

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Machine Learning and Movement Analysis for Cognitive Screening in Older Adults

Vânia Margarida Cardoso Guimarães



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Doctoral Program in Biomedical Engineering

Supervisor: Doctor Miguel Fernando Paiva Velhote Correia

Co-Supervisor: Doctor Inês Nunes de Sousa Soares

October, 2023

Machine Learning and Movement Analysis for Cognitive Screening in Older Adults

Vânia Margarida Cardoso Guimarães

Doctoral Program in Biomedical Engineering

Approved by:

President: Doctor Joaquim Gabriel Magalhães Mendes, University of Porto

Referee: Doctor Manuel Montero-Odasso, University of Western Ontario

Referee: Doctor Cristina Manuela Peixoto dos Santos, University of Minho

Referee: Doctor Jaime dos Santos Cardoso, University of Porto

Supervisor: Doctor Miguel Fernando Paiva Velhote Correia, University of Porto

(Doctor Miguel Fernando Paiva Velhote Correia)

October, 2023

Abstract

As the population ages, the prevalence of cognitive deficits linked to neurocognitive disorders such as Alzheimer’s disease becomes increasingly concerning. The lack of effective treatments amplifies the need to promptly identify and address the condition at its earliest stages, aiming to slow down or prevent the progression to dementia. Once dementia is established, it compromises individuals’ daily functioning and independence. Mild cognitive impairment (MCI) is the first detectable stage of a neurocognitive disorder. However, its symptoms are subtle and often go unnoticed, frequently mistaken for normal age-related cognitive decline.

Current paper-based cognitive tests are inadequate for wide population screening due to their time-consuming nature and the need for evaluation by specialists. Alternatively, recent studies suggest that motor decline can be used as an indicator of cognitive decline. Older adults with MCI tend to walk slower, especially when performing a cognitive task simultaneously (dual-task). These studies reveal a connection between motor and cognitive functions, highlighting the importance of cognition, particularly of executive functions, for effective motor control. Wearable technology, such as inertial sensors, offers a cost-effective and portable alternative for objectively assessing movement parameters associated with MCI, applicable in both clinical and daily life contexts. Yet, this field of study is still in its infancy and requires additional research.

This work aims to explore the use of motion analysis and machine learning techniques to improve the early detection of MCI resorting to wearable technology, including inertial sensors. To do so, we propose to investigate different task performance assessments, including gait tests, dual-task tests, and interaction with games.

Our results show that by using inertial sensors positioned on the feet and employing advanced motion analysis techniques, clinically relevant gait parameters can be measured accurately. Using deep learning techniques allowed to estimate foot trajectories in each stride, providing additional metrics. These approaches exhibit robust performance irrespective of the sensor’s orientation on the foot, hence supporting their application in real-world scenarios. Moreover, inertial sensors on the feet enable the estimation of direction and extraction of various step characteristics during interactive gameplay. The proposed analysis methods compensate for sensor drift over time and allow the analysis of data from both feet without discrimination.

The analysis of game interaction metrics reveals associations with standard measures of cognition and mobility in older adults. These games, originally designed to stimulate the mobility and cognition of older adults, can provide valuable insights into the motor-cognitive state of the players, enabling more frequent, accessible, and enjoyable assessments. The analysis of gait using the developed algorithms proves relevant in evaluating gait metrics affected by cognitive impairment. For instance, the introduction of a cognitive task during gait performance affects both older adults with and without cognitive impairment, with a more pronounced effect observed in older adults with cognitive impairment. Metrics for dual-task assessment and dual-task effect, including some tested for the first time in this study, reveal significant differences between groups and exhibit associations with cognitive test scores.

Machine learning methods, particularly one-class classification with confound control, are effective in detecting MCI using motor, cognitive, and dual tasks. Through this approach, only the data from the majority class (in this case, the control group) are used to train the classifiers, while the remaining data are considered anomalies. These methods outperform traditional binary classification in terms of sensitivity for MCI detection. By combining models trained on different data sources, the performance surpasses that of feature combination and outperforms paper-based screening methods. Additionally, gait and dual-task analysis combined with machine learning methods show promise in identifying cognitive deficits in older adults with cerebrovascular disease.

Overall, this work reinforces the potential of motor tasks in evaluating cognitive deficits in older adults using inertial sensors. The proposed tools for evaluating movement and dual tasks are simpler, faster, and more objective than paper-based screening tests, offering a potential alternative or complement to existing methods. To validate the effectiveness and feasibility of the proposed solutions, additional tests involving a larger sample size are necessary. In the future, these assessments have the potential to enhance the screening for cognitive impairment, contributing to a better understanding of motor-cognitive interactions, the development of novel therapies, and improved disease management.

Keywords: Cognitive Impairment, Cognitive Screening, Dual Tasks, Exergames, Gait Analysis, Inertial Sensors, Machine Learning, Movement Analysis, Older Adults

Resumo

À medida que a população envelhece, a prevalência do déficit cognitivo associado a doenças neurocognitivas, tais como a doença de Alzheimer, torna-se cada vez mais preocupante. A falta de tratamentos eficazes amplifica a necessidade de identificar e intervir precocemente nas fases iniciais da doença, visando retardar ou prevenir a progressão para demência. Uma vez estabelecida, a demência compromete o desempenho e a independência das pessoas nas suas atividades diárias. O déficit cognitivo ligeiro (DCL) é o primeiro estágio detetável de um distúrbio neurocognitivo. No entanto, os seus sintomas são subtis e passam muitas vezes despercebidos, sendo frequentemente confundidos com o declínio cognitivo normal associado à idade.

Os testes cognitivos em papel utilizados nos dias de hoje são inadequados para o rastreio da população em larga escala devido ao seu carácter demorado e à necessidade de avaliação por especialistas. Em alternativa, estudos recentes sugerem que o declínio motor pode ser um indicador de declínio cognitivo. Os idosos com DCL tendem a caminhar mais devagar, especialmente quando realizam uma tarefa cognitiva em simultâneo (dupla-tarefa). Estes estudos revelam uma ligação entre as funções motoras e cognitivas, destacando a importância da cognição, especialmente das funções executivas, para um controlo motor eficaz. Tecnologia vestível, tais como os sensores inerciais, oferecem uma alternativa económica e portátil para avaliar objetivamente parâmetros de movimento associados ao DCL, aplicáveis tanto em contextos clínicos quanto no quotidiano. No entanto, esta área de estudo está ainda em estágios iniciais e requer investigação adicional.

Este trabalho tem como objetivo explorar o uso de técnicas de análise de movimento e aprendizagem automática para melhorar a deteção precoce do DCL com recurso a tecnologia vestível, incluindo sensores inerciais. Para isso, propomos investigar diferentes avaliações de desempenho de tarefas, incluindo testes de marcha, testes de dupla-tarefa e interação com jogos.

Os nossos resultados mostram que, ao utilizar sensores inerciais colocados nos pés e empregar técnicas avançadas de análise de movimento, parâmetros da marcha clinicamente relevantes podem ser medidos com precisão. Técnicas de aprendizagem automática profunda (deep learning) permitem estimar as trajetórias dos pés em cada passada, fornecendo métricas adicionais. Estas abordagens operam de forma robusta qualquer que seja a orientação com que o sensor é colocado no pé, sustentando assim a sua aplicação em cenários reais. Além disso, os sensores inerciais nos pés permitem estimar a direção e extrair várias características dos passos durante a interação com jogos. Os métodos de análise propostos compensam a deriva do sensor ao longo do tempo e permitem a análise dos dados de ambos os pés de forma indiscriminada.

A análise das métricas de interação com jogos revela associações com medidas-padrão de cognição e mobilidade nos idosos. Estes jogos, originalmente projetados para estimular a mobilidade e a cognição dos idosos, podem fornecer informações valiosas sobre o estado motor-cognitivo dos jogadores, tornando assim as avaliações mais frequentes, acessíveis e agradáveis. A análise da marcha usando os algoritmos desenvolvidos mostra-se relevante para avaliar métricas da marcha afetadas pelo déficit cognitivo. Por exemplo, a introdução de uma tarefa cognitiva durante a execução da marcha afeta idosos com e sem déficit cognitivo, sendo que este efeito é mais pronunciado

nos idosos com défice cognitivo. Métricas para avaliação de dupla-tarefa e efeito da dupla-tarefa, algumas destas testadas pela primeira vez neste estudo, revelam diferenças significativas entre os grupos e exibem associações com as pontuações dos testes cognitivos.

Métodos de aprendizagem automática, particularmente classificação de uma classe com controlo de variáveis de confusão, demonstram eficácia na deteção do DCL usando tarefas motoras, cognitivas e dupla-tarefa. Através desta abordagem, apenas os dados da classe maioritária (neste caso, o grupo de controlo) são usados para treinar os classificadores, enquanto que os restantes dados são considerados anomalias. Estes métodos superam a classificação binária tradicional no que diz respeito à sensibilidade para a deteção do DCL. Ao combinar modelos treinados em diferentes fontes de dados, o desempenho é superior à combinação de características e supera os métodos de rastreio baseados em papel. Para além disso, a análise da marcha e da dupla-tarefa combinada com métodos de aprendizagem automática mostram promessa na identificação do défice cognitivo em idosos com doença cerebrovascular.

No geral, este trabalho reforça o potencial das tarefas motoras na avaliação do défice cognitivo em idosos com recurso a sensores inerciais. As ferramentas propostas para avaliar o movimento e as duplas-tarefas são mais simples, rápidas e objetivas do que os testes tradicionais de rastreio em papel, oferecendo uma potencial alternativa ou complemento aos métodos existentes. Para comprovar a eficácia e viabilidade das soluções propostas, são necessários testes adicionais envolvendo um tamanho de amostragem superior. No futuro, estas avaliações podem melhorar o rastreio do défice cognitivo, contribuindo para uma melhor compreensão das interações motoras-cognitivas, desenvolvimento de novas terapias e melhor gestão da doença.

Keywords: Déficit Cognitivo, Rastreio Cognitivo, Dupla-Tarefa, Jogos de Movimento, Análise da Marcha, Sensores Inerciais, Aprendizagem Automática, Análise de Movimento, Idosos

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Miguel V. Correia and Doctor Inês Sousa, for their support and guidance throughout this 4-year journey. I deeply appreciated their commitment and willingness to discuss my work, which helped me to develop new ideas and perspectives. Their support and understanding were instrumental in helping me reach this significant milestone.

I would like to acknowledge Fraunhofer Portugal AICOS and the director Liliana Ferreira for providing the necessary support and conditions to pursue a PhD. I also want to extend special thanks to my current and former colleagues at Fraunhofer Portugal AICOS, especially Elsa Oliveira, Alberto Carvalho, Nuno Cardoso, Ana Pereira, Luís Francisco, Joana Silva, Diana Gomes, Sílvia Rêgo, Tiago Rocha, Eduardo Barbosa, Marcos Liberal, Filipe Sousa, António Santos, António Sousa, Juarez Souza, Mariana Pereira, João Oliveira, Ricardo Peixoto, Paulo Torres, and Ruben Moutinho. I am grateful for their contributions and valuable inputs to the projects VITAAL and Smart-Health-4-All, which created all the necessary conditions to successfully complete this work. I would like to thank Catarina Teixeira and Stefanie Götz for their support with the legal and ethical aspects, the Intelligent Systems and Time-Series Cluster groups for the inspiring discussions, and Maria Costa, Alexandre Nogueira, and the whole purchase department team at AICOS for their help with funding, purchases, and material envoys. For all the support and companionship, I am also grateful to Francisco Nunes, Luís Rosado, Maria Vasconcelos, Rui Castro, Ana Correia de Barros, Filipe Sousa, Manuel Monteiro, Nino Rocha, and João Gonçalves.

I would like to acknowledge VITAAL project partners for their insightful discussions and contributions, including Professor Eling D de Bruin for offering his expertise and assistance, Manuela Omlin for her friendly support, Melanie Thalman for taking the lead on trial planning and implementation, and the physiotherapy clinic Physio SPArtoS, specifically Jacqueline De Jong and Vincent Becker-Hoff for their contributions to participant recruitment, data recordings, and data entry. I am also grateful to the Clinical Academic Center of Braga (2CA-Braga) for facilitating and operationalizing data collection activities in the context of the Smart-Health-4-All project. I am especially grateful to Raquel Cunha and Rosana Magalhães for their remarkable efforts to make data collection possible. Their dedication and diligence were truly exceptional. Additionally, I would like to thank Braga Hospital and, in particular, the Neurology Service of the hospital, for enabling access to their patients, and to the Senior University of the Rotary Club in Póvoa do Lanhoso for facilitating access to healthy older adults, which was crucial for this research.

I would like to express my appreciation to all volunteers who participated in data collection activities. Without them, this research had not been possible. I would like to reiterate my sincere appreciation to Sílvia Rêgo, responsible for the Colaborar network at Fraunhofer Portugal AICOS, for the important role she had in facilitating data collection activities with older adults and ensuring that all ethical requirements were met. I also acknowledge the Ontario Neurodegenerative Disease Research Initiative (PI: Michael Strong, Mario Masellis, Douglas Munoz, and Richard Swartz), the Ontario Brain Institute, the Brain-CODE platform, the Government of Ontario, and the Temerty

Family Foundation, for making their data available.

Finally, I want to thank my beloved friends and family, especially David, my parents, my brother and sister-in-law, and my dearest friends Bruno, Cristiana, Inês, Anthony, Joel, Joana, Ana, and Diogo, for their unwavering support, belief, and encouragement. Your presence and friendship have been a constant source of comfort and strength, whether it was through offering a listening ear, cheering on small victories, or simply providing a much-needed break. Your encouragement, patience, and infinite understanding have been invaluable, and I cannot thank you enough.

This research was supported by the project VITAAL (AAL-2017-066), funded under the Active Assisted Living Program and co-funded by the European Commission and the National Funding Authorities of Portugal, Switzerland, and Belgium. It was also supported by the project Smart-Health-4-All (n.º 046115), co-funded by Portugal 2020, framed under the COMPETE 2020 (Competitiveness and Internationalization Operational Program), Lisbon Regional Operational Programme (LISBOA 2020), and European Regional Development Fund (ERDF) from European Union (EU). It was also supported by the project ConnectedHealth (n.º 46858), funded by Competitiveness and Internationalisation Operational Programme (POCI) and Lisbon Regional Operational Programme (LISBOA 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).



Vânia Guimarães

List of Publications

The work developed in this thesis has been published in journal articles and presented in international conferences. All publications produced during the course of my PhD candidacy are listed below and, when applicable, the corresponding thesis chapter where the article has been included is also presented. The contributions are described for the papers in which I participated as a co-author.

