereb. Circ.*, vol. 46, no. 2, pp. 336–340, Feb. 2015, doi: 10.1161/STROKEAHA.114.007218.

[17] J. Adamson, A. Beswick, and S. Ebrahim, "Is stroke the most common cause of disability?," *J. Stroke Cerebrovasc. Dis. Off. J. Natl. Stroke Assoc.*, vol. 13, no. 4, pp. 171–177, Aug. 2004, doi: 10.1016/j.jstrokecerebrovasdis.2004.06.003.

[18] S. T. Engelter *et al.*, “Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis,” *Stroke*, vol. 37, no. 6, pp. 1379–1384, Jun. 2006, doi: 10.1161/01.STR.0000221815.64093.8c.

[19] M. L. Berthier, “Poststroke aphasia : epidemiology, pathophysiology and treatment,” *Drugs Aging*, vol. 22, no. 2, pp. 163–182, 2005.

[20] L. Dickey, A. Kagan, M. P. Lindsay, J. Fang, A. Rowland, and S. Black, “Incidence and profile of inpatient stroke-induced aphasia in Ontario, Canada,” *Arch. Phys. Med. Rehabil.*, vol. 91, no. 2, pp. 196–202, Feb. 2010, doi: 10.1016/j.apmr.2009.09.020.

[21] “General Practice | Quality and Outcomes Framework (QOF) | Health Topics | ISD Scotland.” https://www.isdscotland.org/Health-Topics/General-Practice/Quality-And-Outcomes-Framework/ (accessed Dec. 14, 2020).

[22] V. L. Feigin *et al.*, “Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010,” *Lancet*, vol. 383, no. 9913, pp. 245–254, Jan. 2014.

[23] “Welsh Government | General medical services contract: Quality and outcomes framework.” Accessed: Dec. 04, 2017. [Online]. Available: http://gov.wales/statistics-and-research/general-medical-services-contract/?lang=en.

[24] “Stroke Happens: Stroke Evolution Days 2-5,” *studylib.net*. https://studylib.net/doc/8429450/stroke-happens--stroke-evolution-days-2-5 (accessed Dec. 14, 2020).

[25] DIPEx, “Stroke recovery: daily activities and personal care,” *healthtalk.org*, Jun. 2017. http://www.healthtalk.org/peoples-experiences/nerves-brain/stroke/stroke-recovery-daily-activities-and-personal-care (accessed Oct. 12, 2017).

[26] D. L. Rudman, D. Hebert, and D. Reid, “Living in a restricted occupational world: the occupational experiences of stroke survivors who are wheelchair users

and their caregivers," *Can. J. Occup. Ther. Rev. Can. Ergother.*, vol. 73, no. 3, pp. 141–152, Jun. 2006, doi: 10.2182/cjot.05.0014.

[27] "The Social Implications of Increases in Wheelchair Use. - Research Portal | Lancaster University." http://www.research.lancs.ac.uk/portal/en/publications/the-social-implications-of-increases-in-wheelchair-use(1178d8c3-3944-4623-9eea-7526676e98e9).html (accessed Nov. 26, 2020).

[28] A. Szeto, "5 - REHABILITATION ENGINEERING AND ASSISTIVE TECHNOLOGY," in *Introduction to Biomedical Engineering (Second Edition)*, J. D. Enderle, S. M. Blanchard, and J. D. Bronzino, Eds. Boston: Academic Press, 2005, pp. 211–254.

[29] D. J. Clarke, "The role of multidisciplinary team care in stroke rehabilitation," *Prog. Neurol. Psychiatry*, vol. 17, no. 4, pp. 5–8, Jul. 2013, doi: 10.1002/pnp.288.

[30] "National-Clinical-Guidelines-for-Stroke-fourth-edi.pdf." Accessed: Dec. 06, 2017. [Online]. Available: https://www.strokeaudit.org/Guideline/Historical-Guideline/National-Clinical-Guidelines-for-Stroke-fourth-edi.aspx.

[31] J. Bernhardt, M. N. Thuy, J. M. Collier, and L. A. Legg, "Very early versus delayed mobilisation after stroke," *Cochrane Database Syst. Rev.*, no. 1, p. CD006187, Jan. 2009, doi: 10.1002/14651858.CD006187.pub2.

[32] G. Kwakkel, B. Kollen, and E. Lindeman, "Understanding the pattern of functional recovery after stroke: facts and theories," *Restor. Neurol. Neurosci.*, vol. 22, no. 3–5, pp. 281–299, 2004.

[33] I. G. L. van de Port, S. Wood-Dauphinee, E. Lindeman, and G. Kwakkel, "Effects of exercise training programs on walking competency after stroke: a systematic review," *Am. J. Phys. Med. Rehabil.*, vol. 86, no. 11, pp. 935–951, Nov. 2007, doi: 10.1097/PHM.0b013e31802ee464.

[34] P. Govender and L. Kalra, “Benefits of occupational therapy in stroke rehabilitation,” *Expert Rev. Neurother.*, vol. 7, no. 8, pp. 1013–1019, Aug. 2007, doi: 10.1586/14737175.7.8.1013.

[35] European Stroke Organisation (ESO) Executive Committee and ESO Writing Committee, “Guidelines for management of ischaemic stroke and transient ischaemic attack 2008,” *Cerebrovasc. Dis. Basel Switz.*, vol. 25, no. 5, pp. 457–507, 2008, doi: 10.1159/000131083.

[36] M. Pickrell, B. Bongers, and E. van den Hoven, “Understanding Persuasion and Motivation in Interactive Stroke Rehabilitation,” in *Persuasive Technology*, Jun. 2015, pp. 15–26, doi: 10.1007/978-3-319-20306-5_2.

[37] T. Sugavanam, G. Mead, C. Bulley, M. Donaghy, and F. van Wijck, “The effects and experiences of goal setting in stroke rehabilitation - a systematic review,” *Disabil. Rehabil.*, vol. 35, no. 3, pp. 177–190, Feb. 2013, doi: 10.3109/09638288.2012.690501.

[38] S. Rosewilliam, C. A. Roskell, and A. Pandyan, “A systematic review and synthesis of the quantitative and qualitative evidence behind patient-centred goal setting in stroke rehabilitation,” *Clin. Rehabil.*, vol. 25, no. 6, pp. 501–514, Jun. 2011, doi: 10.1177/0269215510394467.

[39] I. Hartigan, “Goal setting in stroke rehabilitation: part 1,” *Br. J. Neurosci. Nurs.*, vol. 8, no. 2, pp. 65–69, Apr. 2012, doi: 10.12968/bjnn.2012.8.2.65.

[40] *Cambridge Handbook of Psychology, Health and Medicine*. 1970.

[41] G. Robinson-Smith and E. R. Pizzi, “Maximizing Stroke Recovery Using Patient Self-Care Self-Efficacy,” *Rehabil. Nurs.*, vol. 28, no. 2, pp. 48–51, Mar. 2003, doi: 10.1002/j.2048-7940.2003.tb02028.x.

[42] F. Jones and A. Riazi, “Self-efficacy and self-management after stroke: a systematic review,” *Disabil. Rehabil.*, vol. 33, no. 10, pp. 797–810, Jan. 2011, doi: 10.3109/09638288.2010.511415.

[43] S. Guthrie and A. Harvey, “Motivation and its influence on outcome in rehabilitation,” *Rev. Clin. Gerontol.*, vol. 4, no. 3, pp. 235–243, Aug. 1994, doi: 10.1017/S0959259800003865.

[44] N. Maclean, P. Pound, C. Wolfe, and A. Rudd, “Qualitative analysis of stroke patients’ motivation for rehabilitation,” *BMJ*, vol. 321, no. 7268, pp. 1051–1054, Oct. 2000, doi: 10.1136/bmj.321.7268.1051.

[45] N. Maclean, P. Pound, C. Wolfe, and A. Rudd, “The Concept of Patient Motivation: A Qualitative Analysis of Stroke Professionals’ Attitudes,” *Stroke*, vol. 33, no. 2, pp. 444–448, Feb. 2002, doi: 10.1161/hs0202.102367.

[46] H. I. Krebs, B. Volpe, and N. Hogan, “A working model of stroke recovery from rehabilitation robotics practitioners,” *J. Neuroengineering Rehabil.*, vol. 6, p. 6, Feb. 2009, doi: 10.1186/1743-0003-6-6.

[47] L. von Koch, A. W. Wottrich, and L. W. Holmqvist, “Rehabilitation in the home versus the hospital: the importance of context,” *Disabil. Rehabil.*, vol. 20, no. 10, pp. 367–372, Oct. 1998.

[48] “State of the Nation: stroke statistics,” Feb. 2018. Accessed: Jun. 11, 2018. [Online]. Available: https://www.stroke.org.uk/resources/state-nation-stroke-statistics.

[49] G. Saposnik *et al.*, “iPad technology for home rehabilitation after stroke (iHOME): a proof-of-concept randomized trial,” *Int. J. Stroke Off. J. Int. Stroke Soc.*, vol. 9, no. 7, pp. 956–962, Oct. 2014, doi: 10.1111/ijs.12328.

[50] M. Balaam *et al.*, “Motivating mobility: designing for lived motivation in stroke rehabilitation,” in *CHI 2011: conference proceedings and extended abstracts: 29th Annual Conference on Human Factors in Computing Systems, Vancouver, British Columbia, Canada, May 7-12, 2011*, D. Tan, B. Begole, and W. Kellog, Eds. New York: ACM Press, 2011, pp. 3073–3082.

[51] M. Baran *et al.*, “Design of a home-based adaptive mixed reality rehabilitation system for stroke survivors,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*

Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., vol. 2011, pp. 7602–7605, 2011, doi: 10.1109/IEMBS.2011.6091874.

[52] P. Chaiyawat, K. Kulkantrakorn, and P. Sritipsukho, "Effectiveness of home rehabilitation for ischemic stroke," *Neurol. Int.*, vol. 1, no. 1, Art. no. 1, Aug. 2009, doi: 10.4081/ni.2009.e10.

[53] "UCI stroke therapy study | UCI Health | Orange County, CA." http://www.ucihealth.org/news/2018/05/uci-stroke-therapy-study/ (accessed May 04, 2020).

[54] A. Pennycott, D. Wyss, H. Vallery, V. Klamroth-Marganska, and R. Riener, "Towards more effective robotic gait training for stroke rehabilitation: a review," *J. Neuroengineering Rehabil.*, vol. 9, p. 65, Sep. 2012, doi: 10.1186/1743-0003-9-65.

[55] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 81–94, Mar. 2016, doi: 10.1109/MSP.2015.2503881.

[56] I. Vourganas, V. Stankovic, L. Stankovic, and A. Kerr, "Factors That Contribute to the Use of Stroke Self-Rehabilitation Technologies: A Review," *JMIR Biomed. Eng.*, vol. 4, no. 1, p. e13732, 2019, doi: 10.2196/13732.

[57] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. NeuroEngineering Rehabil.*, vol. 9, p. 21, Apr. 2012, doi: 10.1186/1743-0003-9-21.

[58] J. A. S. Kelso, *Human Motor Behavior: An Introduction*. Psychology Press, 2014.

[59] E. S. Lawrence *et al.*, "Estimates of the Prevalence of Acute Stroke Impairments and Disability in a Multiethnic Population," *Stroke*, vol. 32, no. 6, pp. 1279–1284, Jun. 2001, doi: 10.1161/01.STR.32.6.1279.

[60] L. Brewer, F. Horgan, A. Hickey, and D. Williams, “Stroke rehabilitation: recent advances and future therapies,” *QJM Int. J. Med.*, vol. 106, no. 1, pp. 11–25, Jan. 2013, doi: 10.1093/qjmed/hcs174.

[61] A. A. A. Timmermans, H. A. M. Seelen, R. D. Willmann, and H. Kingma, “Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design,” *J. Neuroengineering Rehabil.*, vol. 6, p. 1, Jan. 2009, doi: 10.1186/1743-0003-6-1.

[62] J. L. Moore, E. J. Roth, C. Killian, and T. G. Hornby, “Locomotor training improves daily stepping activity and gait efficiency in individuals poststroke who have reached a ‘plateau’ in recovery,” *Stroke*, vol. 41, no. 1, pp. 129–135, Jan. 2010, doi: 10.1161/STROKEAHA.109.563247.

[63] T. G. Hornby, D. D. Campbell, J. H. Kahn, T. Demott, J. L. Moore, and H. R. Roth, “Enhanced Gait-Related Improvements After Therapist- Versus Robotic-Assisted Locomotor Training in Subjects With Chronic Stroke: A Randomized Controlled Study,” *Stroke*, vol. 39, no. 6, pp. 1786–1792, Jun. 2008, doi: 10.1161/STROKEAHA.107.504779.

[64] R. F. Macko *et al.*, “Treadmill aerobic exercise training reduces the energy expenditure and cardiovascular demands of hemiparetic gait in chronic stroke patients. A preliminary report,” *Stroke*, vol. 28, no. 2, pp. 326–330, Feb. 1997.

[65] H. Barbeau and M. Visintin, “Optimal outcomes obtained with body-Weight support combined with treadmill training in stroke subjects11No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.,” *Arch. Phys. Med. Rehabil.*, vol. 84, no. 10, pp. 1458–1465, Oct. 2003, doi: 10.1016/S0003-9993(03)00361-7.

[66] H. Schmidt, C. Werner, R. Bernhardt, S. Hesse, and J. Krüger, “Gait rehabilitation machines based on programmable footplates,” *J.*

NeuroEngineering Rehabil., vol. 4, p. 2, Feb. 2007, doi: 10.1186/1743-0003-4-2.

[67] "Walkbot," Nov. 2016. Accessed: Jan. 11, 2018. [Online]. Available: http://exoskeletonreport.com/product/walkbot/.

[68] S.-Y. Kim *et al.*, "Effects of Innovative WALKBOT Robotic-Assisted Locomotor Training on Balance and Gait Recovery in Hemiparetic Stroke: A Prospective, Randomized, Experimenter Blinded Case Control Study With a Four-Week Follow-Up," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 23, no. 4, pp. 636–642, Jul. 2015, doi: 10.1109/TNSRE.2015.2404936.

[69] C. K. Wong, K. Jordan, and M. King, "Robotic arm skate for stroke rehabilitation," in *2011 IEEE International Conference on Rehabilitation Robotics*, Jun. 2011, pp. 1–6, doi: 10.1109/ICORR.2011.5975389.

[70] R. C. V. Loureiro, W. S. Harwin, R. Lamperd, and C. Collin, "Evaluation of Reach and Grasp Robot-Assisted Therapy Suggests Similar Functional Recovery Patterns on Proximal and Distal Arm Segments in Sub-Acute Hemiplegia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 593–602, May 2014, doi: 10.1109/TNSRE.2013.2265263.

[71] C. Yang, A. Kerr, V. Stankovic, L. Stankovic, P. Rowe, and S. Cheng, "Human Upper Limb Motion Analysis for Post-Stroke Impairment Assessment Using Video Analytics," *IEEE Access*, vol. 4, pp. 650–659, 2016, doi: 10.1109/ACCESS.2016.2523803.

[72] S. R. Egglestone *et al.*, "A design framework for a home-based stroke rehabilitation system: Identifying the key components," in *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, Apr. 2009, pp. 1–8, doi: 10.4108/ICST.PERVASIVEHEALTH2009.6049.

[73] S. Saini, D. R. A. Rambli, S. Sulaiman, M. N. Zakaria, and S. R. M. Shukri, "A low-cost game framework for a home-based stroke rehabilitation system," in

2012 International Conference on Computer Information Science (ICCIS), Jun. 2012, vol. 1, pp. 55–60, doi: 10.1109/ICCISci.2012.6297212.

[74] S. Masiero *et al.*, *The value of robotic systems in stroke rehabilitation*, vol. 11. 2014.

[75] A. Koenig *et al.*, “Controlling patient participation during robot-assisted gait training,” *J. NeuroEngineering Rehabil.*, vol. 8, p. 14, Mar. 2011, doi: 10.1186/1743-0003-8-14.

[76] C. N. Schabowsky, S. B. Godfrey, R. J. Holley, and P. S. Lum, “Development and pilot testing of HEXORR: Hand EXOskeleton Rehabilitation Robot,” *J. NeuroEngineering Rehabil.*, vol. 7, p. 36, Jul. 2010, doi: 10.1186/1743-0003-7-36.

[77] R. Secoli, M.-H. Milot, G. Rosati, and D. J. Reinkensmeyer, “Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke,” *J. NeuroEngineering Rehabil.*, vol. 8, p. 21, Apr. 2011, doi: 10.1186/1743-0003-8-21.

[78] P. Staubli, T. Nef, V. Klamroth-Marganska, and R. Riener, “Effects of intensive arm training with the rehabilitation robot ARMin II in chronic stroke patients: four single-cases,” *J. NeuroEngineering Rehabil.*, vol. 6, p. 46, Dec. 2009, doi: 10.1186/1743-0003-6-46.

[79] J. Ziherl, D. Novak, A. Olenšek, M. Mihelj, and M. Munih, “Evaluation of upper extremity robot-assistances in subacute and chronic stroke subjects,” *J. NeuroEngineering Rehabil.*, vol. 7, p. 52, Oct. 2010, doi: 10.1186/1743-0003-7-52.

[80] P. Poli, G. Morone, G. Rosati, and S. Masiero, “Robotic Technologies and Rehabilitation: New Tools for Stroke Patients’ Therapy,” *BioMed Res. Int.*, vol. 2013, 2013, doi: 10.1155/2013/153872.

[81] I. Díaz, J. J. Gil, and E. Sánchez, “Lower-Limb Robotic Rehabilitation: Literature Review and Challenges,” *J. Robot.*, 2011, doi: 10.1155/2011/759764.

[82] O. M. Giggins, U. M. Persson, and B. Caulfield, “Biofeedback in rehabilitation,” *J. NeuroEngineering Rehabil.*, vol. 10, p. 60, Jun. 2013, doi: 10.1186/1743-0003-10-60.

[83] J.-H. Kim, “The effects of training using EMG biofeedback on stroke patients upper extremity functions,” *J. Phys. Ther. Sci.*, vol. 29, no. 6, pp. 1085–1088, Jun. 2017, doi: 10.1589/jpts.29.1085.

[84] “Home |,” *Home | Biomove by Curatronic Ltd.* https://www.biomove.com (accessed Jan. 15, 2018).

[85] R. Proffitt and B. Lange, “Considerations in the Efficacy and Effectiveness of Virtual Reality Interventions for Stroke Rehabilitation: Moving the Field Forward,” *Phys. Ther.*, vol. 95, no. 3, pp. 441–448, Mar. 2015, doi: 10.2522/ptj.20130571.

[86] A. Gaber, M. Taher, and M. Waheb, “A Comparison of Virtual Rehabilitation Techniques | Systems & Biomedical Engineering Department,” Barcelona Spain, 2015, pp. 1–6, Accessed: Jan. 18, 2018. [Online]. Available: http://bmes.cufe.edu.eg/content/comparison-virtual-rehabilitation-techniques.

[87] H. Mousavi Hondori and M. Khademi, “A Review on Technical and Clinical Impact of Microsoft Kinect on Physical Therapy and Rehabilitation,” *J. Med. Eng.*, vol. 2014, p. 846514, 2014, doi: 10.1155/2014/846514.

[88] F. Wittmann *et al.*, “Assessment-driven arm therapy at home using an IMU-based virtual reality system,” in *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*, Aug. 2015, pp. 707–712, doi: 10.1109/ICORR.2015.7281284.