Journal articles

- 2021** V. Guimarães, I. Sousa, and M. V. Correia, “Orientation-Invariant Spatio-Temporal Gait Analysis Using Foot-Worn Inertial Sensors,” *Sensors*, vol. 21, no. 11, p. 3940, Jun. 2021.
<https://doi.org/10.3390/s21113940>
Open access Creative Commons Attribution (CC BY) 4.0 license.
Included in Chapter 3.
- 2021** V. Guimarães, I. Sousa, and M. V. Correia, “A Deep Learning Approach for Foot Trajectory Estimation in Gait Analysis Using Inertial Sensors,” *Sensors*, vol. 21, no. 22, p. 7517, Nov. 2021.
<https://doi.org/10.3390/s21227517>
Open access Creative Commons Attribution (CC BY) 4.0 license.
Included in Chapter 4.
- 2021** V. Guimarães, E. Oliveira, A. Carvalho, N. Cardoso, J. Emerich, C. Dumoulin, N. Swinnen, J. De Jong, E. D. de Bruin, “An Exergame Solution for Personalized Multicomponent Training in Older Adults,” *Applied Sciences*, vol. 11, no. 17, p. 7986, Aug. 2021.
<https://doi.org/10.3390/app11177986>
Open access Creative Commons Attribution (CC BY) 4.0 license.
Included in Chapter 6.
- 2023** V. Guimarães, I. Sousa, E. D. de Bruin, J. Pais, and M. V. Correia, “Using shoe-mounted inertial sensors and stepping exergames to assess the motor-cognitive status of older adults: A correlational study,” *Digital Health*, vol. 9, p. 205520762311670, Jan. 2023.
<https://doi.org/10.1177/20552076231167001>
Open access Creative Commons Attribution-NonCommercial license (CC BY-NC 4.0).
Included in Chapter 6.

- 2023** V. Guimarães, I. Sousa, E. D. de Bruin, J. Pais, and M. V. Correia, “Minding your steps: a cross-sectional pilot study using foot-worn inertial sensors and dual-task gait analysis to assess the cognitive status of older adults with mobility limitations,” *BMC Geriatrics*, vol. 23, no. 1, p. 329, May 2023.
<https://doi.org/10.1186/s12877-023-04042-6>
 Open access Creative Commons Attribution (CC BY) 4.0 license.
 Included in Chapter 7.
- 2023** V. Guimarães, I. Sousa, and M. V. Correia, “Detection of Foot Motions for Interaction With Exergames Using Shoe-Mounted Inertial Sensors,” *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 5, pp. 895–904, Oct. 2023.
<https://doi.org/10.1109/THMS.2023.3292902>
 Copyright © 2023 IEEE.
 Included in Chapter 5.
- 2023** V. Guimarães, I. Sousa, R. Cunha, R. Magalhães, Á. Machado, V. Fernandes, S. Reis, and M. V. Correia, “One-Class Classification with Confound Control for Cognitive Screening in Older Adults Using Gait, Fingertapping, Cognitive, and Dual Tasks,” submitted for publication.
 Included in Chapter 8.
- 2023** V. Guimarães, I. Sousa, and M. V. Correia, “Detecting Cognitive Impairment in Cerebrovascular Disease Using Gait, Dual-tasks and Machine Learning”, submitted for publication.
 Included in Chapter 9.

Conference papers

- 2019** V. Guimarães, I. Sousa and M. V. Correia, “Detection and classification of multidirectional steps for motor-cognitive training in older adults using shoe-mounted inertial sensors,” *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 6926-6929.
<https://doi.org/10.1109/EMBC.2019.8856851>
 Copyright © 2019 IEEE.
 Included in Chapter 5.
- 2021** V. Guimarães, I. Sousa and M. V. Correia, “Gait events detection from heel and toe trajectories: comparison of methods using multiple datasets,” *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2021, pp. 1-6.
<https://doi.org/10.1109/MeMeA52024.2021.9478606>
 Best Women in Engineering Paper Award.

Co-authored works (Journal and Conference)

- 2020** M. Adcock, M. Fankhauser, J. Post, K. Lutz, L. Zizlsperger, A. R. Luft, V. Guimarães, A. Schättin, and E. D. de Bruin, “Effects of an In-home Multicomponent Exergame Training on Physical Functions, Cognition, and Brain Volume of Older Adults: A Randomized Controlled Trial,” *Front. Med.*, vol. 6, p. 321, Jan. 2020.

<https://doi.org/10.3389/fmed.2019.00321>

Author's role: exergame development, trial support, paper revision, and editing.

- 2021** D. Gomes, V. Guimarães, and J. Silva, "A Fully-Automatic Gap Filling Approach for Motion Capture Trajectories," *Applied Sciences*, vol. 11, no. 21, p. 9847, Oct. 2021.

<https://doi.org/10.3390/app11219847>

Author's role: conceptualization, methodology, results analysis, and writing.

- 2021** N. Swinnen, E. D. de Bruin, C. Dumoulin, M. Thalmann, V. Guimarães, J. De Jong, M. Vandenbulcke, and D. Vancampfort, "The VITAAL stepping exergame prototype for older adults with major neurocognitive disorder: a usability study," *Frontiers in Aging Neuroscience*, vol. 13, no. 701319, 2021.

<https://doi.org/10.3389/fnagi.2021.701319>

Author's role: exergame development, paper revision, and editing.

- 2021** M. Thalmann, L. Ringli, M. Adcock, N. Swinnen, J. de Jong, C. Dumoulin, V. Guimarães, and E. D. de Bruin, "Usability Study of a Multicomponent Exergame Training for Older Adults with Mobility Limitations," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, p. 13422, Dec. 2021.

<https://doi.org/10.3390/ijerph182413422>

Author's role: exergame development, paper revision, and editing.

- 2023** N. Swinnen, E. D. de Bruin, V. Guimarães, C. Domoulin, J. De Jong, R. Akkerman, M. Vandenbulcke, B. Stubbs, and D. Vancampfort, "The feasibility of a stepping exergame prototype for older adults with major neurocognitive disorder residing in a long-term care facility: a mixed methods pilot study," *Disability and Rehabilitation*, pp. 1-15, Feb. 2023.

<https://doi.org/10.1080/09638288.2023.2182916>

Author's role: exergame development, paper revision, and editing.

- 2023** M. R. Soares, T. Rocha, D. Oliper, V. Guimarães, and R. Santos, "Novel Bluetooth 5.1 location services for indoor asset tracking using multilateration," *2023 Wireless Telecommunications Symposium (WTS)*, 2023, 9-21 April, Boston, Massachusetts, USA.

Author's role: supervision, results interpretation, paper revision, and editing.

*“Our ultimate goal, after all, is not a good death
but a good life to the very end.”*

Atul Gawande

Contents

List of Publications	vii
List of Figures	xv
List of Tables	xvii
I INTRODUCTION AND THEORETICAL BACKGROUND	1
1 Introduction	3
1.1 Context	3
1.2 Problem	4
1.3 Motivation	5
1.4 Objectives	6
1.5 Contributions	8
1.6 Document Structure	9
2 Background	11
2.1 Human Movement	11
2.2 Aging and Cognition	20
2.3 Linking Cognition and Movement	25
2.4 Automated Detection of Cognitive Impairment	31
2.5 Concluding Remarks	36
II MOVEMENT ANALYSIS USING FOOT-WORN INERTIAL SENSORS	39
3 Orientation-Invariant Spatiotemporal Gait Analysis	41
3.1 Introduction	42
3.2 Materials and Methods	43
3.3 Results	54
3.4 Discussion	57
3.5 Conclusion	62
4 A Deep Learning Approach for Foot Trajectory Estimation	63
4.1 Introduction	64
4.2 Materials and Methods	65
4.3 Results	75
4.4 Discussion	78

4.5	Conclusion	85
5	Stepping Performance Measures for Multidirectional Steps Evaluation	87
5.1	Introduction	88
5.2	Materials and Methods	89
5.3	Results	97
5.4	Discussion	99
5.5	Conclusion	104
III	TOWARD THE DETECTION OF COGNITIVE IMPAIRMENT	105
6	Gameplay Metrics for Motor-Cognitive Assessment: a Correlational Study	107
6.1	Introduction	108
6.2	Materials and Methods	109
6.3	Results	114
6.4	Discussion	117
6.5	Conclusion	124
7	Using Gait to Assess Cognition: a Cross-Sectional Study	125
7.1	Introduction	126
7.2	Materials and Methods	126
7.3	Results	131
7.4	Discussion	134
7.5	Conclusion	139
8	A Machine Learning Approach for the Detection of Mild Cognitive Impairment	147
8.1	Introduction	148
8.2	Methods	149
8.3	Results	156
8.4	Discussion	161
8.5	Conclusion	164
9	Detection of Cognitive Impairment in Post-Stroke Patients	167
9.1	Introduction	168
9.2	Materials and Methods	169
9.3	Results	175
9.4	Discussion	180
9.5	Conclusion	184
IV	FINAL REMARKS	185
10	Conclusion	187
10.1	Overview of the Contributions	188
10.2	Limitations and Future Work	192
10.3	Implications and Significance	195
	References	197

List of Figures

1.1	Cognitive screening based on movement analysis and task performance evaluation	6
1.2	Document Structure	10
2.1	Gait cycle and gait phases	13
2.2	Inertial measurement unit working principles	20
2.3	Neurocognitive domains	21
2.4	Subtypes of mild and major neurocognitive disorders	23
2.5	Determinants of cognitive and motor impairment	26
2.6	Schematic representation of the motoric cognitive risk syndrome	27
2.7	Dual-task interference framework	29
2.8	Receiver Operating Characteristic (ROC) curve	36
3.1	Wearable sensors and reference system	43
3.2	Illustration of spatial gait parameters	45
3.3	Events detection from low pass filtered magnitude and vertical acceleration	49
3.4	Variables involved in the calculation of toe trajectory	50
3.5	Synthetically generated rotations and their relationship with differences in measured turning angles	57
4.1	Placement of the inertial sensor and markers on the shoes	66
4.2	Definition of spatial gait parameters based on heel and toe trajectories	67
4.3	Coordinate systems of the heel and toe stride trajectories	68
4.4	Data analysis process diagram: data preparation, network training, and model performance evaluation	69
4.5	Neural network architecture	71
4.6	Data analysis process diagram: estimating gait parameters from predicted heel and toe trajectories	72
4.7	Events detection from acceleration magnitude (obtained from the inertial sensor data) and vertical acceleration (obtained from predicted toe trajectories)	73
4.8	Mean squared error (loss) achieved while training the model with the best selection of hyperparameters	76
4.9	An example of predicted and reference trajectories, obtained from a sample in the test set	77
5.1	Shoe-mounted inertial sensors and steps directions for interaction with games	91
5.2	Step segmentation state diagram	93
5.3	Illustration of the reflection in plane x-z	95
5.4	Exemplary step segmentation result	97

5.5	Confusion matrices for the classification of steps using the Logistic Regression model	98
5.6	Confusion matrix and accuracy values obtained with the Logistic Regression model for the detection of steps with regards to the annotated displacements	99
5.7	Misclassified steps over time	100
6.1	Stepping-based minigames	112
6.2	Progression within cognitive games	119
6.3	Step speed across difficulty levels presented for all balance and cognitive games .	120
7.1	Group $\times$ task interactions	132
7.2	Regression model for the prediction of MoCA scores	134
8.1	Data collection procedure	151
8.2	Nested cross-validation for OCC	155
8.3	Proportion of features with significant differences between groups before and after confound regression	158
8.4	Comparison of the best OC-SVM results with and without confound control . . .	159
8.5	Comparison between OC-SVM and TC-SVM, with the latter being evaluated both with and without sampling with SMOTE	161
8.6	Distribution of performance metrics and correlations between distances to decision boundaries and MoCA scores	162
8.7	Test predictions in an optimistic dataset partition	163
9.1	Machine learning pipeline with nested cross-validation	173
9.2	Performance using features extracted from the accelerometer	177
9.3	Performance using features extracted from the walkway	178
9.4	Classification performance using MoCA and a majority voting approach	180

List of Tables

2.1	Advantages and disadvantages of non-wearable systems for movement analysis	18
2.2	Advantages and disadvantages of wearable systems for movement analysis	19
2.3	Functional limitations associated with impairment in different cognitive domains	22
2.4	Commonly used cognitive screening tests for mild cognitive impairment	25
2.5	Summary of machine learning approaches using gait tests to distinguish healthy controls from cognitively impaired older adults	33
2.6	Summary of machine learning approaches using other task-oriented movements to distinguish healthy controls from cognitively impaired older adults	34
3.1	Dataset characteristics after splitting	51
3.2	Methods and parameters used in coarse and fine grid-search	53
3.3	Performance on the development set, using the best combination of methods and parameters	54
3.4	Performance on the validation set, including turning strides	55
3.5	Performance on the validation set, excluding turning strides	56
3.6	Orientation invariance results	56
4.1	Dataset splitting	69
4.2	Hyperparameters optimization	71
4.3	Performance of the neural network on validation and test sets	78
4.4	Performance on the test set, including turns	79
4.5	Performance on the test set, excluding turns	80
4.6	Random orientation simulation results	81
4.7	Comparison of the performance achieved with the deep learning approach and the conventional approach	82
5.1	Number of steps in training and test sets	96
5.2	Number of steps missed	97
5.3	Classification performance	98
6.1	Inclusion and exclusion criteria	110
6.2	Demographic and clinical characteristics of the intervention group	115
6.3	Correlations between step characteristics and global physical and cognitive performance tests	116
6.4	Correlations between cognitive games' metrics and MoCA scores	117
6.5	Number of participants playing each game level, stratified by their cognitive status	118
7.1	Demographic and clinical characteristics of study participants stratified by cognitive status	141

7.2	Comparison of average temporal gait parameters within and between groups . . .	142
7.3	Comparison of average spatial gait parameters within and between groups	143
7.4	Comparison of dual-task variability measures between groups	144
7.5	Comparison of dual-task asymmetry measures between groups	144
7.6	Classification performance based on logistic regression and ROC analysis	145
7.7	Associations between gait metrics and MoCA scores in multiple bivariate linear regression analysis	146
8.1	Features extracted from smartphone-guided tasks	152
8.2	Hyperparameters search space	156
8.3	Descriptive statistics	157
8.4	Results of the One-Class SVM	160
8.5	Results of the Two-Class SVM	160
8.6	Performance using MoCA and a majority voting approach	164
9.1	Neuropsychological assessment	172
9.2	Descriptive statistics	176
9.3	Performance on test sets using gait metrics from the accelerometer	177
9.4	Performance on test sets using gait metrics from the electronic walkway	178
9.5	Performance using MoCA and a combination with gait	180

Abbreviations

ACE	Addenbrooke's Cognitive Examination
AD	Alzheimer's disease
ADLs	Activities of daily living
AI	Artificial Intelligence
ALS	Amyotrophic Lateral Sclerosis
AUC	Area Under the Curve
BMI	Body Mass Index
CF	Complementary Filter
CI	Cognitive Impairment
CN	Cognitively Normal
CNN	Convolutional Neural Networks
CoV	Coefficient of Variation
CSF	Cerebrospinal Fluid
CV	Cross-Validation
CVD	Cerebrovascular disease
CWIT	Color Word Interference Test
DAn	Walking while naming animals
DLB	Dementia with Lewy Bodies
DS1	Walking while subtracting serial ones
DS7	Walking while subtracting serial sevens
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (5th ed.)
DT	Dual Task
DTE	Dual-Task Effect
F	Fast walking
FC	Foot Contact
FN	False Negatives
FO	Foot Off
FP	False Positives
FT	Fingertapping
FTD	Frontotemporal disorders
iADLs	Instrumental activities of daily living
ICD	Intra-Class Distance
IMU	Inertial Measurement Unit

IQR	Interquartile Range
KNN	K-Nearest Neighbors
LOOCV	Leave-One-Out Cross-Validation
LOSOCV	Leave-One-Subject-Out Cross-Validation
LR	Logistic Regression
LSTM	Long Short-Term Memory
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MCI	Mild Cognitive Impairment
MCR	Motoric Cognitive Risk Syndrome
MI	Mutual Information
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MRI	Magnetic Resonance Imaging
MS	Mid-Stance
MSE	Mean Squared Error
MSTS	1-Minute Sit to Stand Test
MTC	Minimum Toe Clearance
NN	Neural Networks
OC	One-Class
OCC	One-Class Classification
ONDRI	Ontario Neurodegenerative Disease Research Initiative
PCA	Principal Component Analysis
PD	Parkinson's Disease
PET	Positron Emission Tomography
QIDS	Quick Inventory of Depressive Symptomatology
RBF	Radial Basis Function
RCT	Randomized Controlled Pilot Trial
RF	Random Forest
RFE	Recursive Feature Elimination
RMSD	Root Mean Square Deviation
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RT	Reaction Time
SC	Subjective Complaints
SD	Standard Deviation
SI	Symmetry Index
SL	Stride Length

SMOTE	Synthetic Minority Oversampling Technique
SPPB	Short Physical Performance Battery
SS	Preferred Speed Single-Task Walking
ST	Single Task
SVM	Support Vector Machine
SW	Swing Width
TC	Two-Class
TCC	Two-Class Classification
TMT	Trail Making Test
TN	True Negatives
TP	True Positives
TRIAD	Tri-axial Attitude Determination
VaD	Vascular Dementia
VCI	Vascular Cognitive Impairment
WMSR	Wechsler Memory Scale-Revised
ZVI	Zero Velocity Interval

Part I

INTRODUCTION AND THEORETICAL BACKGROUND

Chapter 1

Introduction

Despite advances in geriatric medicine, cognitive disorders remain a critical aspect of aging. Cognitive impairment leads to a significant decline in quality of life. Moreover, it is associated with increased risk of dementia, mortality [1, 2], and mobility disorders leading to falls, injuries, and loss of independence [3].

Early diagnosis of cognitive impairment remains the most promising strategy to prevent subsequent disability and improve prognosis through early intervention [4, 5]. However, the subtle presentation of symptoms, the high workload of specialists, and the lack of adequate screening tools constitute barriers to diagnosis. The detection of cognitive impairment is thus likely to be missed or delayed [4, 6].

This thesis focuses on the development of novel tools that explore the use of movement analysis and machine learning to improve the early detection of cognitive impairment. In this introductory chapter, we start by describing the context (Section 1.1), the problem (Section 1.2), and the motivation (Section 1.3) for this work. Then, we outline its objectives (Section 1.4) and scientific contributions (Section 1.5). In Section 1.6, we describe document structure.

1.1 Context

The aging of the population and the high prevalence of age-related conditions, including cognitive and mobility disorders, are bringing new health, economic, and social challenges [7, 8]. The global prevalence of cognitive impairment is estimated to be between 5% and 41% in older adults [9], affecting cognitive domains such as memory, executive functions, attention, language, or visuospatial skills [2].

When cognitive decline exceeds what is expected due to the normal aging process, the decline is diagnosed based on severity as Mild Cognitive Impairment (MCI) or dementia [10–12]. In MCI, cognitive decline is greater than expected for age, but the ability to perform everyday activities is still preserved [10, 12]. Dementia, or major cognitive impairment, is established when

the loss in cognitive functioning is significant enough to interfere with a person's functional independence [11, 12]. People with MCI have a higher risk of conversion to dementia. Studies suggest that 8% to 15% of the MCI cases progress to dementia each year [13].

Dementia is caused by damage or changes in the brain, often stemming from conditions such as Alzheimer's disease, vascular disease, Parkinson's disease, Lewy body disease, or frontotemporal lobar degeneration [12]. Dementia is a major cause of disability and dependency causing a significant social and economic impact worldwide [14]. Around 50 million people worldwide are living with dementia, and its prevalence is expected to triple in the next 30 years resulting from the rapid aging of the population [14]. In Portugal, 6% of the population aged 60 years and older are estimated to have dementia, which represents a cost of around €37M per year, just with medication [15]. The World Health Organization (WHO) describes dementia as a public health priority [16].

Currently, there are no treatments capable of reversing or stopping the progression of dementia. However, current treatments can stabilize or slow the progression of cognitive decline [17, 18]. Therefore, the earlier cognitive decline is detected, the sooner appropriate care can be administered, minimizing patients' suffering and improving their quality of life [19]. To this purpose, the study of cognitive decline and MCI constitutes a priority.

1.2 Problem

People with MCI are more likely to develop dementia. However, MCI remains significantly overlooked. Recognizing early signs of dementia can be challenging due to the subtlety of its symptoms, which can go unnoticed until they start to disrupt daily functioning [20]. Although comprehensive neuropsychological testing is the gold standard for assessing cognitive function, it requires substantial time and expertise to administer and score, making it unsuitable for widespread population screening [21].

The screening for cognitive impairment is currently based on faster paper-and-pencil tests, such as the Montreal Cognitive Assessment (MoCA) [22]. Even though these tests were designed for screening, they still require a significant time investment of a trained questioner, being incompatible with the busy pace of clinical environments. Besides being performed in unfamiliar settings, these methods are also unsuitable for self-administration on a regular basis [23]. Moreover, older adults often perceive their cognitive decline as a natural consequence of aging and may not actively seek medical advice [24, 25]. General practitioners often lack the necessary specialized knowledge to recognize abnormal cognitive decline, and the low frequency of appointments may further limit their ability to do so [26]. At the same time, referrals for neuropsychological testing can be costly and time-consuming. Cognitive screening is thus delayed leading to abnormal declines being detected and intervened late, even though the detection of early signs of impairment is crucial to prevent and better manage further decline [4, 5].

Alternative screening methods have been proposed in the literature, such as computerized assessments designed for self-administration at home [27], serious games [28], speech analysis [29, 30], unobtrusive assessments of daily activities [31], and smartphone interactions [32]. While some of these methods largely depend on patients' compliance, others are still too far from being adopted and routinely applied in clinical practice. Additionally, the lack of objective, robust, and fast screening tools poses significant barriers to the assessment of cognitive decline in primary care settings. As a result, researchers keep joining their efforts to explore alternative biomarkers and develop new strategies for more frequent screening [33]. As the population ages, the prevalence of age-related cognitive impairments is anticipated to rise significantly, imposing an increasingly substantial burden on our society [34, 35].

1.3 Motivation

Although cognitive decline is the main clinical sign of cognitive impairment, recent studies show that motor decline can also occur, either concurrently with or before the onset of cognitive impairment [36–38]. In individuals with cognitive impairment, lower motor performance is characterized by a significant reduction in gait speed, stride length, and cadence, and an increase in gait variability [39–41]. Worsening of gait performance determined through these parameters is associated with global imaging markers of a “deteriorating brain” [42].

Studies in cognitive neuroscience have established a link between motor and cognitive functions, demonstrating that effective motor control requires intact cognition, specifically executive functions [43, 44]. Common underlying brain regions and shared pathological mechanisms have been associated with both gait and cognition [45–48]. The term “Motoric Cognitive Risk Syndrome” has been used to describe the state of the simultaneous presence of slow walking speed and self-reported cognitive decline [49]. The Motoric Cognitive Risk Syndrome was shown to predict cognitive impairment and dementia, suggesting that it may be used to improve screening in primary care settings [50].

The quality of walking while simultaneously performing a cognitive attention-demanding task such as naming animals, or doing serial subtractions, also known as dual-tasking, unmasks underlying gait disturbances that may only be evident under cognitive stress [36, 51]. The quality of gait under dual-task testing is affected by cognitive impairment and can distinguish MCI from preserved cognition in older adults, exposing symptoms that would otherwise be invisible [52]. The results of a meta-analysis indicated that various gait parameters, including speed, stride length, and stride time, could differentiate MCI from healthy controls under single-task conditions. However, the discriminative power of these parameters was further increased when assessed under dual-task conditions [53]. According to another meta-analysis, the effects of the dual-task on gait speed are highly sensitive for detecting MCI [52]. The findings from previous studies thus suggest that dual-task testing could serve as a sensitive measure for improving the early detection of cognitive impairment [36, 46, 53–56].

Conventional gait and balance tests based on timed assessments only enable evaluation of a restricted set of metrics and may not be sufficiently sensitive to identify the subtle motor deficits associated with MCI [57, 58]. Additionally, some studies suggest that a comprehensive and objective gait analysis, involving multiple gait characteristics, is necessary to effectively evaluate gait deficits associated with cognitive impairment [59, 60]. However, the technology required for such an assessment, for instance, instrumented walkways or optical motion capture systems, is expensive and requires dedicated spaces that are not ideal for routine screening within clinical settings [58]. Additionally, these systems have limited coverage areas [58], which compromises the assessment of metrics like gait variability due to insufficient walking distance [53].

Wearable technology based on inertial sensors offers a low-cost and less restricted alternative for the assessment of movement, allowing it to be captured both at the clinic and in free-living conditions [58]. Due to advances in data analytics, wearables are now widely proposed as practical tools to aid the diagnosis and treatment of a range of neurological disorders [61]. Despite extensive work in this area, the problem of accurately measuring movement parameters associated with cognitive impairment by resorting to inertial sensors is not yet entirely solved and remains an ongoing area of investigation [62, 63]. Thus, it remains to be established whether wearables and their associated algorithms are capable of detecting cognitive impairment [58, 64]. Despite not yet being standardized for community-wide screening, dual-task tests are easy to administer, low-cost, and non-invasive, which supports their use, for instance, in primary care settings [46].

1.4 Objectives

This work aims to explore the use of movement analysis and machine learning to improve the early detection of cognitive impairment. For that purpose, we explore different task performance assessments, envisioning two different application scenarios (Figure 1.1): i) a cognitive screening tool for supervised use (e.g., at the clinic) relying on the analysis of gait and dual-task tests, and ii) a cognitive screening solution for unsupervised use (e.g., at home) exploring motor and cognitive metrics extracted from an exergame.

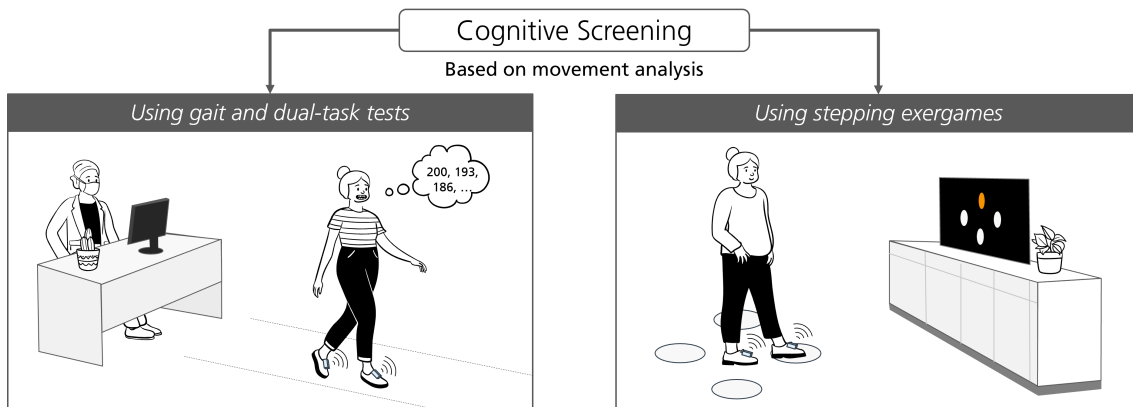


Figure 1.1: Cognitive screening based on movement analysis and task performance evaluation.

To achieve this overarching objective, the following specific objectives and research questions are addressed in this work:

Objective 1: Development of a reliable inertial sensor-based gait analysis solution for the extraction of a comprehensive set of gait parameters (**Chapters 3 and 4**).

Which signal processing techniques, or, which combination of methods and parameters, can be used to improve the extraction of spatiotemporal gait parameters?

To what extent can deep learning models outperform traditional gait analysis pipelines in accurately estimating foot trajectories?

What approaches can be employed to measure and improve the robustness of the gait analysis under different sensor-body alignment conditions?

Objective 2: Development of an inertial sensor-based stepping evaluation method to evaluate foot motions while interacting with stepping-based motor-cognitive games (**Chapter 5**).

How can inertial sensors be used to evaluate foot motions and extract stepping characteristics while interacting with exergames?

Objective 3: Studying the pertinence of stepping and gameplay metrics to assess the motor-cognitive status of older adults (**Chapter 6**).

What stepping and gameplay performance metrics can be relevant to assess the motor-cognitive status of older adults?

Objective 4: Studying the pertinence of gait metrics in distinguishing older adults with and without cognitive impairment (**Chapter 7**).

Can foot-worn inertial sensors be used to capture and differentiate gait parameters in the presence of cognitive impairment?

Which gait parameters are more relevant?

Objective 5: Development of machine learning models for the automatic classification of cognitive impairment based on movement data (**Chapters 8 and 9**).

What machine learning algorithms or models are suitable for effectively classifying cognitive impairment based on movement data?

Which movement data features or tests are the most informative for the classification of cognitive impairment?

How do the machine learning models compare to traditional paper-based screening tests in terms of performance?

By addressing these research questions, this work aims to improve our understanding of how movement and machine learning techniques can be used to improve the detection of cognitive impairment, relying on objective, non-invasive, task performance assessments. In particular, we investigate whether easily applicable, inexpensive, and unobtrusive instruments can be used in the detection of motor changes associated with cognitive impairment. We explore the use of inertial sensors placed on the feet and advanced movement analysis techniques to measure clinically relevant gait and stepping parameters in older adults. Moreover, we combine gait assessments and gameplay measurements with challenging cognitive tasks to unmask motor-cognitive disorders that may only be evident under cognitive stress. We investigate the ability of motor, cognitive, and dual tasks to differentiate older adults with and without cognitive impairment, as an alternative or complement to existing paper-and-pencil screening tools. Finally, we propose machine learning models for the automated detection of cognitive impairment.

1.5 Contributions

The main contributions of this work are as follows:

- We developed a gait analysis solution based on foot-worn inertial sensors for the extraction of clinically relevant gait parameters in older adults, whose results are invariant to the orientation with which the sensors are placed on the feet. Moreover, we provided a comparison of methods for orientation estimation and double integration and proposed a method to objectively assess robustness to different sensor orientations (**Chapter 3**).
- We proposed a novel deep learning-based approach for the estimation of heel and toe trajectories while walking, relying on foot-worn inertial sensor data. The estimations were performed on a stride-by-stride basis, using an inter-stride independent coordinate frame that removes any dependencies from previous strides and external inputs. The estimations were shown to be robust to imperfect sensor-foot alignment conditions (**Chapter 4**).
- We acquired a dataset for the validation of the proposed inertial sensor-based gait analysis solutions using an optical motion capture system as a reference (**Chapters 3 and 4**).
- We proposed a method to estimate the direction of steps and extract stepping characteristics while interacting with exergames using inertial sensors. The method incorporates a novel calibration approach to compensate for sensor orientation drift over time, as well as a novel foot reflection method that allows for the analysis of both feet without distinction (**Chapter 5**).
- We investigated the potential of stepping and gameplay metrics to assess the motor-cognitive status of older adults, by analyzing the associations between these metrics and standard cognitive and mobility assessments. We used data from older adults with mobility limitations who played the exergame (**Chapter 6**).