[89] A. Slijper, K. E. Svensson, P. Backlund, H. Engström, and K. S. Sunnerhagen, “Computer game-based upper extremity training in the home environment in stroke persons: a single subject design,” *J. Neuroengineering Rehabil.*, vol. 11, p. 35, Mar. 2014, doi: 10.1186/1743-0003-11-35.

[90] R. V. Kenyon, J. Leigh, and E. A. Keshner, “Considerations for the future development of virtual technology as a rehabilitation tool,” *J. NeuroEngineering Rehabil.*, vol. 1, p. 13, Dec. 2004, doi: 10.1186/1743-0003-1-13.

[91] S. V. Adamovich, G. G. Fluet, A. Mathai, Q. Qiu, J. Lewis, and A. S. Merians, “Design of a complex virtual reality simulation to train finger motion for persons with hemiparesis: a proof of concept study,” *J. Neuroengineering Rehabil.*, vol. 6, p. 28, Jul. 2009, doi: 10.1186/1743-0003-6-28.

[92] J.-A. Gil-Gómez, R. Lloréns, M. Alcañiz, and C. Colomer, “Effectiveness of a Wii balance board-based system (eBaViR) for balance rehabilitation: a pilot randomized clinical trial in patients with acquired brain injury,” *J. NeuroEngineering Rehabil.*, vol. 8, p. 30, May 2011, doi: 10.1186/1743-0003-8-30.

[93] S. T. Smith, A. Talaei-Khoei, M. Ray, and P. Ray, “Electronic Games for Aged Care and Rehabilitation,” in *2009 11th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Dec. 2009, pp. 42–47, doi: 10.1109/HEALTH.2009.5406197.

[94] N. Lehrer, S. Attygalle, S. L. Wolf, and T. Rikakis, “Exploring the bases for a mixed reality stroke rehabilitation system, Part I: A unified approach for representing action, quantitative evaluation, and interactive feedback,” *J. NeuroEngineering Rehabil.*, vol. 8, p. 51, Aug. 2011, doi: 10.1186/1743-0003-8-51.

[95] S. M. Nijenhuis *et al.*, “Feasibility study into self-administered training at home using an arm and hand device with motivational gaming environment in chronic stroke,” *J. NeuroEngineering Rehabil.*, vol. 12, p. 89, Oct. 2015, doi: 10.1186/s12984-015-0080-y.

[96] J. C. Stewart *et al.*, “Intervention to enhance skilled arm and hand movements after stroke: A feasibility study using a new virtual reality system,” *J. NeuroEngineering Rehabil.*, vol. 4, p. 21, Jun. 2007, doi: 10.1186/1743-0003-4-21.

[97] M. Gorsic and D. Novak, "Design and pilot evaluation of competitive and cooperative exercise games for arm rehabilitation at home," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2016, pp. 4690–4694, 2016, doi: 10.1109/EMBC.2016.7591774.

[98] R. Colombo *et al.*, "Design strategies to improve patient motivation during robot-aided rehabilitation," *J. NeuroEngineering Rehabil.*, vol. 4, p. 3, Feb. 2007, doi: 10.1186/1743-0003-4-3.

[99] B. R. Webster, P. A. Celnik, and L. G. Cohen, "Noninvasive brain stimulation in stroke rehabilitation," *NeuroRX*, vol. 3, no. 4, pp. 474–481, Oct. 2006, doi: 10.1016/j.nurx.2006.07.008.

[100] F. Moniche *et al.*, "Clinical Studies of Bone Marrow-Derived Stem Cell Therapy in Stroke Patients," in *Bone marrow stem cell therapy for stroke*, Springer, Singapore, 2017, pp. 241–259.

[101] M. Invernizzi, S. Negrini, S. Carda, L. Lanzotti, C. Cisari, and A. Baricich, "The value of adding mirror therapy for upper limb motor recovery of subacute stroke patients: a randomized controlled trial," *Eur. J. Phys. Rehabil. Med.*, vol. 49, no. 3, pp. 311–317, Jun. 2013.

[102] "Canadian Partnership for Stroke Recovery." http://www.canadianstroke.ca/en/ (accessed Jan. 16, 2018).

[103] Y. Zhang, S. Guo, G. Cao, S. Zhang, and Y. Liu, "A novel variable stiffness actuator-based exoskeleton device for home rehabilitation," in *2016 IEEE International Conference on Mechatronics and Automation*, Aug. 2016, pp. 878–883, doi: 10.1109/ICMA.2016.7558678.

[104] F. Amirabdollahian *et al.*, "Design, development and deployment of a hand/wrist exoskeleton for home-based rehabilitation after stroke - SCRIPT project," *Robotica*, vol. 32, no. 8, pp. 1331–1346, Dec. 2014, doi: 10.1017/S0263574714002288.

[105] H. Yano, N. Tanaka, K. Kamibayashi, H. Saitou, and H. Iwata, “Development of a Portable Gait Rehabilitation System for Home-Visit Rehabilitation,” *Sci. World J.*, 2015, doi: 10.1155/2015/849831.

[106] S. Mohamaddan *et al.*, “Development of Upper Limb Rehabilitation Robot Device for Home Setting,” *Procedia Comput. Sci.*, vol. 76, pp. 376–380, Jan. 2015, doi: 10.1016/j.procs.2015.12.312.

[107] M. Díaz, J. Saez-Pons, M. Heerink, and C. Angulo, “Emotional factors in robot-based assistive services for elderly at home,” in *2013 IEEE RO-MAN*, Aug. 2013, pp. 711–716, doi: 10.1109/ROMAN.2013.6628396.

[108] M. White, M. Vining Radomski, M. Finkelstein, D. A. S. Nilsson, E. Oddsson, and L. Ingimar, “Assistive/Socially Assistive Robotic Platform for Therapy and Recovery: Patient Perspectives,” *Int. J. Telemed. Appl.*, 2013, doi: 10.1155/2013/948087.

[109] M. Sivan *et al.*, “Home-based Computer Assisted Arm Rehabilitation (hCAAR) robotic device for upper limb exercise after stroke: results of a feasibility study in home setting,” *J. NeuroEngineering Rehabil.*, vol. 11, p. 163, Dec. 2014, doi: 10.1186/1743-0003-11-163.

[110] M. J. Johnson, X. Feng, L. M. Johnson, and J. M. Winters, “Potential of a suite of robot/computer-assisted motivating systems for personalized, home-based, stroke rehabilitation,” *J. NeuroEngineering Rehabil.*, vol. 4, p. 6, Mar. 2007, doi: 10.1186/1743-0003-4-6.

[111] A. González, P. Fraisse, and M. Hayashibe, “A personalized balance measurement for home-based rehabilitation,” in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, Apr. 2015, pp. 711–714, doi: 10.1109/NER.2015.7146722.

[112] J.-P. Wang, L.-Q. Guo, T.-Y. Sheng, and D.-X. Xiong, “Automated Brunnstrom Assessment for Home Rehabilitation Based on GRNN Model,” *ITM Web Conf.*, vol. 12, p. 01021, 2017, doi: 10.1051/itmconf/20171201021.

[113] B.-C. Lee, J. Kim, S. Chen, and K. H. Sienko, “Cell phone based balance trainer,” *J. Neuroengineering Rehabil.*, vol. 9, p. 10, Feb. 2012, doi: 10.1186/1743-0003-9-10.

[114] F. Kohler, T. Schmitz-Rode, and C. Disselhorst-Klug, “Introducing a feedback training system for guided home rehabilitation,” *J. NeuroEngineering Rehabil.*, vol. 7, p. 2, Jan. 2010, doi: 10.1186/1743-0003-7-2.

[115] N. Friedman, V. Chan, D. Zondervan, M. Bachman, and D. J. Reinkensmeyer, “MusicGlove: motivating and quantifying hand movement rehabilitation by using functional grips to play music,” *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 2359–2363, 2011, doi: 10.1109/IEMBS.2011.6090659.

[116] T. Nimmy John and D. Hepsiba, “Rehabilitation Exercise Monitoring System,” *Int J Adv Res Electron Commun Eng*, vol. 2, no. 3, pp. 389–391, 2013.

[117] R. E. Jorge, L. Acion, D. Moser, H. P. Adams, and R. G. Robinson, “Escitalopram and Enhancement of Cognitive Recovery Following Stroke,” *Arch. Gen. Psychiatry*, vol. 67, no. 2, pp. 187–196, Feb. 2010, doi: 10.1001/archgenpsychiatry.2009.185.

[118] T. Hoffmann, S. Bennett, C.-L. Koh, and K. McKenna, “A Systematic Review of Cognitive Interventions to Improve Functional Ability in People Who Have Cognitive Impairment Following Stroke,” *Top. Stroke Rehabil.*, vol. 17, no. 2, pp. 99–107, Mar. 2010, doi: 10.1310/tsr1702-99.

[119] T. S. Sathyanarayana Rao and C. Andrade, “Exploring new frontiers in neuropsychopharmacology: SSRIs for stroke,” *Indian J. Psychiatry*, vol. 53, no. 4, pp. 283–286, 2011, doi: 10.4103/0019-5545.91899.

[120] M. C. Brady, H. Kelly, J. Godwin, and P. Enderby, “Speech and language therapy for aphasia following stroke,” *Cochrane Database Syst. Rev.*, no. 5, p. CD000425, May 2012, doi: 10.1002/14651858.CD000425.pub3.

[121] Dr Zisis Papanikolaou, “Speed and strength training with plyometrcis.” Dr Zisis Papanikolaou, [Online]. Available: http://www.pe.uth.gr/cms/phocadownload/epeaek/anaptiksi_taxythtas_MK1012/dialekseis/8.pdf.

[122] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, “The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance,” *Scand. J. Rehabil. Med.*, vol. 7, no. 1, pp. 13–31, 1975.

[123] R. Teasell and N. H. M. MRehabMed, “4b. Rehab of Hemiplegic Upper Extremity Post Stroke,” p. 22, 2016.