- In a cross-sectional study, we investigated the potential of the proposed gait analysis solutions to evaluate gait metrics affected by cognitive impairment based on single- and dual-task gait assessments. In this study, we explored differences between cognitively impaired and control groups considering a comprehensive set of gait parameters, some of them not yet explored in the literature. We also assessed their discrimination ability and association with cognitive performance (**Chapter 7**).
- We defined a protocol for data acquisition comprised of a cognitive task, single- and dual-task gait tests, and single- and dual-task fingertapping tests, acquired from older adults with and without a diagnosis of mild cognitive impairment. The dataset comprised inertial sensor data from gait tasks, touch events from fingertapping on a smartphone, and audio transcriptions from the cognitive task (**Chapter 8**).
- We proposed a machine learning approach based on one-class classification with confound control for the detection of mild cognitive impairment in an imbalanced dataset comprised of gait, fingertapping, cognitive, and dual tasks (as described above). We compared the performance of the different tasks, as well as the combination of features (early fusion) and models (late fusion). Moreover, we compared the performance of these models with traditional binary classification approaches and assessed their performance in comparison with conventional paper-based screening tests (**Chapter 8**).
- We proposed a machine learning approach for the detection of cognitive impairment in cerebrovascular disease, comparing different dual-task tests and gait metrics. Classification performance was assessed using data from post-stroke older adults. Also, we proposed a majority voting approach to combine the predictions of gait-based models with a cognitive screening test. A comprehensive neuropsychological assessment was used to categorize participants as impaired or as cognitively normal (**Chapter 9**).

1.6 Document Structure

This document is organized into four parts, and a total of 10 chapters, structured as follows:

- In **Part I**, we provide an introduction to the topic (Chapter 1) and present the main concepts and fundamentals needed to understand this work (Chapter 2).
- In **Part II**, we present our contributions to movement analysis based on foot-worn inertial sensors. Chapters 3 and 4 are dedicated to the development of a gait analysis solution, whereas Chapter 5 focuses on the extraction of stepping performance measures while interacting with exergames.

- In **Part III**, we present our contributions toward the use of movement and task performance assessments to detect cognitive impairment. Chapter 6 investigates the potential of gameplay metrics to assess the motor-cognitive status of older adults, whereas Chapters 7, 8, and 9 explore the detection of cognitive impairment using gait metrics and dual-tasks.
- In **Part IV**, comprised of Chapter 10, we provide the main conclusions of this work, reflecting on its contributions, limitations, and future directions of research.

The document structure and organization are depicted in Figure 1.2.

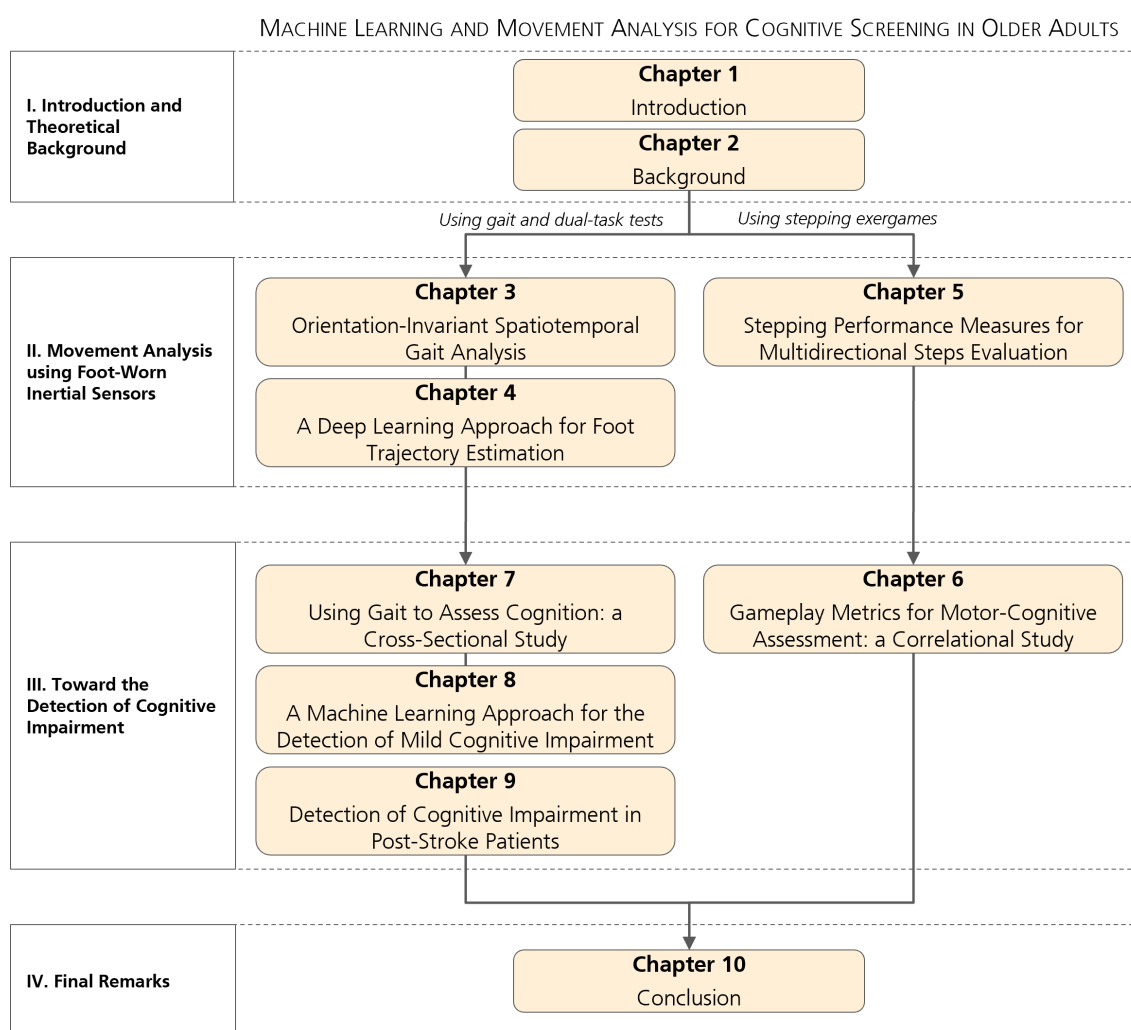


Figure 1.2: Document Structure.

Chapter 2

Background

This chapter offers a thorough understanding of the relationship between movement and cognition, with a focus on the latest and most promising tools and technology for assessing cognitive and motor abilities in older adults. We start by describing movement and gait, as well as the instruments that can be used to evaluate them (Section 2.1). Then, we delve into the effects of aging on cognition, clarify the definition of cognitive impairment, and identify the most commonly used instruments for cognitive screening (Section 2.2). We then elucidate the connection between cognition and movement and explain how assessments of movement and dual-task performance can be used to evaluate cognitive function (Section 2.3). Finally, we describe machine learning approaches commonly used to detect cognitive impairment in older adults resorting to movement and task performance assessments (Section 2.4). We conclude the chapter by critically evaluating current challenges in neurocognitive research, identifying unmet clinical demands, and suggesting potential research directions (Section 2.5).

2.1 Human Movement

Human movement is crucial to maintaining an independent life. Mobility, defined as the ability to move in an environment without difficulties or restrictions, is essential for the effective performance of activities of daily living, such as walking or climbing stairs. However, with age and various medical conditions, mobility can become impaired, leading to difficulties while performing these activities [65, 66]. Mobility limitations are prevalent in older adults, with around 23% to 47% experiencing mobility problems, most commonly due to chronic conditions such as arthritis [67]. Problems with ambulation and mobility are frequent and a major concern in older adults, leading to functional limitations, increased dependency, and reduced quality of life [68, 69].

Gait, or walking, is a fundamental activity required for daily life activities [70]. Changes in gait and obstacle avoidance strategies frequently lead to the occurrence of falls and reduced mobility [65, 67, 71]. The incidence of gait and balance disorders is expected to rise with the aging population, as their prevalence markedly increases with age [72].

Gait disorders in older adults can be part of the normal aging process or due to underlying pathologies such as neurological (e.g. multiple sclerosis, Parkinson's disease) and orthopedic problems (e.g. osteoarthritis and skeletal deformities), heart failure, respiratory insufficiency, and obesity [35, 66, 73]. Lower muscle mass (sarcopenia), less strength and power production (dynapenia), declines in proprioceptive function, and joint deformities are commonly present in older adults, contributing to gait disorders [74, 75].

In 2000, Bloem et al. [25] used the term “Senile Gait Disorders” to describe gait disturbances with unknown origin. The occurrence of senile gait disorders, characterized by cautious gait with slower speed, increased instability, and excessive fear of falling, has been associated with increased risk of cardiovascular disease, dementia, institutionalization, and death [25, 35, 76]. Changes in walking patterns can be used as an early sign of declining health and disability, and taking action early can help prevent further limitations, maintain or restore daily activity levels, and promote healthy aging and overall quality of life in older adults. However, identifying those at risk can be challenging since early signs of limitations do not necessarily require medical services and, as a result, they may not be adequately documented in medical diagnosis codes [67].

2.1.1 Characterization of Gait

Despite its apparent simplicity, gait is a highly complex and attention-demanding task. It involves three primary components: initiating and maintaining a rhythmic stepping (locomotion), maintaining balance, and adapting to the environment. Achieving a normal gait requires the integration of sensory input (such as visual, proprioceptive, and vestibular) and effective motor planning, initiation, execution, integration, and coordination. Gait control involves all levels of the neural pathway, including cortical, subcortical, spinal, and peripheral levels, as well as the musculoskeletal and cardiovascular systems [3, 35].

During walking, the body weight shifts from one leg to the other. The gait cycle, also known as stride, comprises a sequence of events between two successive contacts of the same foot and involves forward displacement of the body. It includes two main phases: the stance phase and the swing phase. The stance phase comprises approximately 60% of the gait cycle and can be subdivided into initial contact (or heel strike), loading response, mid-stand, terminal stance, and pre-swing. At the beginning and end of the stance phase, both feet are in contact with the ground. This period, called the double support phase, lasts approximately 10-12% of the gait cycle. The remaining 40% of the gait cycle corresponds to the swing phase, which is further divided into the initial swing, mid-swing, and terminal swing. The swing phase starts when the foot leaves the ground (toe-off) and finishes when the heel touches the ground to initiate the stance phase again [35]. A schematic representation of the gait cycle and its phases can be found in Figure 2.1.

The following list details some of the most relevant and clinically interpretable gait measures:

- Spatial Parameters
 - **Step length (m):** Linear distance between two consecutive footprints.

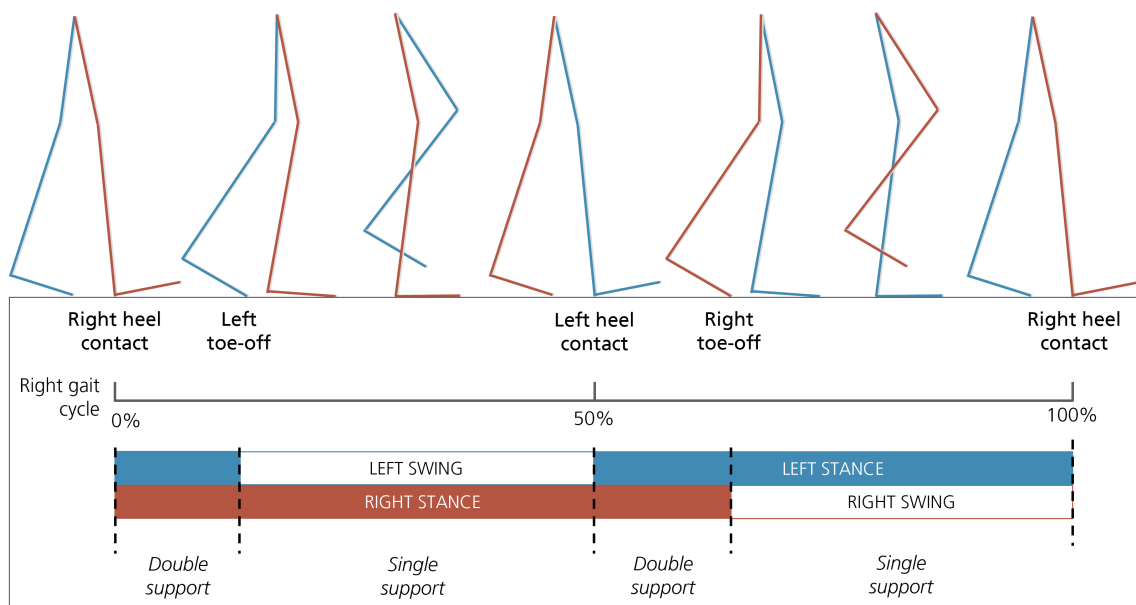


Figure 2.1: Gait cycle and gait phases.

- **Stride length (m):** Linear distance between two consecutive footprints of the same foot (left to left, right to right).
- **Swing width (cm):** Maximal lateral displacement of the foot during the swing phase, i.e. the greatest lateral distance between the actual path and the intended forward path of the foot. Also referred to as circumduction [77].
- **Stride width (cm):** Medio-lateral distance between the heels in two consecutive footprints [78].
- **Foot clearance (m):** Vertical displacement of the foot during swing. Parameters such as maximum heel clearance, maximum toe clearance, and minimum toe clearance all represent foot clearance [79].

- Temporal Parameters

- **Cadence (steps/min):** Number of steps per minute, or number of completed steps during the walking time, or number of strides per minute [78].
- **Step Time (s):** Time elapsed from the initial contact of one foot to the initial contact of the opposite foot.
- **Stride Time (s):** Time elapsed between the initial contacts of two consecutive footfalls of the same foot.
- **Stance Time (s):** The period of the gait cycle during which the foot is in contact with the ground. It begins when the foot makes contact with the ground (heel strike) and ends when the same foot leaves the ground (toe-off).

- **Swing Time (s):** The period of the gait cycle during which the foot is not in contact with the ground. It begins when the foot leaves the ground (toe-off) and ends when the same foot makes contact with the ground again (heel strike).
 - **Single Support Time (s):** It refers to the duration of the gait cycle during which only one foot is in contact with the ground. It begins when the opposite foot leaves the ground and ends when it makes contact with the ground again.
 - **Double Support Time (s):** It refers to the duration of the gait cycle during which both feet are in contact with the ground. It occurs twice during each gait cycle.
- **Temporophasic Parameters**
 - **Stance Ratio (%):** Stance time normalized to stride time.
 - **Swing Ratio (%):** Swing time normalized to stride time.
 - **Single Support Ratio (%):** Single support time normalized to stride time.
 - **Double Support Ratio (%):** Double support time normalized to stride time.
 - **Loading (%):** Time loading (heel strike to foot-flat) as a percentage of stance time.
 - **Foot-flat (%):** Time in foot-flat (full contact with the ground), as a percentage of stance time.
 - **Pushing (%):** Time pushing (foot-flat to toe-off) as a percentage of stance time.
 - **Spatiotemporal Parameters**
 - **Stride speed (m/s):** Calculated by dividing stride length by the stride time.
 - **Walk-Ratio (cm/steps/min):** The walk-ratio describes the spatiotemporal coordination of walking, representing the control of amplitude and frequency of steps. It is calculated as the ratio between step length and step frequency [80].
 - **Variability and asymmetry**
 - **Variability (%):** Variability describes the stride-to-stride fluctuations in a certain gait parameter. It is usually expressed as the Coefficient of Variation (CoV) of a metric, calculated as the standard deviation divided by the average of the metric. Standard deviation is also sometimes used to quantify the variability of a parameter [80]. Usually, more than 20 stride cycles are required to properly evaluate gait variability [81].
 - **Asymmetry (%):** Asymmetry compares a gait parameter measured on the left (X_L) and right (X_R) feet. Although several coefficients may be used to assess symmetry—for instance, the Ratio Index, the Symmetry Index (SI), the Gait Asymmetry, or the Symmetry Angle—Błażkiewicz et al. [82] recommended using the SI, which is defined as $SI = \frac{|X_L - X_R|}{0.5 \times |X_L + X_R|} \times 100$.

By being a strong indicator of general health and survival, gait speed is one of the most important gait parameters [35, 83]. Declines in gait speed can stem from clinical and subclinical impairments that accumulate across multiple systems, such as the peripheral nervous system and cardiovascular system, as well as functions such as vision and hearing [84]. Demographic and lifestyle-related factors (e.g. education and physical activity), acute (e.g. infections) and chronic conditions (e.g. obesity), and external factors (e.g. medications) may all affect gait speed [84].

The usual or preferred gait speed remains consistent throughout adult life, but it tends to decrease with advanced age [84]. In apparently healthy older adults, the preferred walking speed decreases by approximately 1% per year from an average of 1.3 m/s in the seventh decade to an average of 0.95 m/s in individuals aged over 80 years old [85]. The literature suggests that individuals unconsciously choose these speeds in order to optimize walking efficiency and reduce energy expenditure. Higher energetic costs are linked to a greater likelihood of developing a slow gait [86]. A change of 0.1 m/s in the usual walking speed is considered the minimal clinically important difference for assessing health conditions and evaluating the efficacy of interventions [80, 87].

Changes in gait speed may be used to assess physical function [88], quality of life [89], frailty [90], risk of falling [91, 92], risk of dementia [37, 49, 93] and even risk of early mortality [83]. In 2014, a Mobility Working Group recommended the adoption of a dismobility diagnosis, which involves using a gait speed cut point of 0.6 m/s to differentiate individuals with a higher risk of disability and other negative health-related outcomes [94]. The working group further recommended that gait speed should be incorporated into medical practice as a routine measure. They noted that the lack of a diagnosis for very slow gait speed renders impaired mobility unnoticed. By implementing these measures, it should be possible to encourage investigations into treatable causes of slow gait speed, enhance clinical care, and expedite regulatory approval of interventions aimed at improving mobility [94].

In addition to gait speed, changes in other gait parameters have also been recognized as independent predictors of fall risk. Reduced stride length, reduced cadence, reduced walk-ratio and increased double-support time have been associated with increased risk of future falls [95]. Reduced stride length, reduced speed, and increased double-support time have been associated with increased fear of falling [91]. Widening of the step width has been observed as a protective strategy to prevent falls [96] and the variability in step width as a global indicator of balance control [80]. Adequate foot clearance was considered crucial to avoid fall-related tripping since most falls occur when the moving foot hits the ground unexpectedly, such as when encountering an unseen obstacle. Foot drop or weakness of the pelvic muscles causes the foot to be transferred in a circumduction [80]. Foot clearance and, in particular, minimum toe clearance, have been connected to an increased risk of falling in older adults [97]. Moreover, increased gait variability in multiple gait metrics, such as stride length, step width, or stride time, have been associated with future falls in older persons [91, 92, 98, 99].

Gait changes may be subtle and difficult to detect through clinical observation alone, thus requiring the aid of instruments and quantitative assessments as noted by Bridenbaugh et al. [100]. Visual observation is a subjective method that relies entirely on the perception of the observer,

which can result in low reliability [66, 101, 102]. A thorough investigation of such changes, incorporating quantitative methods, is thus necessary to identify underlying causes as early as possible and apply proper interventions [35].

2.1.2 Instruments

Movement analysis refers to the scientific investigation of human motion. Its applications are not limited to clinical and medical studies such as rehabilitation and health diagnosis, but also extend to the fields of sports and robotics research, among others [103]. Even though movement analysis may include the study of kinematics (e.g., position, motion, trajectories), kinetics (e.g. moments and force obtained through inverse dynamic analysis), and muscle activation [70, 103], this work focuses solely on the instruments used for the analysis of kinematics and related spatiotemporal parameters.

In the clinical setting, gait and balance are commonly evaluated by visual observation, questionnaires, or simple functional performance assessments, such as the Timed Up and Go test (TUG), the Six-Minute Walk test (6MWT), or the Performance-Oriented Mobility Assessment (POMA), which are either timed or scored semiquantitatively [104]. These evaluations do not require sophisticated equipment and are easy to perform. However, they may provide limited information of suboptimal quality, which can result in subtle changes going unnoticed [105–107].

Multiple systems and technology can be used for human movement analysis, including non-wearable and wearable systems.

2.1.2.1 Non-Wearable Systems

Optical motion capture systems are widely regarded as the gold standard for human motion analysis [105]. They employ multiple cameras to capture the movements and they may or not use optical markers to aid in body movement detection [70]. Marker-based optical motion capture systems are commonly found in gait laboratories and are widely used in research [108]. They require reflective markers to be placed in strategic positions of the human body, providing high-precision three-dimensional spatial information [105]. They allow an in-depth analysis of a large number of kinematic parameters, and, when synchronized with ground reaction forces obtained from force platforms, they allow computation of kinetic information [105, 109]. However, their high cost, complexity, and time-intensive nature limit their application to laboratory settings. Moreover, given their limited coverage area, they restrict patients' walking distance and are not suitable to evaluate parameters like gait variability which require the execution of several strides [110]. Due to these characteristics, they are unsuitable for use in clinical settings that are subject to space, time, and cost constraints [62, 109].

Alternative motion capture solutions exist nowadays that do not require the placement of markers in the body. Microsoft KinectTM, originally designed as a video gaming device, is a prominent example of such type of solution. Although it is more practical, it also has some limitations, for instance, the inability to capture the movement of joints not directed to the camera (e.g., due to

occlusion), or the constrained space within which it can record movements. Moreover, it relies on advanced algorithms for the proper tracking of the human skeleton, background removal, and identification of joint positions, which are prone to failure and frequently a source of movement evaluation error. Kinect has been used in many different contexts, including movement-based game interaction, and rehabilitation [111–114].

Electronic walkways have embedded pressure-sensitive switches that allow faster gait analysis without all preparation and post-processing needed in traditional optical motion capture systems. GAITRite is the most commonly used electronic walkway. With a few meters in length, it is able to identify foot contacts and measure several spatiotemporal parameters of gait. However, these systems can only measure some parameters, excluding the analysis of swing and other dynamic information [115, 116].

Portable balance boards, such as the well-known Nintendo Wii™ Balance Board (©Nintendo, Kyoto, Japan), provide a low-cost alternative to laboratory-based force platforms. They use force transducers from which the center of pressure position can be estimated. Balance boards are primarily used for balance assessment, offering a very limited ability to measure other movements [116].

Table 2.1 outlines the advantages and disadvantages of the most commonly used non-wearable systems for movement analysis.

2.1.2.2 Wearable Systems

Wearable systems enable long-term analysis outside laboratory settings. They are usually cheaper than non-wearable solutions and do not limit walking conditions [116]. Pressure insoles, for instance, can be integrated into the shoes to measure changes in pressure between the foot and the ground. However, unlike force platforms, they cannot distinguish the directions of the forces. They allow the measurement of temporal gait variables but provide no information on the dynamic movement observed during the swing, being also restricted to certain types of movements [116]. In contrast, inertial sensors are the most widely used type of wearable system for movement evaluation, as they are not restricted to a specific type of movement. They allow the measurement of movement in any body part, including, spatiotemporal gait parameters, and joint and segment angles of the lower and upper limbs [116]. The advantages and disadvantages of these systems are summarized in Table 2.2.

Inertial Sensors

Given their versatility and potential in the context of movement analysis, inertial sensors stood out from the alternative wearable and non-wearable solutions. They became highly attractive in the past years and constitute a promising approach to assist clinical decision-making for gait-related disorders in aging [105, 118]. Due to their compact size, low cost, portability, and low energy requirements, inertial sensors are well suited to measure human movement [119, 120], for instance,

Table 2.1: Advantages and disadvantages of non-wearable systems for movement analysis.

Instrument	Advantages	Disadvantages
Marker-based optical motion capture systems [62]	• Widely used for research purposes • Very accurate • Can capture complex movements, including full-body motion	• Restricted to the lab • Requires markers to be placed on the subject • Long set-up time • Difficult to operate • Long time to analyze the data • High cost • Limited path of walking
Markerless motion capture (e.g., Kinect™) [113]	• Does not require markers • Does not interfere with normal movement • Easy set-up and use • Low-cost • Portable	• Limited acquisition area • Affected by movement occlusion • Subject to errors in movement evaluation
Electronic walkways (e.g., GAITRite) [115]	• Does not interfere with normal movement • Portable	• May not capture movements other than walking • No dynamic information about the swing phase • Limited resolution of sensing points • Limited path of walking • Only suitable to the laboratory setting
Force platforms [117]	• Can provide information about balance and postural control • Widely used in gait analysis (kinetics)	• Can only capture movements performed on the platform • Lack of portability • Multiple repetitions needed until only one leg hits the plate (for gait analysis) • Contact times depend on a threshold • High cost
Balance boards (e.g. Wii™ balance board) [116]	• Can be used to assess balance and postural control • Portable • Low-cost	• Restricted to balance evaluation • Cannot evaluate gait

to recognize human activities, measure biomechanical parameters (e.g., joint angles, spatiotemporal gait metrics), or track pedestrian trajectories [119]. Inertial sensors have been used in older adults [121], neurological populations [122–124], Parkinson’s disease [121, 125], multiple sclerosis [126], foot drop [127], among others. They generally show good agreement with classical gait analysis for several clinically interpretable spatiotemporal gait parameters, supporting their use over an extended space and time and even outside the laboratory in real-life conditions [62]. The literature also supports their use in people with mobility impairments, being the preferred technology to monitor everyday life motor activities in people with altered movement patterns [120].

Unlike stationary systems used in laboratory settings, inertial sensors are placed on the body, enabling unobtrusive movement analysis outside the laboratory. Nowadays, inertial sensor-based movement analysis is performed using three-axis accelerometers and/or three-axis gyroscopes,

Table 2.2: Advantages and disadvantages of wearable systems for movement analysis.

Instrument	Advantages	Disadvantages
Pressure insoles and footswitches [108]	• Portable, not restricted to the lab	• Sometimes uncomfortable • Difficult to properly position them on the shoes • Less effective for uneven surfaces • Do not provide dynamic information about the swing phase of walking
Inertial sensors [62]	• Low cost • Small size • Light weight • Low consumption • Portable, not restricted to the lab • Little subject preparation • Wide range of applications	• Sensor uncertainty • Complex error handling mechanisms are required

made with Micro-Electro-Mechanical Systems (MEMS) technology. The assembly of both devices is typically called an Inertial Measurement Unit (IMU) [62, 128].

The functioning of accelerometers is based on the fundamental principles of Newton's laws of motion. Newton's first law of motion, also called the principle of inertia, states that "a body with no net force acting on it will either remain at rest or continue to move with uniform speed in a straight line, according to its initial condition of motion". Newton's second law of motion, also called the law of acceleration, states that "when a force is applied to the body, it will accelerate in the same direction of the force". Accelerometers are thus composed of a suspended mass inside a casing that moves in response to acceleration, generating an electrical signal that is proportional to the acceleration (Figure 2.2a). Accelerometers measure static (gravity) and dynamic (linear movement) acceleration. At rest, they will only measure the Earth's gravitational acceleration, which is approximately 9.81 m/s^2 [70, 129]. Gyroscopes take advantage of the Coriolis effect to measure angular velocity: a vibrating structure (proof-mass) is suspended in a way that Coriolis forces can be sensed as the mass undergoes a rotation relative to the inertial space. The underlying physical principle is that a vibrating object tends to continue vibrating in the same plane even if its support rotates. The Coriolis effect causes the object to exert a force on its support, and by measuring this force the rate of rotation can be determined (Figure 2.2b) [130]. Accelerometers and gyroscopes provide complementary information that allows evaluation of linear and angular movement [128]. Bosch[®] BMI160 is an example of an IMU combining an accelerometer and a gyroscope that was used in our experiments (Figure 2.2c).

Sensors can be placed in multiple body positions, including, the feet/shoes [131–133], lower trunk [134], shank [108] and ankles [135]. Both the accelerometer and the gyroscope are subject to errors, and specific techniques are commonly applied to mitigate these errors. For instance, the angular displacement of the device is typically calculated through the numerical integration of the angular velocity data acquired by the gyroscope. Small measurement errors in velocity accumulate over time during integration, resulting in progressively increasing errors in the computed displacement [136]. This is known as integration drift. Also, the noise inherent to the accelerom-

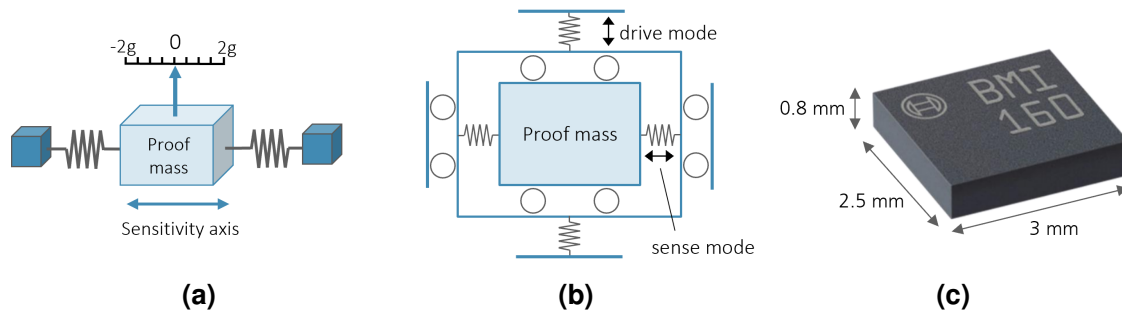


Figure 2.2: Inertial measurement unit working principles. (a) Model of a capacitive accelerometer, (b) Two-dimensional model of a Coriolis vibratory gyroscope, and (c) Bosch® BMI160 IMU model used in our experiments.

eter and the difficulty in isolating gravity from linear acceleration due to movement constitute a challenge [137].

Inertial sensors placed on the feet allow the complete evaluation of spatiotemporal gait parameters, being currently accepted as one of the best positions to evaluate gait [138–140]. Parameters such as the stride length, gait speed, cadence, swing width, and foot clearance, are important to assess walking performance in older adults [141] and can be obtained from this position [131, 132, 142, 143].

Despite significant efforts toward mobile gait assessment, the problem of accurately measuring gait parameters using inertial sensors is not yet completely solved [62, 123, 144]. Even when sensors are placed on the foot, errors may occur, necessitating the implementation of intelligent error-handling mechanisms [107, 132, 145, 146]. Additionally, algorithms and error correction mechanisms often rely on assumptions about movement characteristics that may not apply to all subjects and conditions [144, 147]. Hannink et al. [144] have noted that current solutions’ precision is still inadequate to discern various clinically significant phenomena, such as the decline of stride length due to the progression of neurological disorders. Thus, the problem of detecting clinically relevant changes in gait resorting to inertial sensors remains an open area of research [62, 123, 144].

2.2 Aging and Cognition

Modern medicine and technology enable us to maintain relatively good health until the very late stages of our life. However, the apparent cognitive decline remains one of the most central aspects of aging [148]. The term “cognition” is broadly used in psychology to refer to thought and multiple related processes [12]. As individuals get older, they may experience subtle changes in their cognitive ability. Though all brains age, the course of cognitive aging varies among older individuals and can potentially impair their independence, well-being, and overall quality of life [149, 150].