[124] “The Brunnstrom Stages of Stroke Recovery,” *Saebo*, Sep. 09, 2015. https://www.saebo.com/the-stages-of-stroke-recovery/ (accessed Mar. 23, 2018).

[125] G. Marchionini, “Exploratory Search: From Finding to Understanding,” *Commun ACM*, vol. 49, no. 4, pp. 41–46, Apr. 2006, doi: 10.1145/1121949.1121979.

[126] G. Wulf, “Self-controlled practice enhances motor learning: implications for physiotherapy,” *Physiotherapy*, vol. 93, no. 2, pp. 96–101, Jun. 2007, doi: 10.1016/j.physio.2006.08.005.

[127] N. Yozbatiran, L. Der-Yeghiaian, and S. C. Cramer, “A Standardized Approach to Performing the Action Research Arm Test,” *Neurorehabil. Neural Repair*, vol. 22, no. 1, pp. 78–90, Jan. 2008, doi: 10.1177/1545968307305353.

[128] M. Penta, L. Tesio, C. Arnould, A. Zancan, and J. L. Thonnard, “The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment,” *Stroke*, vol. 32, no. 7, pp. 1627–1634, Jul. 2001, doi: 10.1161/01.str.32.7.1627.

[129] “Mirror Therapy - upper extremity,” *Stroke Engine*. https://www.strokengine.ca/en/intervention/mirror-therapy/ (accessed Apr. 28, 2020).

[130] M. Ye, C. Yang, V. Stankovic, L. Stankovic, and A. Kerr, “A Depth Camera Motion Analysis Framework for Tele-rehabilitation: Motion Capture and Person-Centric Kinematics Analysis,” *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 5, pp. 877–887, Aug. 2016, doi: 10.1109/JSTSP.2016.2559446.

[131] C. Yang *et al.*, “Autonomous Gait Event Detection with Portable Single-Camera Gait Kinematics Analysis System,” *J. Sens.*, vol. 2016, p. 5036857, Jan. 2016, doi: 10.1155/2016/5036857.

[132] D. Peters, R. A. Calvo, and R. M. Ryan, “Designing for Motivation, Engagement and Wellbeing in Digital Experience,” *Front. Psychol.*, vol. 9, 2018, doi: 10.3389/fpsyg.2018.00797.

[133] M. A. Pérez-Quiñones and J. L. Sibert, “A collaborative model of feedback in human-computer interaction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, Canada, Apr. 1996, pp. 316–323, doi: 10.1145/238386.238535.

[134] K. He, D. Jakovetic, B. Zhao, V. Stankovic, L. Stankovic, and S. Cheng, “A Generic Optimisation-Based Approach for Improving Non-Intrusive Load Monitoring,” *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6472–6480, Nov. 2019, doi: 10.1109/TSG.2019.2906012.

[135] Michael Fell, Harry Kennard, Gesche Huebner, and Moira Nicolson, Simon Elam, David Shipworth, “Energising Health: A review of the health and care applications of smart meter data,” May 2017, [Online]. Available: https://www.smartenergygb.org/en/resources/press-centre/press-releases-folder/energising-healthcare.

[136] A. Yassine, S. Singh, and A. Alamri, “Mining Human Activity Patterns From Smart Home Big Data for Health Care Applications,” *IEEE Access*, vol. 5, pp. 13131–13141, 2017, doi: 10.1109/ACCESS.2017.2719921.

[137] S. Samarah, M. G. A. Zamil, A. F. Aleroud, M. Rawashdeh, M. F. Alhamid, and A. Alamri, “An Efficient Activity Recognition Framework: Toward Privacy-

Sensitive Health Data Sensing," *IEEE Access*, vol. 5, pp. 3848–3859, 2017, doi: 10.1109/ACCESS.2017.2685531.

[138] Q. Ni, A. B. García Hernando, and I. P. de la Cruz, "The Elderly's Independent Living in Smart Homes: A Characterization of Activities and Sensing Infrastructure Survey to Facilitate Services Development," *Sensors*, vol. 15, no. 5, pp. 11312–11362, May 2015, doi: 10.3390/s150511312.

[139] E. Ebeid, R. Heick, and R. H. Jacobsen, "Deducing Energy Consumer Behavior from Smart Meter Data," *Future Internet*, vol. 9, no. 3, p. 29, Jul. 2017, doi: 10.3390/fi9030029.

[140] K. Gajowniczek and T. Ząbkowski, "Data Mining Techniques for Detecting Household Characteristics Based on Smart Meter Data," *Energies*, vol. 8, no. 7, pp. 7407–7427, Jul. 2015, doi: 10.3390/en8077407.

[141] F. N. Melzi, M. H. Zayani, A. B. Hamida, A. Same, and L. Oukhellou, "Identifying Daily Electric Consumption Patterns from Smart Meter Data by Means of Clustering Algorithms," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2015, pp. 1136–1141, doi: 10.1109/ICMLA.2015.18.

[142] A. Al-Wakeel, J. Wu, and N. Jenkins, "k-means based load estimation of domestic smart meter measurements," *Appl. Energy*, vol. 194, pp. 333–342, May 2017, doi: 10.1016/j.apenergy.2016.06.046.

[143] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Trans. Smart Grid*, pp. 1–1, 2018, doi: 10.1109/TSG.2018.2818167.

[144] "3rings," *3rings*. https://www.3rings.co.uk/ (accessed Apr. 03, 2018).

[145] "Envirotxt." https://tekview-solutions.com/envirotxt.php (accessed Apr. 03, 2018).

[146] “Kemuri Smart Power Sockets.” http://www.kemurisense.com/index.html (accessed Apr. 03, 2018).

[147] “CurrentCare - Welcome.” http://currentcare.uk.com/#system (accessed Apr. 03, 2018).

[148] “Howz Smart Home.” https://www.howz.com/ (accessed Apr. 03, 2018).

[149] “Smart meter tracks when the kettle’s on to check grandpa’s OK | New Scientist.” https://www.newscientist.com/article/2121709-smart-meter-tracks-when-the-kettles-on-to-check-grandpas-ok/ (accessed Apr. 03, 2018).

[150] F. Cavallo, M. Aquilano, L. Odetti, M. Arvati, and M. C. Carrozza, “A first step toward a pervasive and smart ZigBee sensor system for assistance and rehabilitation,” in *2009 IEEE International Conference on Rehabilitation Robotics*, Jun. 2009, pp. 632–637, doi: 10.1109/ICORR.2009.5209471.

[151] “Guideline for the prevention of falls in older persons. American Geriatrics Society, British Geriatrics Society, and American Academy of Orthopaedic Surgeons Panel on Falls Prevention,” *J. Am. Geriatr. Soc.*, vol. 49, no. 5, pp. 664–672, May 2001.

[152] G. Sprint, D. Cook, and D. Weeks, “Towards Automating Clinical Assessments: A Survey of the Timed Up and Go (TUG),” *IEEE Rev. Biomed. Eng.*, vol. 8, pp. 64–77, 2015, doi: 10.1109/RBME.2015.2390646.

[153] C. U. Persson, A. Danielsson, K. S. Sunnerhagen, A. Grimby-Ekman, and P.-O. Hansson, “Timed Up & Go as a measure for longitudinal change in mobility after stroke - Postural Stroke Study in Gothenburg (POSTGOT),” *J. Neuroengineering Rehabil.*, vol. 11, p. 83, May 2014, doi: 10.1186/1743-0003-11-83.

[154] A. Dubois, T. Bihl, and J.-P. Bresciani, “Automating the Timed Up and Go Test Using a Depth Camera,” *Sensors*, vol. 18, no. 1, Dec. 2017, doi: 10.3390/s18010014.

[155] N. Kitsunezaki, E. Adachi, T. Masuda, and J. Mizusawa, "KINECT applications for the physical rehabilitation," in *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2013, pp. 294–299, doi: 10.1109/MeMeA.2013.6549755.

[156] M. H. M. Chan, D. T. F. Keung, S. Y. T. Lui, and R. T. H. Cheung, "A validation study of a smartphone application for functional mobility assessment of the elderly," *Hong Kong Physiother. J.*, vol. 35, pp. 1–4, Dec. 2016, doi: 10.1016/j.hkpj.2015.11.001.

[157] R. C. van Lummel, S. Walgaard, A. B. Maier, E. Ainsworth, P. J. Beek, and J. H. van Dieën, "The Instrumented Sit-to-Stand Test (iSTS) Has Greater Clinical Relevance than the Manually Recorded Sit-to-Stand Test in Older Adults," *PLOS ONE*, vol. 11, no. 7, p. e0157968, Jul. 2016, doi: 10.1371/journal.pone.0157968.

[158] J. A. Merchán-Baeza, M. González-Sánchez, and A. Cuesta-Vargas, "Mobile Functional Reach Test in People Who Suffer Stroke: A Pilot Study," *JMIR Rehabil. Assist. Technol.*, vol. 2, no. 1, p. e6, Jun. 2015, doi: 10.2196/rehab.4102.

[159] L. Blum and N. Korner-Bitensky, "Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review," *Phys. Ther.*, vol. 88, no. 5, pp. 559–566, May 2008, doi: 10.2522/ptj.20070205.

[160] pubmeddev and A. S. al et, "Validity of the performance-oriented mobility assessment in predicting fall of stroke survivors: a retrospective cohort study. - PubMed - NCBI." https://www.ncbi.nlm.nih.gov/pubmed/24850058 (accessed Apr. 29, 2020).

[161] B. Chinsongkram, N. Chaikeeree, V. Saengsirisuwan, F. B. Horak, and R. Boonsinsukh, "Responsiveness of the Balance Evaluation Systems Test (BESTest) in People With Subacute Stroke," *Phys. Ther.*, vol. 96, no. 10, pp. 1638–1647, Oct. 2016, doi: 10.2522/ptj.20150621.

[162] "The sensitivity and specificity of the Falls Efficacy Scale and the Activities-specific Balance Confidence Scale for hemiplegic stroke patients. - PubMed -

NCBI.” https://www.ncbi.nlm.nih.gov/pubmed/29950756 (accessed Apr. 29, 2020).