The aging process affects both gray and white matter regions of the brain, resulting in a decrease in overall brain size [151, 152]. Initially, it was believed that the reduction in gray matter volume was due to neuronal loss. However, recent research has linked the age-related cognitive

decline to the loss of synapses related to changes in the structure of neurons and in the functioning of neurotransmitter receptors, without significant neuronal loss [150, 151]. Although these changes may impact daily function, it is important to note that synapse loss may be reversible, whereas neuron death commonly associated with neurodegenerative diseases is not [150, 153].

Cognitive performance is typically conceptualized in terms of domains of functioning, although a complete consensus is lacking on the definition of these domains. The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines six basic neuropsychological domains: learning and memory, perceptual-motor function, language, executive function, complex attention, and social cognition [12]. Each domain has its own sub-domains, as shown in Figure 2.3.

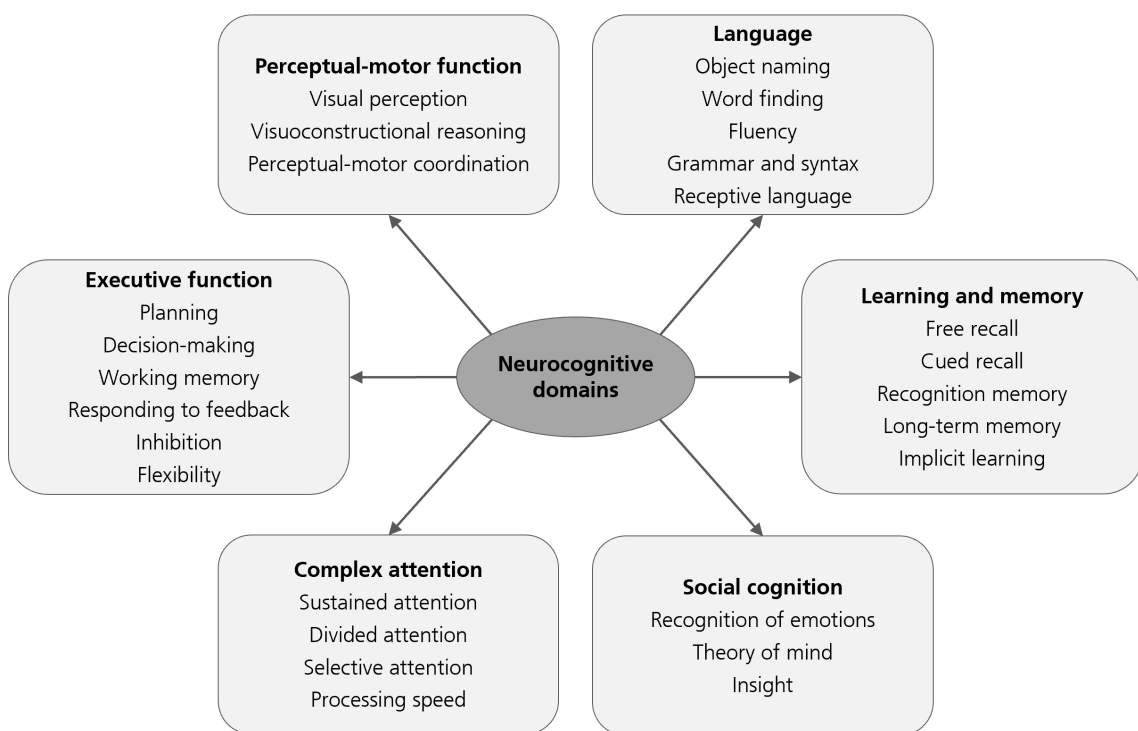


Figure 2.3: Neurocognitive domains (adapted from [12]).

The DSM-5 also provides examples of symptoms and observational criteria for each cognitive domain, as well as objective assessment examples. However, it refrains from endorsing any proprietary tests [12]. Table 2.3 presents a summary of functional limitations that are often associated with impairment in various cognitive domains.

Cognitive domains are interconnected and work together to carry out everyday activities such as solving problems, planning, thinking abstractly, learning, or making choices [150]. Although some cognitive abilities may decline with age, older adults can still maintain high levels of competency in daily living [150]. The brain utilizes cognitive reserve and neural plasticity to adapt to age-related or disease-related changes [154]. Individuals with a greater cognitive reserve can better tolerate brain changes without experiencing significant cognitive decline. Several factors can contribute to enhancing cognitive reserve, such as maintaining a healthy diet, limiting alcohol

Table 2.3: Functional limitations associated with impairment in different cognitive domains (adapted from [11]).

Cognitive domains	Examples of changes in everyday activities
Complex attention	Tasks take longer, especially when there are competing stimuli; individuals get easily distracted; tasks need to be simplified; difficulties holding information in mind to do calculations or dial a phone number.
Executive function	Difficulty with multi-stage tasks, planning, organizing, multi-tasking, following directions, and keeping up with shifting conversations.
Learning and memory	Individuals have difficulty recalling recent events, repeating themselves, misplacing objects, losing track of actions already performed, and increasing reliance on lists and reminders.
Language	Word-finding difficulty, use of general phrases or wrong words, grammatical errors, difficulty with comprehension of others' language or written material.
Perceptual-motor function	Getting lost in familiar places, more use of notes and maps, difficulty using familiar tools and appliances.
Social cognition	Disinhibition or apathy, loss of empathy, inappropriate behavior, and loss of judgment.

consumption, engaging in social activities, participating in intellectual activities, and remaining physically active [150, 151].

2.2.1 Cognitive Impairment

The DSM-5 provides a framework for the classification of neurocognitive disorders based on three syndromes: delirium, mild neurocognitive disorder, and major neurocognitive disorder [12]. Delirium is defined as a disturbance in attention and awareness that develops over a short period of time (typically hours to days) and can be caused by an underlying medical condition, substance intoxication or withdrawal, or a combination of these factors [12]. In contrast to delirium, mild and major neurocognitive disorders do not have disturbances in attention as a core feature, and individuals usually retain awareness of their environment. Both disorders share a common characteristic of a decline in one or more key cognitive domains from a prior level of functioning. They represent distinct diagnostic constructs that are based on a continuum of cognitive impairment ranging from normality to severe cognitive deficits and functional impairment [12].

Mild neurocognitive disorder, also referred to as Mild Cognitive Impairment (MCI), is characterized by a modest cognitive decline (beyond what is expected from normal cognitive aging, typically falling below 1–2 SD of the normative mean on neuropsychological tests) in one or more cognitive domains that do not compromise the ability to perform activities of the daily living [12]. The term was developed to describe the long pre-dementia stage in neurodegenerative diseases, resulting from advancements in early diagnosis, and the increasing focus on early intervention to prevent progression to dementia [12].

The term MCI has been traditionally used to characterize the transitional zone between normal cognitive function and probable dementia due to Alzheimer's disease (AD). However, it is now recognized as a transitional state for other possible etiologies, such as other degenerative conditions, as well as vascular, metabolic, traumatic, or psychiatric causes [155]. In 2004, an International Working Group on MCI proposed further classifying MCI into Amnesic Single Domain (only memory is impaired), Amnesic Multiple Domain (memory and other cognitive domains are impaired), Non-Amnesic Single Domain (a single cognitive domain other than memory is impaired), and Non-Amnesic Multiple Domain (multiple cognitive domains—other than memory—are impaired) [13, 155]. The diagnostic criteria for MCI proposed by the International Working Group are very similar to the concept of mild neurocognitive disorder proposed by DSM-5 [13, 155].

The term major neurocognitive disorder is proposed in DSM-5 as an alternative to the concept of dementia [12]. The key distinction between mild and major neurocognitive disorders relies on the criterion of independent functioning. Dementia is therefore characterized by a significant cognitive decline (typically 2 SD above the normative mean in neuropsychological testing) in one or more cognitive domains that interfere with independence in everyday activities [12].

Mild and major neurocognitive disorders can be subtyped according to their etiology, as shown in Figure 2.4a. The main subtypes of dementia include Alzheimer's disease (AD), vascular dementia (VaD), dementia with Lewy bodies (DLB), and frontotemporal dementia (FTD). AD is the most common form of dementia, accounting for at least 50% of the dementia cases [156]. Vascular dementia is the second most common cause of dementia [157–159] (Figure 2.4b).

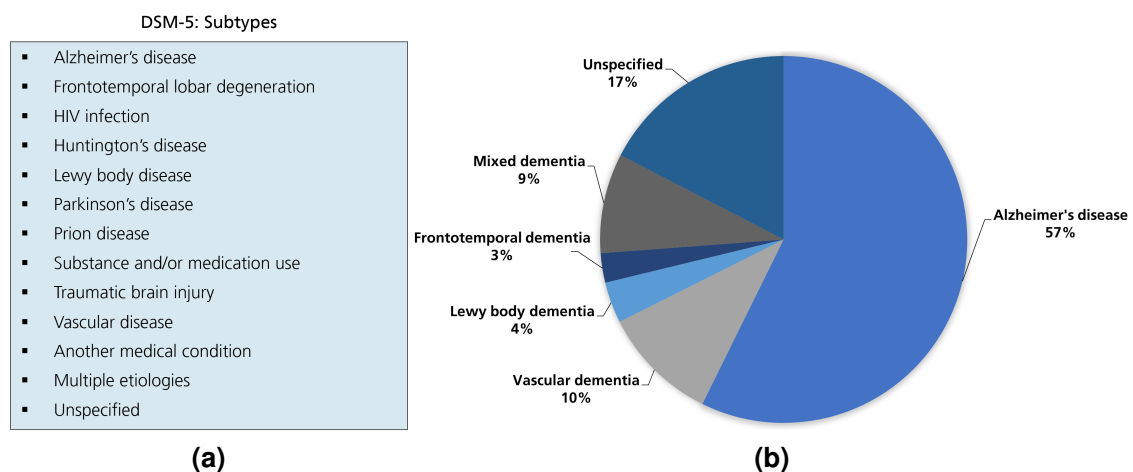


Figure 2.4: Subtypes of mild and major neurocognitive disorders. **(a)** Etiological subtypes, according to the DSM-5 [12] and **(b)** prevalence of dementia subtypes, according to [160].

Neurocognitive disorders are often associated with a causative disorder, such as Parkinson's disease, Huntington's disease, traumatic brain injury, HIV infection, or stroke. However, in some cases, the cognitive and behavioral symptoms are the initial manifestations, and the underlying cause is revealed through the longitudinal course of the disorder, which may include etiologies such as AD, VaD, DLB, or FTD [12, 161]. While differentiating among these multiple types of

neurocognitive disorders can be challenging, they become more evident at the level of major neurocognitive disorder than at the level of mild neurocognitive disorder [161]. Differential diagnosis may involve a detailed neuropsychological assessment, genetic testing, cerebrospinal fluid (CSF) analysis, and structural neuroimaging using techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) [12].

Currently, there are no treatments capable of reversing or stopping the progression of dementia, although its evolution may be slowed down [18]. Much effort is therefore being made to improve early detection when treatments could be potentially more effective in slowing or preventing disease progression [2].

2.2.2 Screening Instruments

Differentiating normal from abnormal cognitive decline is one of the central questions of current research. Neurocognitive disorders have a huge impact on people's lives, and early detection and intervention are key to slowing down or preventing progression. Growing evidence supports that maintaining a healthy lifestyle, both physically and mentally, can decrease the rate of cognitive decline associated with aging, and potentially delay the onset of cognitive symptoms in age-related diseases [151, 152].

Clinical assessments for cognitive impairment typically begin with a concern raised by the individual or a reliable informant [11, 13, 155, 162]. However, cognitive symptoms are often dismissed by patients and families as a normal part of aging and are thus underestimated. On the other hand, some may overemphasize minor changes and express fears of developing dementia, which can influence the clinician's impressions [11]. Standardized and objective assessments are thus required to accurately diagnose cognitive impairment.

The objective assessment of cognitive performance requires the administration of neuropsychological tests, interpreted in accordance with normative data appropriate for the patient's age, education, and cultural-linguistic background [11, 12, 163]. Cognitive testing is the primary way of establishing the severity of cognitive impairment and is therefore a necessary component of a neuropsychological evaluation [163]. The whole diagnostic process is time-consuming, expensive, and somewhat subjective, as it relies heavily on the clinician's interpretation [162].

When a comprehensive neuropsychological assessment is unavailable, objective testing can be achieved using a global screening scale, such as the well-known Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA), or the Addenbrooke's Cognitive Examination (ACE) [11]. These tests were developed to allow more objective and faster screening for cognitive impairment in clinical settings, for instance, in the context of primary care settings. These tests are final score-based and their sensitivity and specificity are defined according to pre-established cut-offs [162]. Table 2.4 provides an overview of existing tests for the screening of mild cognitive impairment.

The MMSE is widely used as a screening tool for dementia [22], but it has limited sensitivity for detecting MCI [165]. In contrast, the MoCA was developed specifically to detect cognitive

Table 2.4: Commonly used cognitive screening tests for mild cognitive impairment.

Test	Duration (min)	Sensitivity (%)	Specificity (%)	References
MMSE	5–10	65–85	71–85	[164] [165]
MoCA	10–20	81–97	60–86	[166] [22] [167] [168] [169]
ACE-III	15–20	75–92	63–92	[170] [171] [172]

MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; ACE-III: Ad-denbrooke's Cognitive Examination–III.

impairment at an earlier stage and has consistently shown superior psychometric properties compared to the MMSE, with higher diagnostic accuracy for discriminating between MCI and AD patients [173]. A recent systematic review found that the MoCA is the most commonly used and preferred tool for MCI screening in primary care, while the ACE-III is better for dementia screening [22]. Although the ACE-III takes longer to administer than the MMSE, it addresses additional neuropsychological domains and has excellent diagnostic utility in specialist settings [174]. According to Roeck et al. [175], the MoCA is the most suitable instrument for detecting MCI and AD in population-wide screening.

Screening tests require clinicians to interpret the results and are thus somewhat subjective [162]. A recent study found that standardization of test administration is often not achieved in practice due to variations in the way the clinicians interact with the patients to introduce the test and implement it. These variations affect the way the patient understands the tasks, can create confusion and anxiety, and ultimately affect test results [176]. The results are also influenced by demographic factors, such as age, gender, or education, and, thus, standardized normative data are required for comparison. Additionally, the cultural characteristics of the participants require test adaptation and translation for use in different countries [162]. Besides the administration times presented in Table 2.4, the time needed for scoring the test also needs to be taken into account. Although some of the tests offer one or two alternative versions, administering the same questions repeatedly can limit their utility [23].

2.3 Linking Cognition and Movement

Until recently, healthcare providers and researchers conducted separate assessments for gait and cognitive functions when evaluating older adults. However, recent studies have shown that gait and cognition are interrelated in older adults. Contrary to previous beliefs held for several years, poor mobility is associated with poor cognitive function [177–179].

It is common for older adults to experience both cognitive decline and gait abnormalities, which have been linked to disabilities, loss of independence, falls, and even mortality [49, 74]. However, approaching cognition and motor functions as separate entities hinders our understanding of the shared underlying mechanisms and limits the potential for integrated preventive strategies and treatments. For this reason, several research groups have recently investigated the interrelationships between cognition and motor functions [84]. Evidence has been derived from imaging

studies, observational studies, and interventional studies [3]. The most important contributions to these topics are summarized next.

2.3.1 Common Underlying Mechanisms

Gait is not merely an automatic motor task, but rather, it requires input from higher cognitive areas [46, 180, 181]. Associations have been found between gait and multiple cognitive domains, including executive functions [44, 179, 182], memory [178, 183, 184], processing speed [179, 183, 185], attention [183, 185–188], language [189] and visuospatial skills [186]. Morphological and functional neuroimaging studies have offered additional evidence supporting the relationship between gait and cognitive resources [42, 190, 191]. Hence, a change in gait pattern may indicate possible structural damage in the brain, which could also be responsible for a concurrent decline in cognitive function [84].

Motor and cognitive decline share a number of pathophysiological pathways. Although these mechanisms are not yet entirely understood, several potential causes or pathways may be involved in both dysfunctions, as illustrated in Figure 2.5.

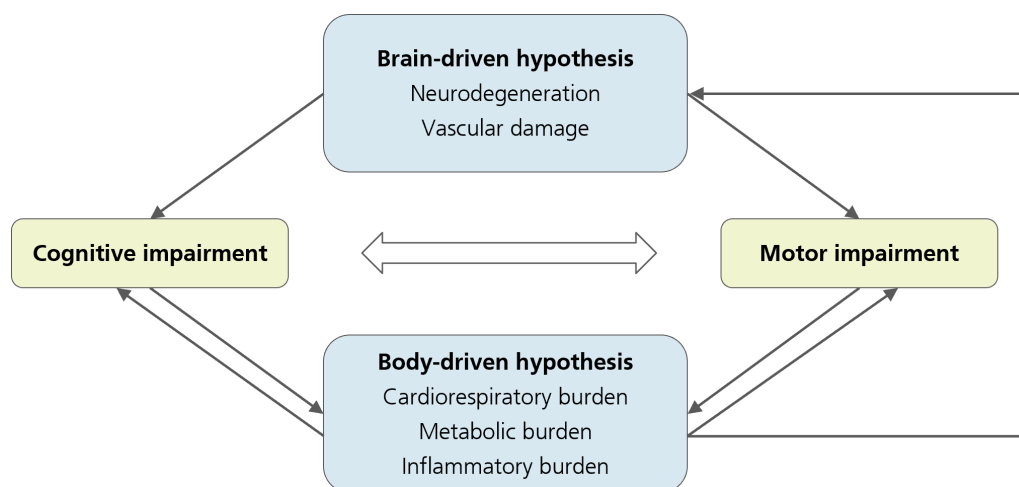


Figure 2.5: Determinants of cognitive and motor impairment (adapted from [84]).

According to the brain-driven hypothesis, greater neurodegeneration or vascular damage is responsible for both cognitive and motor dysfunction [84]. For example, individuals with MCI or AD often have slower and more variable gait compared to those without cognitive impairment [74]. On the other hand, the body-driven hypothesis suggests that higher levels of inflammation, metabolic dysfunction, and cardiorespiratory burden outside of the central nervous system may also contribute to declines in both cognitive and motor function [84]. For instance, heart failure has been considered a risk factor for motor and cognitive decline. Older adults often experience multiple coexisting diseases, which can contribute to a higher burden associated with accelerated motor and cognitive decline [84].

2.3.2 Relationship with Gait

The term “Motoric Cognitive Risk Syndrome” (MCR) refers to the co-occurrence of slow walking speed, defined as one standard deviation below age- and sex-specific average gait speed, and self-reported (subjective) cognitive impairment [49], as illustrated in Figure 2.6. Numerous cohorts and populations worldwide have been studied to determine the prevalence of MCR, which has been found to range from 2% to 27% [192]. The MCR was shown to increase the risk for further cognitive decline and dementia, also contributing to other negative age-related outcomes. Unlike MCI, MCR can be diagnosed without formal neuropsychological assessments, making it a more accessible and cost-effective screening option for use in primary care settings [50, 192].

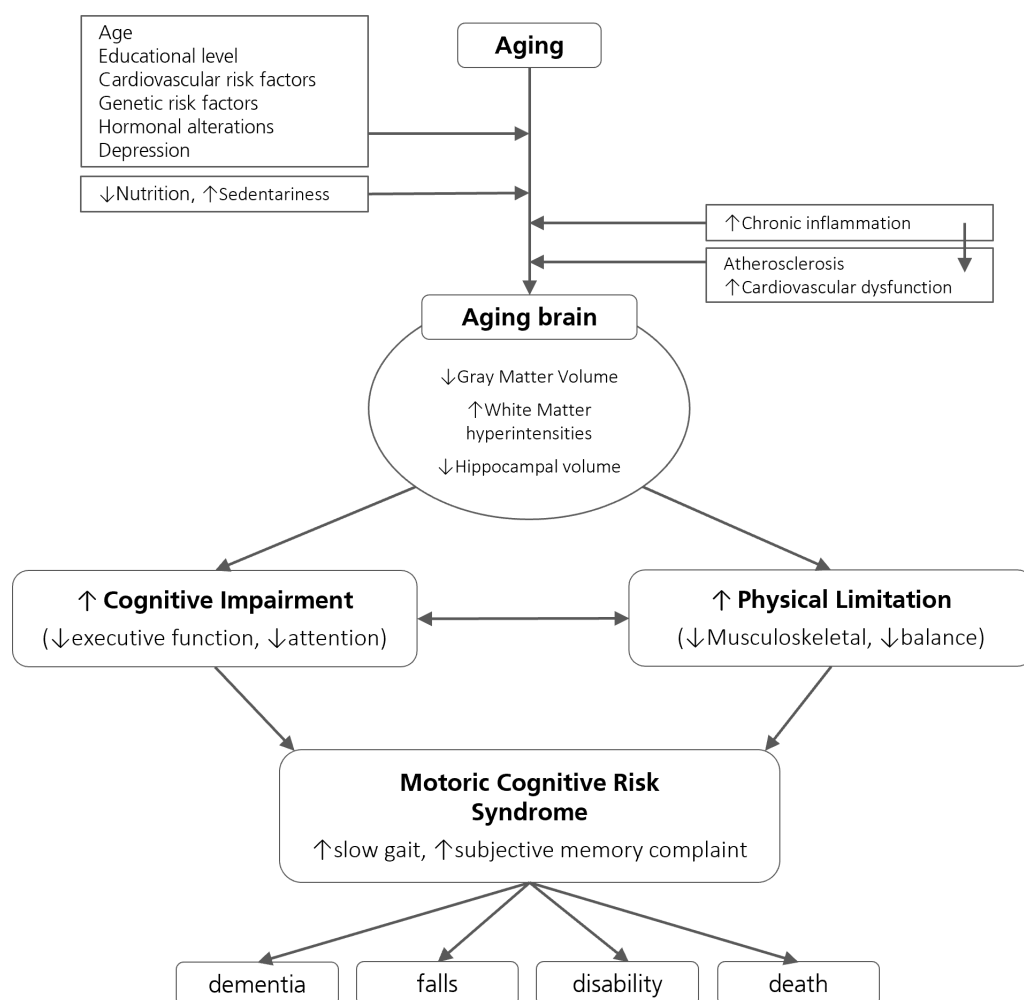


Figure 2.6: Schematic representation of the motoric cognitive risk syndrome (adapted from [49]).

Dementia is often accompanied by lower mobility, which can greatly affect an individual’s independence and quality of life [49]. Studies have shown that the severity of cognitive impairment is directly linked to increased gait abnormalities. For example, a systematic review conducted by van Iersel et al. [193] revealed that individuals with dementia exhibit a decrease in walking

speed compared to healthy controls and that this decrease worsens with the progression of dementia severity. In addition, De Cock et al. [194] found statistically significant differences in gait characteristics across different stages of cognitive impairment. Bahureksa et al. [53] conducted a meta-analysis that showed that gait parameters such as speed, stride length, and stride time were effective in distinguishing between older adults with MCI and healthy controls. In another meta-analysis, Beauchet et al. [39] found that higher stride time variability was related to both MCI and dementia. Moreover, higher stride time variability was linked to episodic memory impairments and executive dysfunction. The authors concluded that higher stride time variability is a motor phenotype of cognitive decline [39].

Longitudinal studies have shown that gait slowing can even precede cognitive impairment in older adults by 7 to 12 years [37, 195, 196]. According to Dumurgier et al. [195], the decrease in gait speed can occur up to 7 years before the clinical onset of dementia, while Buracchio et al. [37] found that motor decline could already be observed 12 years before the diagnosis of MCI. While gait speed at baseline was considered a risk factor for cognitive decline, baseline cognition did not predict changes in gait speed [93]. Kikkert et al. [197] reviewed longitudinal studies and concluded that gait slowing (0.68–1.1 m/s) precedes cognitive decline and the development of dementia. In a meta-analysis, Beauchet et al. [198] provided additional evidence that poor gait performance predicts the future occurrence of dementia. Taken together, these findings suggest that changes in gait performance may serve as early markers of cognitive decline and the development of dementia, and underscore the importance of monitoring gait performance in older adults.

2.3.3 The Dual-Task Paradigm

Walking is governed by automaticity, a defining characteristic of healthy and well-functioning adults. Automaticity in walking refers to the ability of the nervous system to regulate routine walking without relying heavily on attention-demanding executive control resources. Research suggests that impairments in walking can arise from a shift in locomotor control from healthy automaticity to compensatory executive control. This shift can be detrimental to walking performance, as the executive control approach is not optimized for locomotor control and depletes a limited pool of executive resources. As a result, it impairs the capacity to perform basic and complex walking tasks and increases the risk of negative mobility outcomes, such as falls. The loss of automaticity can interfere with walking performance while multitasking, resulting in a “cost” of performing two tasks simultaneously. Dual-task assessments can be employed to differentiate between automatic and executive control in walking [199].

Dual-task interference occurs when the performance of one or both tasks deteriorates as a result of attempting to perform two different tasks simultaneously. Cognitive-motor interference is a type of dual-task interference that occurs when a motor task, such as walking, is combined with a cognitive task, like counting backwards [200, 201]. The change in performance associated with dual-tasking is referred to as dual-task cost or dual-task effect (DTE) [200] and can be calculated

as shown in the following equation:

$$DTE(\%) = \frac{\text{dual task variable} - \text{single task variable}}{\text{single task variable}} \times 100 \quad (2.1)$$

where a minus sign may be added to the numerator to ensure that a negative DTE indicates performance deterioration [200].

The framework proposed by Plummer et al. [200] classifies the dual-task interference into nine possible patterns. According to this framework, the pattern of dual-task interference can be classified by plotting the gait and cognitive DTE against one another, as illustrated in Figure 2.7. The nine patterns arise from the three possible interference patterns observed in gait and cognition under dual-task, which can be categorized as having no effect, an interference (detrimental) effect, or a facilitation (positive) effect. For instance, in the mutual interference scenario, both tasks, gait and cognition, deteriorate under dual-task [200]. The accurate interpretation of the dual-task interference thus requires measuring both tasks in both single and dual-task conditions [200]. However, most studies in the literature focus on the analysis of the dual-task effect on gait, neglecting its effects on the cognitive task [54, 200, 202].

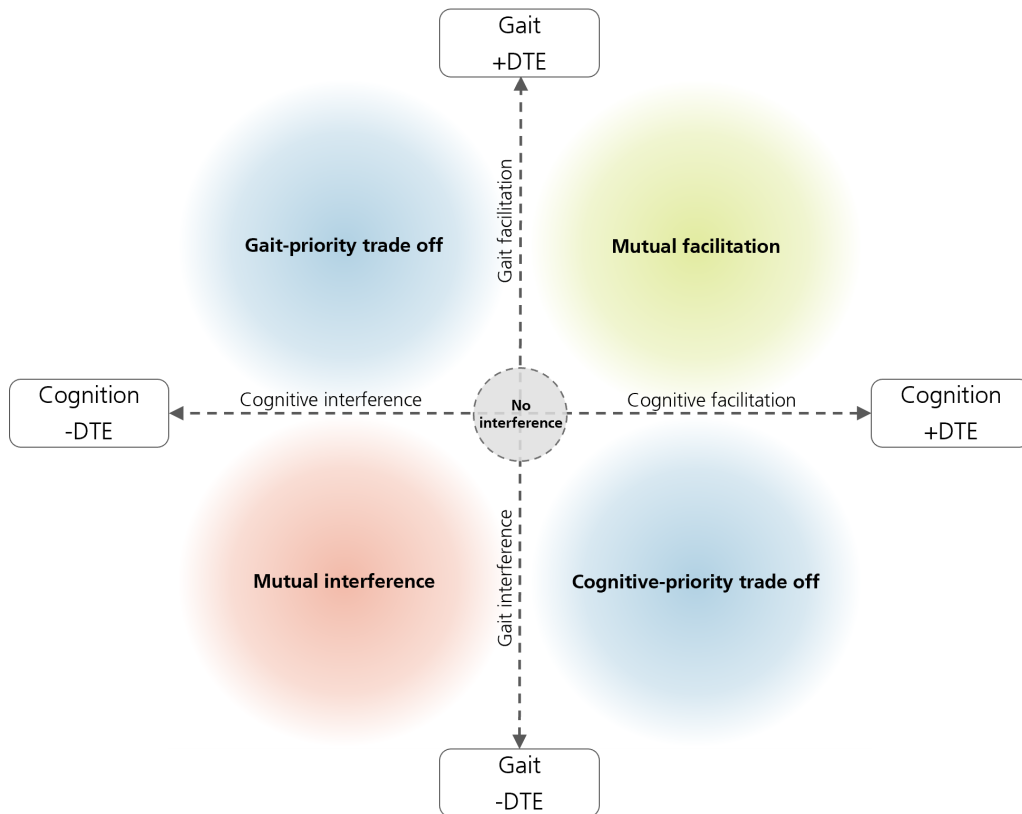


Figure 2.7: Dual-task interference framework (adapted from [200]).

Research has demonstrated that cognitive impairment affects gait performance under dual-task conditions such as when walking while spelling words backward, doing serial subtractions, or naming animals. The use of the dual-task paradigm could help distinguish older adults with

MCI from those with preserved cognition [57, 203–206], revealing gait impairments that would otherwise be invisible [36, 51, 52, 203]. The results of a meta-analysis indicated that dual-task assessments of gait performance, including gait variables such as gait speed, stride length, and stride time, allow better discriminatory ability than the same parameters under single-task conditions [53]. In a more recent meta-analysis, the effects of the dual-task on gait speed were shown to be highly sensitive for the detection of MCI [52]. Dual-task effects on other gait parameters could also differentiate older adults with cognitive impairment [55, 56]. Furthermore, dual-task gait performance has been associated with longitudinal progression to dementia in patients with MCI [207], could significantly differentiate patients with MCI from those with subjective cognitive complaints [55], and could distinguish between amnesic and non-amnesic MCI [51].

While the majority of studies have found that dual-task assessments can improve the differentiation between older adults with MCI and healthy older adults, there have been some inconsistent findings [202, 208, 209]. The contradictory results can be attributed to variations in instruments, protocols, and subject populations [202, 208]. For example, the insensitivity of current tools in detecting early signs of cognitive decline may result in a wrong stratification of the groups [54, 202]. Additionally, many studies fail to consider the effects of dual tasks on cognitive performance, making it difficult to assess the magnitude and direction of interference [54, 200, 202]. Furthermore, some studies rely on simplistic measures, such as a stopwatch, which cannot adequately capture differences in gait performance [208]. Finally, inappropriate concurrent tasks are often used to challenge motor-cognitive interaction, further complicating the interpretation of results [54, 205].