[163] G. L. Iverson and M. S. Koehle, “Normative data for the balance error scoring system in adults,” *Rehabil. Res. Pract.*, vol. 2013, p. 846418, 2013, doi: 10.1155/2013/846418.

[164] Y. Mong, T. W. Teo, and S. S. Ng, “5-repetition sit-to-stand test in subjects with chronic stroke: reliability and validity,” *Arch. Phys. Med. Rehabil.*, vol. 91, no. 3, pp. 407–413, Mar. 2010, doi: 10.1016/j.apmr.2009.10.030.

[165] NHS, “NHS VIDEOS | Risk of falls: the timed get up and go test - NHS,” Apr. 10, 2018. https://www.nhs.uk/Video/Pages/timed-get-up-and-go-test.aspx (accessed Apr. 28, 2020).

[166] C. L. Lysack, “Chapter 23 - Household and Neighborhood Safety, Mobility,” in *Handbook of Assessment in Clinical Gerontology (Second Edition)*, P. A. Lichtenberg, Ed. San Diego: Academic Press, 2010, pp. 619–646.

[167] A. Ibrahim, D. K. A. Singh, and S. Shahar, “‘Timed Up and Go’ test: Age, gender and cognitive impairment stratified normative values of older adults,” *PLOS ONE*, vol. 12, no. 10, p. e0185641, Oct. 2017, doi: 10.1371/journal.pone.0185641.

[168] T. A. de Melo, A. C. M. Duarte, T. S. Bezerra, F. França, N. S. Soares, and D. Brito, “The Five Times Sit-to-Stand Test: safety and reliability with older intensive care unit patients at discharge,” *Rev. Bras. Ter. Intensiva*, vol. 31, no. 1, pp. 27–33, 2019, doi: 10.5935/0103-507X.20190006.

[169] B. Pieper, T. N. Templin, and A. Goldberg, “A comparative study of the five-times-sit-to-stand and timed-up-and-go tests as measures of functional mobility in persons with and without injection-related venous ulcers,” *Adv. Skin Wound Care*, vol. 27, no. 2, pp. 82–92; quiz 93–94, Feb. 2014, doi: 10.1097/01.ASW.0000442876.94332.26.

[170] W. K. Durfee, L. Savard, and S. Weinstein, "Technical Feasibility of Teleassessments for Rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 23–29, Mar. 2007, doi: 10.1109/TNSRE.2007.891400.

[171] D. Berrada, M. Romero, G. Abowd, M. Blount, and J. Davis, "Automatic administration of the get up and go test," in *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*, San Juan, Puerto Rico, Jun. 2007, pp. 73–75, doi: 10.1145/1248054.1248075.

[172] Z. Skrba *et al.*, "Objective real-time assessment of walking and turning in elderly adults," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2009, pp. 807–810, 2009, doi: 10.1109/IEMBS.2009.5333934.

[173] F. Wang, M. Skubic, C. Abbott, and J. M. Keller, "Quantitative analysis of 180 degree turns for fall risk assessment using video sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 7606–7609, doi: 10.1109/IEMBS.2011.6092005.

[174] A. Ghose, P. Sinha, C. Bhaumik, A. Sinha, A. Agrawal, and A. Dutta Choudhury, "UbiHeld: ubiquitous healthcare monitoring system for elderly and chronic patients," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, Zurich, Switzerland, Sep. 2013, pp. 1255–1264, doi: 10.1145/2494091.2497331.

[175] O. Lohmann, T. Luhmann, and A. Hein, "Skeleton Timed Up and Go," in *2012 IEEE International Conference on Bioinformatics and Biomedicine*, Oct. 2012, pp. 1–5, doi: 10.1109/BIBM.2012.6392610.

[176] A. Salarian, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian, "iTUG, a Sensitive and Reliable Measure of Mobility," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 3, pp. 303–310, Jun. 2010, doi: 10.1109/TNSRE.2010.2047606.

[177] M. R. Narayanan, S. R. Lord, M. M. Budge, B. G. Celler, and N. H. Lovell, “Falls Management: Detection and Prevention, using a Waist-mounted Triaxial Accelerometer,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2007, pp. 4037–4040, doi: 10.1109/IEMBS.2007.4353219.

[178] M. R. Narayanan, M. E. Scalzi, S. J. Redmond, S. R. Lord, B. G. Celler, and N. H. Lovell, “A wearable triaxial accelerometry system for longitudinal assessment of falls risk,” in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, pp. 2840–2843, doi: 10.1109/IEMBS.2008.4649794.

[179] M. R. Narayanan, S. J. Redmond, M. E. Scalzi, S. R. Lord, B. G. Celler, and N. H. Lovell *, “Longitudinal Falls-Risk Estimation Using Triaxial Accelerometry,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 534–541, Mar. 2010, doi: 10.1109/TBME.2009.2033038.

[180] Barry. R. Greene, A. O’Donovan, R. Romero-Ortuno, L. Cogan, C. N. Scanaill, and R. A. Kenny, “Quantitative Falls Risk Assessment Using the Timed Up and Go Test,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2918–2926, Dec. 2010, doi: 10.1109/TBME.2010.2083659.

[181] Y. Higashi, K. Yamakoshi, T. Fujimoto, M. Sekine, and T. Tamura, “Quantitative evaluation of movement using the timed up-and-go test,” *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 4, pp. 38–46, Jul. 2008, doi: 10.1109/MEMB.2008.919494.

[182] S. Gillain *et al.*, “The value of instrumental gait analysis in elderly healthy, MCI or Alzheimer’s disease subjects and a comparison with other clinical tests used in single and dual-task conditions,” *Ann. Phys. Rehabil. Med.*, vol. 52, no. 6, pp. 453–474, Jul. 2009, doi: 10.1016/j.rehab.2008.10.004.

[183] C. Tacconi, S. Mellone, and L. Chiari, “Smartphone-based applications for investigating falls and mobility,” in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and*

Workshops, May 2011, pp. 258–261, doi: 10.4108/icst.pervasivehealth.2011.246060.

[184] S. Mellone, C. Tacconi, and L. Chiari, “Suitability of a Smartphone accelerometer to instrument the Timed Up and Go test: A preliminary study,” *Gait Posture*, vol. Supplement 1, no. 33, pp. S50–S51, 2011, doi: 10.1016/j.gaitpost.2010.10.060.

[185] L. Palmerini, S. Mellone, L. Rocchi, and L. Chiari, “Dimensionality reduction for the quantitative evaluation of a smartphone-based Timed Up and Go test,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 7179–7182, doi: 10.1109/IEMBS.2011.6091814.

[186] S. Mellone, C. Tacconi, L. Schwickert, J. Klenk, C. Becker, and L. Chiari, “Smartphone-based solutions for fall detection and prevention: the FARSEEING approach,” *Z. Für Gerontol. Geriatr.*, vol. 45, no. 8, pp. 722–727, Dec. 2012, doi: 10.1007/s00391-012-0404-5.

[187] J. Fontecha, Fco. J. Navarro, R. Hervás, and J. Bravo, “Elderly frailty detection by using accelerometer-enabled smartphones and clinical information records,” *Pers. Ubiquitous Comput.*, vol. 17, no. 6, pp. 1073–1083, Aug. 2013, doi: 10.1007/s00779-012-0559-5.

[188] M. Milosevic, E. Jovanov, and A. Milenković, “Quantifying Timed-Up-and-Go test: A smartphone implementation,” in *2013 IEEE International Conference on Body Sensor Networks*, May 2013, pp. 1–6, doi: 10.1109/BSN.2013.6575478.

[189] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, “A Survey on Ambient Intelligence in Healthcare,” *Proc. IEEE*, vol. 101, no. 12, pp. 2470–2494, Dec. 2013, doi: 10.1109/JPROC.2013.2262913.

[190] D. Austin, T. L. Hayes, J. Kaye, N. Mattek, and M. Pavel, “Unobtrusive monitoring of the longitudinal evolution of in-home gait velocity data with applications to elder care,” in *2011 Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society, Aug. 2011, pp. 6495–6498, doi: 10.1109/IEMBS.2011.6091603.

[191] S. Hagler, D. Austin, T. L. Hayes, J. Kaye, and M. Pavel, "Unobtrusive and Ubiquitous In-Home Monitoring: A Methodology for Continuous Assessment of Gait Velocity in Elders," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 813–820, Apr. 2010, doi: 10.1109/TBME.2009.2036732.

[192] T. Frenken, B. Vester, M. Brell, and A. Hein, "aTUG: Fully-automated timed up and go assessment using ambient sensor technologies," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, May 2011, pp. 55–62, doi: 10.4108/icst.pervasivehealth.2011.245985.

[193] G. R. H. Regterschot, M. Folkersma, W. Zhang, H. Baldus, M. Stevens, and W. Zijlstra, "Sensitivity of sensor-based sit-to-stand peak power to the effects of training leg strength, leg power and balance in older adults," *Gait Posture*, vol. 39, no. 1, pp. 303–307, Jan. 2014, doi: 10.1016/j.gaitpost.2013.07.122.

[194] B. R. Greene, E. P. Doheny, R. A. Kenny, and B. Caulfield, "Classification of frailty and falls history using a combination of sensor-based mobility assessments," *Physiol. Meas.*, vol. 35, no. 10, pp. 2053–2066, Sep. 2014, doi: 10.1088/0967-3334/35/10/2053.

[195] N. Millor, P. Lecumberri, M. Gómez, A. Martínez-Ramírez, and M. Izquierdo, "An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit," *J. NeuroEngineering Rehabil.*, vol. 10, no. 1, p. 86, Aug. 2013, doi: 10.1186/1743-0003-10-86.

[196] A. Kanai *et al.*, "Use of the sit-to-stand task to evaluate motor function of older adults using telemetry," *BMC Geriatr.*, vol. 16, no. 1, p. 121, Jun. 2016, doi: 10.1186/s12877-016-0294-2.

[197] A. Ejupi, M. Brodie, Y. J. Gschwind, S. R. Lord, W. L. Zagler, and K. Delbaere, "Kinect-Based Five-Times-Sit-to-Stand Test for Clinical and In-Home

Assessment of Fall Risk in Older People," *Gerontology*, vol. 62, no. 1, pp. 118–124, 2016, doi: 10.1159/000381804.

[198] U. Lindemann, R. Muche, M. Stuber, W. Zijlstra, K. Hauer, and C. Becker, "Coordination of Strength Exertion During the Chair-Rise Movement in Very Old People," *J. Gerontol. Ser. A*, vol. 62, no. 6, pp. 636–640, Jun. 2007, doi: 10.1093/gerona/62.6.636.

[199] A. Arcelus, R. Goubran, H. Sveistrup, M. Bilodeau, and F. Knoefel, "Context-aware smart home monitoring through pressure measurement sequences," in *2010 IEEE International Workshop on Medical Measurements and Applications*, Apr. 2010, pp. 32–37, doi: 10.1109/MEMEA.2010.5480223.

[200] D. K. Dey, E. Rothenberg, V. Sundh, I. Bosaeus, and B. Steen, "Height and body weight in elderly adults: a 21-year population study on secular trends and related factors in 70-year-olds," *J. Gerontol. A. Biol. Sci. Med. Sci.*, vol. 56, no. 12, pp. M780-784, Dec. 2001, doi: 10.1093/gerona/56.12.m780.

[201] "Five Times Sit to Stand Test," *Shirley Ryan AbilityLab*. https://www.sralab.org/rehabilitation-measures/five-times-sit-stand-test (accessed Apr. 30, 2020).

[202] B. Carstensen, J. Simpson, and L. C. Gurrin, "Statistical models for assessing agreement in method comparison studies with replicate measurements," *Int. J. Biostat.*, vol. 4, no. 1, p. Article 16, 2008, doi: 10.2202/1557-4679.1107.

[203] L. A. Critchley and J. A. Critchley, "A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques," *J. Clin. Monit. Comput.*, vol. 15, no. 2, pp. 85–91, Feb. 1999, doi: 10.1023/a:1009982611386.

[204] J. L. Carrasco, B. R. Phillips, J. Puig-Martinez, T. S. King, and V. M. Chinchilli, "Estimation of the concordance correlation coefficient for repeated measures using SAS and R," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 293–304, Mar. 2013, doi: 10.1016/j.cmpb.2012.09.002.

[205] P. E. Shrout and J. L. Fleiss, “Intraclass correlations: uses in assessing rater reliability,” *Psychol. Bull.*, vol. 86, no. 2, pp. 420–428, Mar. 1979, doi: 10.1037//0033-2909.86.2.420.

[206] Kalra, “Decoding the Bland–Altman plot: Basic review.” http://www.j-pcs.org/article.asp?issn=2395-5414;year=2017;volume=3;issue=1;spage=36;epage=38;aulast=Kalra (accessed May 01, 2020).

[207] H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, and Q. Yang, “Building ethics into artificial intelligence,” in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, Jul. 2018, pp. 5527–5533, Accessed: Sep. 12, 2020. [Online].

[208] S. Taylor *et al.*, “Responsible AI – Key themes, concerns & recommendations for European research and innovation,” Jul. 02, 2018. https://eprints.soton.ac.uk/426307/ (accessed Sep. 11, 2020).

[209] S. Larsson and F. Heintz, “Transparency in artificial intelligence,” *Internet Policy Rev.*, vol. 9, no. 2, May 2020, Accessed: Sep. 12, 2020. [Online]. Available: https://policyreview.info/concepts/transparency-artificial-intelligence.

[210] R. Iphofen and M. Kritikos, “Regulating artificial intelligence and robotics: ethics by design in a digital society,” *Contemp. Soc. Sci.*, pp. 1–15, Jan. 2019, doi: 10.1080/21582041.2018.1563803.

[211] L. Abrardi, C. Cambini, and L. Rondi, “The Economics of Artificial Intelligence: A Survey,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3425922, Jul. 2019. doi: 10.2139/ssrn.3425922.

[212] M. Coeckelbergh, “Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability,” *Sci. Eng. Ethics*, vol. 26, no. 4, pp. 2051–2068, Aug. 2020, doi: 10.1007/s11948-019-00146-8.

[213] I. Vourganas, V. Stankovic, L. Stankovic, and A. L. Michala, “Evaluation of home-based rehabilitation sensing systems with respect to standardised clinical tests,” *Sensors*, vol. 20, no. 1, p. 26, 2020.

[214] admin, “Briefing: Health and Care of Older People in England 2017,” *A Healthier Future*, Feb. 20, 2017. https://www.healthierfuture.org.uk/publications/2017/february/briefing-health-and-care-older-people-england-2017 (accessed Dec. 02, 2020).

[215] N. Sambasivan and J. Holbrook, “Toward responsible AI for the next billion users.,” *interactions*, vol. 26, no. 1, pp. 68–71, 2019.

[216] A. A. Tubella, A. Theodorou, V. Dignum, and F. Dignum, “Governance by glass-box: implementing transparent moral bounds for AI behaviour,” *ArXiv Prepr. ArXiv190504994*, 2019.

[217] J. Bryson and A. Winfield, “Standardizing ethical design for artificial intelligence and autonomous systems,” *Computer*, vol. 50, no. 5, pp. 116–119, 2017.

[218] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck, “Fair, transparent, and accountable algorithmic decision-making processes,” *Philos. Technol.*, vol. 31, no. 4, pp. 611–627, 2018.

[219] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” *ArXiv Prepr. ArXiv170208608*, 2017.

[220] V. Dignum, “Responsible artificial intelligence: designing AI for human values,” *ITU J. ICT Discov.*, vol. 25, no. 1, 2017.

[221] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, 2019.

[222] A. Choudhary, *Decoding the Black Box: An Important Introduction to Interpretable Machine Learning Models in Python*. 2019.

[223] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A Survey on Bias and Fairness in Machine Learning,” *ArXiv190809635 Cs*, Sep. 2019, Accessed: Sep. 14, 2020. [Online]. Available: http://arxiv.org/abs/1908.09635.

[224] D. Gunning and D. W. Aha, “DARPA’s Explainable Artificial Intelligence Program,” *AI Mag.*, vol. 40, no. 2, pp. 44–58, 2019.

[225] The Centre for Data Ethics and Innovation, “Review into bias in algorithmic decision-making. Interim review.),” UK Government, 2019. [Online]. Available: https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation/interim-report-review-into-bias-in-algorithmic-decision-making.

[226] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[227] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, “DeepFool: a simple and accurate method to fool deep neural networks. CoRR abs/1511.04599 (2015),” *ArXiv Prepr. ArXiv151104599*, 2015.

[228] D. Leslie, “Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector,” *Alan Turing Inst.*, vol. 6, 2019.

[229] A. B. Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[230] N. Vollmer, “Article 25 EU General Data Protection Regulation (EU-GDPR),” *Retrieved January*, vol. 15, p. 2018, 2017.

[231] A. Rockall, “From hype to hope to hard work: developing responsible AI for radiology,” *Clin. Radiol.*, 2019.

[232] E. Tjoa and C. Guan, “A survey on explainable artificial intelligence (XAI): towards medical XAI,” *ArXiv Prepr. ArXiv190707374*, 2019.

[233] O. Lahav, N. Mastronarde, and M. van der Schaar, “What is interpretable? Using machine learning to design interpretable decision-support systems,” *ArXiv Prepr. ArXiv181110799*, 2018.

[234] J. Cassens and R. Wegener, “Ambient Explanations: Ambient Intelligence and Explainable AI,” in *European Conference on Ambient Intelligence*, 2019, pp. 370–376.

[235] J. C. Augusto, “Ambient intelligence: opportunities and consequences of its use in smart Classrooms,” *Innov. Teach. Learn. Inf. Comput. Sci.*, vol. 8, no. 2, pp. 53–63, Jun. 2009, doi: 10.11120/ital.2009.08020053.

[236] X. Chen, Q. Shi, L. Yang, and J. Xu, “ThriftyEdge: Resource-Efficient Edge Computing for Intelligent IoT Applications,” *IEEE Netw.*, vol. 32, no. 1, pp. 61–65, Jan. 2018, doi: 10.1109/MNET.2018.1700145.

[237] O. Pauly, “Random forests for medical applications,” PhD Thesis, Technische Universität München, 2012.

[238] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, “Decision Trees: An Overview and Their Use in Medicine,” *J. Med. Syst.*, vol. 26, no. 5, pp. 445–463, Oct. 2002, doi: 10.1023/A:1016409317640.

[239] P. Kazienko, E. Lughofer, and B. Trawinski, “Editorial on the special issue ‘Hybrid and ensemble techniques in soft computing: recent advances and emerging trends,’” *Soft Comput.*, vol. 19, no. 12, pp. 3353–3355, Dec. 2015, doi: 10.1007/s00500-015-1916-x.

[240] P.-Y. Yin, C.-C. Tsai, R.-F. Day, C.-Y. Tung, and B. Bhanu, “Ensemble learning of model hyperparameters and spatiotemporal data for calibration of low-cost PM2. 5 sensors.,” *Math. Biosci. Eng. MBE*, vol. 16, no. 6, pp. 6858–6873, 2019.

[241] A. Stojić *et al.*, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition," *Sci. Total Environ.*, vol. 653, pp. 140–147, 2019.

[242] M. V. Shenoy, "Sensor Information Processing for Wearable IoT Devices," in *Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm*, Springer, 2020, pp. 177–200.

[243] D. J. Norris, "Predictions using CNNs and MLPs for medical research," in *Machine Learning with the Raspberry Pi*, Springer, 2020, pp. 453–500.

[244] J. Aveleira-Mata and H. Alaiz-Moreton, "Functional Prototype for Intrusion Detection System Oriented to Intelligent IoT Models," in *International Symposium on Ambient Intelligence*, 2019, pp. 179–186.

[245] T. Chen, "Introduction to boosted trees," *Univ. Wash. Comput. Sci.*, vol. 22, p. 115, 2014.

[246] N. Sharma, Anju, and A. Juneja, "Extreme Gradient Boosting with Squared Logistic Loss Function," in *Machine Intelligence and Signal Analysis*, Singapore, 2019, pp. 313–322, doi: 10.1007/978-981-13-0923-6_27.

[247] L. Feng, Y. Li, Y. Wang, and Q. Du, "Estimating hourly and continuous ground-level PM2. 5 concentrations using an ensemble learning algorithm: The ST-stacking model," *Atmos. Environ.*, vol. 223, p. 117242, 2020.

[248] H. K. Jabbar and R. Khan, "Methods to Avoid Over-Fitting and Under-Fitting in Supervised Machine Learning (Comparative Study)," 2014, doi: 10.3850/978-981-09-5247-1_017.

[249] K. Baumann, "Cross-validation as the objective function for variable-selection techniques," *TrAC Trends Anal. Chem.*, vol. 22, no. 6, pp. 395–406, Jun. 2003, doi: 10.1016/S0165-9936(03)00607-1.

[250] S. Basu, J. H. Faghmous, and P. Doupe, "Machine Learning Methods for Precision Medicine Research Designed to Reduce Health Disparities: A

Structured Tutorial," *Ethn. Dis.*, vol. 30, no. Suppl 1, Art. no. Suppl 1, Apr. 2020, doi: 10.18865/ed.30.S1.217.

[251] G. C. Cawley and N. L. C. Talbot, "Preventing Over-Fitting during Model Selection via Bayesian Regularisation of the Hyper-Parameters," *J. Mach. Learn. Res.*, vol. 8, no. Apr, pp. 841–861, 2007.

[252] J. L. Andrews, "Addressing overfitting and underfitting in Gaussian model-based clustering," *Comput. Stat. Data Anal.*, vol. 127, pp. 160–171, Nov. 2018, doi: 10.1016/j.csda.2018.05.015.

[253] V. Dignum, "Ethics in artificial intelligence: introduction to the special issue," *Ethics Inf. Technol.*, vol. 20, no. 1, pp. 1–3, Mar. 2018, doi: 10.1007/s10676-018-9450-z.

[254] N. Laranjeiro, S. N. Soydemir, and J. Bernardino, "A survey on data quality: classifying poor data," in *2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC)*, 2015, pp. 179–188.

[255] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *J. Manag. Inf. Syst.*, vol. 12, no. 4, pp. 5–33, Mar. 1996, doi: 10.1080/07421222.1996.11518099.

[256] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," in *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 191–198.

[257] J. Grus, *Data science from scratch: first principles with python*. O'Reilly Media, 2019.

[258] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[259] J. Jordon, J. Yoon, and M. van der Schaar, "Measuring the quality of Synthetic data for use in competitions," *ArXiv180611345 Cs Stat*, Jun. 2018, Accessed: Dec. 02, 2020. [Online]. Available: http://arxiv.org/abs/1806.11345.

[260] A. Goncalves, P. Ray, B. Soper, J. Stevens, L. Coyle, and A. P. Sales, "Generation and evaluation of synthetic patient data," *BMC Med. Res. Methodol.*, vol. 20, no. 1, p. 108, May 2020, doi: 10.1186/s12874-020-00977-1.

[261] A. R. Benaim *et al.*, "Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies," *JMIR Med. Inform.*, vol. 8, no. 2, p. e16492, 2020, doi: 10.2196/16492.

[262] "Timed Up and Go," *Shirley Ryan AbilityLab*. https://www.sralab.org/rehabilitation-measures/timed-and-go (accessed Sep. 25, 2020).

[263] M. Boonstra, M. D. W. Malefijt, and N. Verdonschot, "How to quantify knee function after total knee arthroplasty?," *The Knee*, vol. 15, no. 5, pp. 390–395, 2008.

[264] D. Brooks, A. M. Davis, and G. Naglie, "Validity of 3 physical performance measures in inpatient geriatric rehabilitation," *Arch. Phys. Med. Rehabil.*, vol. 87, no. 1, pp. 105–110, 2006.

[265] K. J. Brusse, S. Zimdars, K. R. Zalewski, and T. M. Steffen, "Testing functional performance in people with Parkinson disease," *Phys. Ther.*, vol. 85, no. 2, pp. 134–141, 2005.

[266] H. Chen, X. Zheng, H. Huang, C. Liu, Q. Wan, and S. Shang, "The effects of a home-based exercise intervention on elderly patients with knee osteoarthritis: a quasi-experimental study," *BMC Musculoskelet. Disord.*, vol. 20, no. 1, p. 160, 2019.

[267] V. Dal Bello-Haas, L. Klassen, M. S. Sheppard, and A. Metcalfe, "Psychometric properties of activity, self-efficacy, and quality-of-life measures

in individuals with Parkinson disease," *Physiother. Can.*, vol. 63, no. 1, pp. 47–57, 2011.

[268] U. B. Flansbjer, A. M. Holmbäck, D. Downham, C. Patten, and J. Lexell, "Reliability of gait performance tests in men and women with hemiparesis after stroke," *J. Rehabil. Med.*, vol. 37, no. 2, pp. 75–82, 2005.

[269] K. B. Foreman, O. Addison, H. S. Kim, and L. E. Dibble, "Testing balance and fall risk in persons with Parkinson disease, an argument for ecologically valid testing," *Parkinsonism Relat. Disord.*, vol. 17, no. 3, pp. 166–171, 2011.

[270] K. M. Gill-Body, M. Beninato, and D. E. Krebs, "Relationship among balance impairments, functional performance, and disability in people with peripheral vestibular hypofunction," *Phys. Ther.*, vol. 80, no. 8, pp. 748–758, 2000.

[271] M. Katz-Leurer, H. Rotem, H. Lewitus, O. Keren, and S. Meyer, "Functional balance tests for children with traumatic brain injury: within-session reliability," *Pediatr. Phys. Ther.*, vol. 20, no. 3, pp. 254–258, 2008.

[272] D. M. Kennedy, P. W. Stratford, J. Wessel, J. D. Gollish, and D. Penney, "Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty," *BMC Musculoskelet. Disord.*, vol. 6, no. 1, p. 3, 2005.

[273] S. Knorr, B. Brouwer, and S. J. Garland, "Validity of the Community Balance and Mobility Scale in community-dwelling persons after stroke," *Arch. Phys. Med. Rehabil.*, vol. 91, no. 6, pp. 890–896, 2010.

[274] J. Lemay and S. Nadeau, "Standing balance assessment in ASIA D paraplegic and tetraplegic participants: concurrent validity of the Berg Balance Scale," *Spinal Cord*, vol. 48, no. 3, pp. 245–250, 2010.

[275] M. Maly, P. Costigan, and S. Olney, "Contribution of psychosocial and mechanical variables to mobility outcome measures in knee OA," *Phys Ther*, vol. 85, pp. 1318–1328, 2005.

[276] J. D. Ries, J. L. Echternach, L. Nof, and M. Gagnon Blodgett, "Test-retest reliability and minimal detectable change scores for the timed 'up & go' test, the six-minute walk test, and gait speed in people with Alzheimer disease," *Phys. Ther.*, vol. 89, no. 6, pp. 569–579, 2009.

[277] F. Sabirli, N. Paker, and D. Bugdayci, "The relationship between Knee Injury and Osteoarthritis Outcome Score (KOOS) and Timed Up and Go test in patients with symptomatic knee osteoarthritis," *Rheumatol. Int.*, vol. 33, no. 10, pp. 2691–2694, 2013.

[278] T. M. Steffen, T. A. Hacker, and L. Mollinger, "Age-and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds," *Phys. Ther.*, vol. 82, no. 2, pp. 128–137, 2002.

[279] A. A. Wright, C. E. Cook, G. D. Baxter, J. D. Dockerty, and J. H. Abbott, "A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis," *J. Orthop. Sports Phys. Ther.*, vol. 41, no. 5, pp. 319–327, 2011.

[280] M. Beninato, L. G. Portney, and P. E. Sullivan, "Using the International Classification of Functioning, Disability and Health as a framework to examine the association between falls and clinical assessment tools in people with stroke," *Phys. Ther.*, vol. 89, no. 8, pp. 816–825, 2009.

[281] R. W. Bohannon, M. E. Shove, S. R. Barreca, L. M. Masters, and C. S. Sigouin, "Five-repetition sit-to-stand test performance by community-dwelling adults: A preliminary investigation of times, determinants, and relationship with self-reported physical performance," *Isokinet. Exerc. Sci.*, vol. 15, no. 2, pp. 77–81, 2007.

[282] C. L. Christiansen and J. E. Stevens-Lapsley, "Weight-bearing asymmetry in relation to measures of impairment and functional mobility for people with knee osteoarthritis," *Arch. Phys. Med. Rehabil.*, vol. 91, no. 10, pp. 1524–1528, 2010.

[283] R. P. Duncan, A. L. Leddy, and G. M. Earhart, “Five times sit-to-stand test performance in Parkinson’s disease,” *Arch. Phys. Med. Rehabil.*, vol. 92, no. 9, pp. 1431–1436, 2011.

[284] Y.-C. Lin, R. C. Davey, and T. Cochrane, “Tests for physical function of the elderly with knee and hip osteoarthritis,” *Scand. J. Med. Sci. Sports*, vol. 11, no. 5, pp. 280–286, 2001.

[285] S. R. Lord, S. M. Murray, K. Chapman, B. Munro, and A. Tiedemann, “Sit-to-Stand Performance Depends on Sensation, Speed, Balance, and Psychological Status in Addition to Strength in Older People,” *J. Gerontol. Ser. A*, vol. 57, no. 8, pp. M539–M543, 2002, doi: 10.1093/gerona/57.8.M539.

[286] S. S. Paul, C. G. Canning, C. Sherrington, and V. S. Fung, “Reproducibility of measures of leg muscle power, leg muscle strength, postural sway and mobility in people with Parkinson’s disease,” *Gait Posture*, vol. 36, no. 3, pp. 639–642, 2012.

[287] S. R. Piva *et al.*, “Contribution of hip abductor strength to physical function in patients with total knee arthroplasty,” *Phys. Ther.*, vol. 91, no. 2, pp. 225–233, 2011.

[288] A. Tiedemann, H. Shimada, C. Sherrington, S. Murray, and S. Lord, “The comparative ability of eight functional mobility tests for predicting falls in community-dwelling older people,” *Age Ageing*, vol. 37, no. 4, pp. 430–435, 2008.

[289] S. L. Whitney, D. M. Wrisley, G. F. Marchetti, M. A. Gee, M. S. Redfern, and J. M. Furman, “Clinical measurement of sit-to-stand performance in people with balance disorders: validity of data for the Five-Times-Sit-to-Stand Test,” *Phys. Ther.*, vol. 85, no. 10, pp. 1034–1045, 2005.

[290] R. W. Bohannon, “Reference Values for the Five-Repetition Sit-to-Stand Test: A Descriptive Meta-Analysis of Data from Elders,” *Percept. Mot. Skills*, vol. 103, no. 1, pp. 215–222, 2006, doi: 10.2466/pms.103.1.215-222.

[291] M. Brandkvist, J. H. Bjørngaard, R. A. Ødeg\aard, B. O. \AAsvold, E. R. Sund, and G. \AAberge Vie, "Quantifying the impact of genes on body mass index during the obesity epidemic: longitudinal findings from the HUNT Study," *bmj*, vol. 366, p. l4067, 2019.

[292] C. Fryar, D. Kruszon-Moran, Q. Gu, and C. Ogden, "Mean body weight, height, waist circumference, and body mass index among adults: United States, 1999–2000 through 2015–2016 (National Health Statistics Reports No. 122)," *Retrieved Natl. Cent. Health Stat. Httpswww Cdc Govnchsdatanhsrnhsr122-508 Pdf*, 2018.

[293] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[294] S. M. Afifi, H. Gholamhosseini, and S. Poopak, "Hardware Implementations of SVM on FPGA: A State-of-the-Art Review of Current Practice," vol. 2, no. 11, pp. 733–752, Nov. 2015.

[295] Y. Zhang, "Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM," *Chem. Eng. Sci.*, vol. 64, no. 5, pp. 801–811, Mar. 2009, doi: 10.1016/j.ces.2008.10.012.

[296] M. Praveena and V. Jaiganesh, "A Literature Review on Supervised Machine Learning Algorithms and Boosting Process," 2017, doi: 10.5120/IJCA2017914816.

[297] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[298] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[299] Y.-P. Huang and M.-F. Yen, “A new perspective of performance comparison among machine learning algorithms for financial distress prediction,” *Appl. Soft Comput.*, vol. 83, p. 105663, Oct. 2019, doi: 10.1016/j.asoc.2019.105663.

[300] J. Fan *et al.*, “Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China,” *Energy Convers. Manag.*, vol. 164, pp. 102–111, May 2018, doi: 10.1016/j.enconman.2018.02.087.

[301] Y. Lee, D. Han, M.-H. Ahn, J. Im, and S. J. Lee, “Retrieval of Total Precipitable Water from Himawari-8 AHI Data: A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network,” *Remote Sens.*, vol. 11, no. 15, Art. no. 15, Jan. 2019, doi: 10.3390/rs11151741.

[302] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, Jun. 2006, pp. 161–168, doi: 10.1145/1143844.1143865.

[303] Y. Xing, J. Weng, Y. Wang, L. Sui, Y. Shan, and Y. Wang, “An In-depth Comparison of Compilers for Deep Neural Networks on Hardware,” *2019 IEEE Int. Conf. Embed. Softw. Syst. ICESS*, 2019, doi: 10.1109/ICESS.2019.8782480.

[304] N. Memon, S. B. Patel, and D. P. Patel, “Comparative Analysis of Artificial Neural Network and XGBoost Algorithm for PolSAR Image Classification,” in *Pattern Recognition and Machine Intelligence*, Dec. 2019, pp. 452–460, doi: 10.1007/978-3-030-34869-4_49.

[305] R. A. Stein, P. A. Jaques, and J. F. Valiati, “An Analysis of Hierarchical Text Classification Using Word Embeddings,” *Inf. Sci.*, vol. 471, pp. 216–232, Jan. 2019, doi: 10.1016/j.ins.2018.09.001.

[306] D. Chakraborty and H. Elzarka, “Advanced machine learning techniques for building performance simulation: a comparative analysis,” *J. Build. Perform.*

Simul., vol. 12, no. 2, pp. 193–207, Mar. 2019, doi: 10.1080/19401493.2018.1498538.

[307] A. G. Jivani, “The Novel k Nearest Neighbor Algorithm,” in *2013 International Conference on Computer Communication and Informatics*, Jan. 2013, pp. 1–4, doi: 10.1109/ICCCI.2013.6466287.

[308] R. A. Ashley and C. F. Parmeter, “Sensitivity analysis of an OLS multiple regression inference with respect to possible linear endogeneity in the explanatory variables, for both modest and for extremely large samples,” *Econometrics*, vol. 8, no. 1, p. 11, Mar. 2020, doi: 10.3390/econometrics8010011.

[309] F. Kane, *Hands-on data science and python machine learning*. Packt Publishing Ltd, 2017.

[310] T. Wong and P. Yeh, “Reliable Accuracy Estimates from k-Fold Cross Validation,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, Aug. 2020, doi: 10.1109/TKDE.2019.2912815.

[311] M. R. Haley, “K-fold cross validation performance comparisons of six naive portfolio selection rules: how naive can you be and still have successful out-of-sample portfolio performance?,” *Ann. Finance*, vol. 13, no. 3, pp. 341–353, 2017.

[312] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[313] A. Rao, F. Palaci, and W. Chow, “A practical guide to Responsible Artificial Intelligence (AI),” PwC, 2019. [Online]. Available: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf.

[314] E. Acorn, N. Dipsis, T. Pincus, and K. Stathis, “Sit-to-Stand Movement Recognition Using Kinect,” in *Statistical Learning and Data Sciences*, Cham, 2015, pp. 179–192, doi: 10.1007/978-3-319-17091-6_13.

[315] S. Hellmers *et al.*, “Measurement of the Chair Rise Performance of Older People Based on Force Plates and IMUs,” *Sensors*, vol. 19, no. 6, Art. no. 6, Jan. 2019, doi: 10.3390/s19061370.

[316] J. Silva, “Comparing machine learning approaches for fall risk assessment,” 2017, vol. 5, Accessed: Apr. 10, 2021. [Online]. Available: https://scholar.google.com/scholar_lookup?title=Comparing+machine+learning+approaches+for+fall+risk+assessment&conference=Proceedings+of+the+10th+International+Conference+on+Bio-inspired+Systems+and+Signal+Processing&author=Silva,+J.&author=Madureira,+J.&author=Tonelo,+C.&author=Baltazar,+D.&author=Silva,+C.&author=Martins,+A.&author=Alcobia,+C.&author=Sousa,+I.&publication_year=2017&pages=223%E2%80%93230#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3An1dq_p_4WA0J%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den.

[317] D. V. Gruzenkin *et al.*, “Neural networks to solve modern artificial intelligence tasks,” *J. Phys. Conf. Ser.*, vol. 1399, p. 033058, Dec. 2019, doi: 10.1088/1742-6596/1399/3/033058.

[318] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, “Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda,” in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–18.

Appendix

Scanned copies of questionnaires responses are presented in this Appendix section.

Experiment 1

User questionnaire

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?						✓
Was the feedback sufficient?						✓
Would you use the device again?						✓
Was the device engaging?						✓
Did the device increase your motivation for performing the task ?						✓

Any other thoughts??

Thank you

Experiment 2.

User questionnaire

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?						✓
Was the feedback sufficient?						✓
Would you use the device again?					✓	
Was the device engaging?						✓
Did the device increase your motivation for performing the task ?						✓

Any other thoughts??

Thank you

User questionnaire

Experiment 3

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?					α	
Was the feedback sufficient?						α
Would you use the device again?					α	
Was the device engaging?				α		
Did the device increase your motivation for performing the task ?					α	

Any other thoughts??

Thank you

Experiment 4

User questionnaire

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?						✓
Was the feedback sufficient?		✓				
Would you use the device again?					✓	
Was the device engaging?					✓	
Did the device increase your motivation for performing the task ?				✓		

Any other thoughts??

Thank you

Experiment 5

User questionnaire

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?						✓
Was the feedback sufficient?						✓
Would you use the device again?						✓
Was the device engaging?						✓
Did the device increase your motivation for performing the task ?						✓

Any other thoughts??

Thank you

Experiment 6

User questionnaire

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?						✓
Was the feedback sufficient?						✓
Would you use the device again?						✓
Was the device engaging?						✓
Did the device increase your motivation for performing the task ?						✓

Any other thoughts??

Thank you

User questionnaire

Experimen 7.

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?					✓	
Was the feedback sufficient?					✓	
Would you use the device again?				✓		
Was the device engaging?				✓		
Did the device increase your motivation for performing the task ?				✓		

Any other thoughts??

Thank you

Experiment 8

User questionnaire

Please answer the following questions with score from 0 – 5, with 0 very poor and 5 excellent

Question	score					
	0	1	2	3	4	5
Was the device easy to use and set up?						✓
Was the feedback sufficient?						✓
Would you use the device again?						✓
Was the device engaging?						✓
Did the device increase your motivation for performing the task ?						✓

Any other thoughts??

Thank you