The selection of an appropriate cognitive task is a crucial consideration when using the dual-task approach, as it can affect the degree of dual-task interference. Tasks involving verbal fluency rely on semantic memory, which tends to remain relatively unimpaired in the early stages of AD. Conversely, mathematical subtraction tasks involve working memory, a function that is typically impaired early in the course of AD [57]. Studies have found that counting backward by 7s elicits greater gait disruption than the verbal fluency task in older adults and more disruption than counting backward by ones [203, 204, 210].

In addition to gait, other motor tasks have been explored to assess cognitive impairment. For example, Ehsani et al. [211] evaluated the upper-extremity function while participants counted backward by 1s or 3s. In their study, the dual-tasking with upper-extremity function was found to have better discriminatory ability than gait in identifying cognitive impairment in older adults [211]. The task of stepping in place while solving a cognitive challenge has also been explored showing some promising results [23, 212–214]. Mancioffi et al. [215] proposed two innovative motor-cognitive dual tasks, the forefinger tapping and the toe-tapping, to discriminate older adults with and without MCI. Their results showed good correlations with MMSE and high performances in identifying MCI.

The mechanisms underlying the dual-task interference are still unclear and remain a question of ongoing debate. Two major theories try to explain why dual-task interference occurs. The *serial bottleneck model* proposes that due to limitations in the brain's processing capacity, it can only effectively process one task at a time. Consequently, when two tasks require simultaneous

processing, a bottleneck arises, leading to a reduction in overall performance. Conversely, the *capacity-sharing model* suggests that multiple tasks can be processed in parallel, but only up to a certain limit. The interference between the tasks occurs when the demands of the tasks exceed the available processing resources [54, 200, 201].

Although dual-task assessment appears to be a promising solution for enhancing clinical diagnosis, much remains to be understood about the motor-cognitive interplay and its utility in identifying individuals with early signs of cognitive impairment [36, 54, 216]. Even though dual-task assessments are relatively inexpensive, non-invasive, and easy to administer, the integration of this assessment framework into clinical practice remains to be further investigated [200].

2.4 Automated Detection of Cognitive Impairment

The field of artificial intelligence (AI) has progressed rapidly in recent years, generating considerable interest in various domains, including healthcare. As dementia syndromes become an urgent concern, AI has the potential to assist in the creation of new affordable and accessible digital tools that can be used to improve the screening for cognitive disorders [162, 217].

Recently, various AI-based tests have been introduced to assist with the early detection of cognitive impairment, including computerized cognitive tests, computer-assisted interpretation of brain scans, analysis of movement, and speech [162]. AI allows automated analysis of the data, resulting in time and cost-efficient assessments. Moreover, AI-based tests can provide improved detection accuracy, due to their capability of capturing additional features from a larger amount of data. Compared to a manual analysis by clinicians, AI-based assessments can also be more objective [162]. Despite the promising potential of AI in improving cognitive screening, AI-based tests have yet to be widely implemented in routine clinical practice and remain primarily in the research setting [162].

In this section we provide a short overview of machine learning approaches for the detection of cognitive impairment, focusing on those that use movement and task performance evaluation metrics (including, dual tasks) to distinguish healthy from cognitively impaired older adults.

2.4.1 Machine Learning Approaches

Machine learning is a subset of AI that involves training algorithms to recognize patterns and relationships in data and make predictions based on unseen data. In supervised learning—a subcategory of machine learning—the algorithm learns from labeled data, meaning that the training data provided to the algorithm includes both the input data and the labels. The goal is to learn a mapping function (a model) that can predict the correct output for new, unseen (and unlabeled) input data. When the label is continuous, the problem is referred to as regression and can be used for instance to predict MMSE or MoCA scores. In contrast, when the labels are categorical, the problem is referred to as classification and can be used, for instance, to predict whether a person is “healthy” or “cognitively impaired”. This is typically done by learning decision boundaries that

separate the two classes based on features of the input data [217–220]. An overview of some of the most commonly used supervised learning algorithms is presented below.

- **Decision Tree** is a hierarchical tree-like prediction model represented as a set of if-then rules in which each node represents a test on a feature, each branch represents the outcome of the test, and each leaf node represents the class label [218, 219, 221, 222];
- **Random Forest (RF)** is an ensemble learning method that combines the outcomes of several individual trees trained on random subsets of the data and features [222];
- **Logistic Regression (LR)** estimates class membership probabilities using a logit-transformed linear regression [217];
- **Naïve Bayes** makes classification decisions based on the probability of each class given the input features, calculated using the Bayes’ theorem and the “naïve” assumption of conditional independence between features [218, 221, 222];
- **Support Vector Machine (SVM)** is a binary classifier that finds the hyperplane in the feature space that best separates the margin between data classes. It can handle non-linearly separable data using kernel methods [217–219, 221];
- **K-Nearest neighbors (KNN)** predicts the label or value of a new data point based on the majority label or value of its k closest instances in the feature space [219, 221];
- **Neural Networks (NN)** use a set of interconnected nodes (called neurons) organized into layers to learn complex patterns and relationships in the data by adjusting the weights and biases of the interconnected nodes. A deep neural network can have many hidden layers, typically ranging from a few to dozens or even hundreds of layers. Deep neural networks include specialized models like convolutional neural networks (CNNs) for image analysis, recurrent neural networks (RNNs) for sequence modeling, and long short-term memory (LSTM) networks for handling long-term dependencies in sequential data [217].

Most of these algorithms have been employed to detect cognitive impairment, some of them based on movement analysis. Unlike traditional cognitive evaluations, movement-based tests offer an opportunity to study the decline in cognitive functions using a method that is free of language, and thus, more universally accessible. However, this area of study is still in its early stages [162].

Tables 2.5 and 2.6 provide a summary of machine learning approaches that use movement and task performance evaluations to distinguish cognitively normal from cognitively impaired older adults. Table 2.5 aggregates the studies on gait analysis, whereas Table 2.6 presents studies in which other tasks and body movements have been explored. The surge of recent publications on this topic, with the majority of them appearing since 2019, underscores the growing interest and research attention devoted to this area. Although gait is the type of movement most commonly studied (Table 2.5), other types of movements, including arm movements and stepping, have also been investigated (Table 2.6). Various instruments have been utilized to assess movement, including inertial sensors, electronic walkways, and markerless or optical motion capture systems.

Table 2.5: Summary of machine learning approaches using gait tests to distinguish healthy controls from cognitively impaired older adults.

Ref.	Groups	Walking Test	Instrument	Model	Validation	Performance
Gwak (2018) [223]	CN (n=26) MCI (n=27)	Without cognitive load	Smartwatch	LR	5-fold CV	Acc = 86.0%
Gonzalo (2019) [224]	CN (n=74) MCI (n=28)	Without cognitive load	Optical MoCap	RF	LOOCV	Acc = 84.8%
Boettcher (2020) [225]	CN and SC (n=31) MCI (n=61)	Without cognitive load, while reciting the alphabet and subtracting 3s	Electronic walkway	SVM	80% training 20% test	Se = 82.0% Sp = 67.7%
Ricciardi (2020) [226]	Patients with PD: CN (n=22) MCI (n=23)	Walking while subtracting 7s	Optical MoCap	RF	LOOCV	Se = 85.3% Sp = 79.4%
You (2020) [227]	CN (n=35) MCI and AD (n=52)	Without cognitive load	Kinect	STGCN	75% training 25% test	Acc = 93.1%
You (2021) [228]	CN (n=35) MCI and AD (n=53)	Without cognitive load	Kinect	LSTM	75% training 25% test	Se = 92.0% Sp = 88.2%
Ghoraani (2021) [229]	CN (n=32) MCI and AD (n=46)	Without cognitive load, while reciting the alphabet and subtracting 3s	Electronic walkway	SVM	5-fold CV	F1 = 88.0% Acc = 86.0%
Jung (2021) [230]	MMSE ≥ 24 (n=60) 20 $\leq$ MMSE $\leq$ 23 (n=36) MMSE ≤ 19 (n=12)	Usual and fast pace	Inertial sensors (foot dorsum)	LSTM	70% training 30% test	Se = 97.5% F1 = 97.4%
Yamada (2021) [231]	CN (n=47) MCI (n=45)	Without cognitive load	Optical MoCap	SVM	10-fold CV (20x)	Se = 84.9% Sp = 69.8%
Shahzad (2022) [232]	CN (n=30) MCI (AD type) (n=30)	Walking while naming animals	Inertial sensors (shank)	SVM	LOSOCV	Se = 83.3% Sp = 56.7%
Seifallahi (2022) [233]	CN (n=80) MCI (n=70)	Walking while subtracting 1s	Kinect	ANFIS	None mentioned	Se = 90.0% Sp = 99.0%

CN: cognitively normal; SC: subjective complaints; ANFIS: adaptive neuro-fuzzy inference system; LOOCV: Leave-One-Out Cross-Validation; LOSOCV: Leave-One-Subject-Out Cross-Validation; STGCN: Spatio-Temporal Graph Convolutional Networks; Se: Sensitivity; Sp: Specificity; Acc: Accuracy.

2.4.2 Training and Evaluation

The process of training a machine learning model can be seen as an optimization problem that involves minimizing the training error. For that, a training set is required. What distinguishes machine learning from optimization is the requirement of the model to perform well not only on the training data but also on new and previously unseen inputs. This ability is referred to as generalization. To evaluate the generalization error of a machine learning model, we evaluate its performance on a test set comprising examples that are distinct from those in the training set. For that purpose, the data are usually split into two separate sets, the training set, and the test set. Because most machine learning algorithms have hyperparameters that affect their behavior, the training set is typically further split into two distinct subsets. One of these subsets is used to train the model by learning its parameters, while the other subset is used as a validation set to

Table 2.6: Summary of machine learning approaches using other task-oriented movements to distinguish healthy controls from cognitively impaired older adults.

Ref.	Groups	Test	Instrument	Model	Validation	Performance
Aoki (2019) [212]	MMSE $\geq$ 25 (n=88) MMSE $<$ 25 (n=15)	Stepping in place while subtracting 1s	Kinect	SVM	5-fold CV (200x)	AUC = 74.7%
Matsuura (2019) [23]	MMSE $>$ 23 (n=64) MMSE $\leq$ 23 (n=26)	Calculation, stepping, in place and dual task	Kinect and floor pressure sensor	NN	LOSOVC	Se = 75.3% Sp = 79.9%
Toosizadeh (2019) [234]	CN (n=35) MCI and AD (n=56)	Elbow flexion while subtracting 3s	Inertial sensors (arm and wrist)	LR	10-fold CV	Se = 80.0% Sp = 75.0%
Ehsani (2020) [235]	CN (n=35) MCI and AD (n=56)	Elbow flexion while subtracting 1s	Inertial sensors (arm and wrist)	LR	10-fold CV	Se = 82.0% Sp = 85.0%
Wu (2021) [213]	MMSE $>$ 27 (n=22) MMSE $\leq$ 27 (n=16)	Calculation, stepping, in place and dual task	Kinect and floor pressure sensor	LSTM	LOSOVC	Se = 94.0% Sp = 52.0%
Wu (2021) [214]	MMSE $>$ 27 (n=118) MMSE $\leq$ 27 (n=53)	Calculation, stepping, in place and dual task	Kinect and floor pressure sensor	STGCN	None mentioned	Se = 92.0% Sp = 84.0%
Konstantinidis (2021) [236]	CN (n=38) MCI (n=64)	Aerobic, resistance, flexibility, and balance exergames	Balance board	NN	10-fold CV	Se = 73.4% Sp = 68.4%

CN: cognitively normal; LOOCV: Leave-One-Out Cross-Validation; LOSOCV: Leave-One-Subject-Out Cross Validation; STGCN: Spatio-Temporal Graph Convolutional Networks.

evaluate the model’s performance under various hyperparameter configurations. The test set is kept separate and only used at the end of training to assess the model’s ability to generalize to unseen data [217].

When the test set is too small, the estimated test error can be statistically uncertain, and alternative splitting methods are necessary. One such method is k -fold cross-validation, which involves dividing the dataset into k non-overlapping subsets. The model is trained on $k - 1$ subsets and the remaining subset is used to estimate the test error. The process is repeated k times, each time using a different subset for testing. The overall test error is then estimated by computing the average of the test errors across the k trials [217–220]. Leave-one-out cross-validation (LOOCV) is a special case of k -fold cross-validation in which each test subset consists of an individual sample or instance so that the number of test sets is equal to the number of samples in the dataset. The method can be computationally expensive but provides more accurate estimates of the model’s performance when the dataset is too small. In leave-one-subject-out cross-validation (LOSOVC), the data are split by leaving out all samples from a single person for the test set and using the remaining data for training. This method is useful in cases where the data come from multiple people or subjects, and we want to ensure that the model generalizes well to new individuals [217–220].

Several metrics can then be used to evaluate classifier performance. The cases in which the model predicted a positive class (e.g., “cognitively impaired”) and the actual class was also positive are considered the true positives (TP). The true negatives (TN) will correspond to the negative cases (e.g., “cognitively normal”) correctly classified as such. False positives (FP) and false negatives (FN) correspond to cases incorrectly classified as positive and negative. These four outcomes can be used to calculate the evaluation metrics in Equation 2.2 to 2.7.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (2.2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2.4)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.5)$$

$$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (2.6)$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \quad (2.7)$$

Accuracy (Equation 2.2) is commonly used to evaluate the performance of machine learning models, as it measures the percentage of all predictions that are correct. In health-related problems, a further distinction is required concerning the behavior of models on positive and negative classes. For that purpose, *Sensitivity* (Equation 2.3), also called recall or true positive rate, can be used to measure how well a model identifies the true positive cases among all the cases that actually have a condition. *Specificity* (Equation 2.4), or true negative rate, on the other hand, can be used to measure how well a model identifies true negative cases among all the cases that do not have the condition. A highly sensitive test is required for the effective screening of a certain condition or disease [237]. A test with high specificity will correctly identify most of the people who do not have the disease as negative. *Precision* (or positive predictive value) is used to quantify the proportion of true positives among all predicted positives (Equation 2.5). In binary classification, it is also common to report the *F1-score*, defined as the harmonic mean of precision and sensitivity (Equation 2.6). F1-score is used as a single measure of performance for the positive class. When a considerable imbalance exists between positive and negative samples in a dataset, balanced accuracy can be used to better describe classification performance. *Balanced accuracy* (Equation 2.7) is obtained by averaging sensitivity and specificity.

The Receiver Operating Characteristic (ROC) curve is a graphical representation that shows how well a model is able to distinguish between the positive and negative classes across all possible threshold values. The curve is generated by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings for the model (Figure 2.8). The points near the upper left corner of the ROC curve represent better thresholds with higher sensitivity and specificity. The diagonal line represents a classifier that performs no better than random guessing. The area under the ROC curve (AUC) summarizes the overall performance of the classifier, where 1 indicates perfect classification performance and 0.5 indicates random guessing [219]. ROC and AUC take into account the trade-off between true positives and false positives, which can be particularly important in situations where the cost of false positives and false negatives is different [238]. In medical diagnosis, for example, a false negative (missed diagnosis) can delay necessary interventions that can have serious negative consequences on patients' health. On the

other side, having too many false positives (false alarms) can result in high costs, discomfort, and additional risks due to unnecessary referrals to specialist areas for further evaluation [239]. Finding the right balance between sensitivity and specificity in medical testing is critical to ensuring accurate and timely diagnoses while avoiding unnecessary harm and expenses.

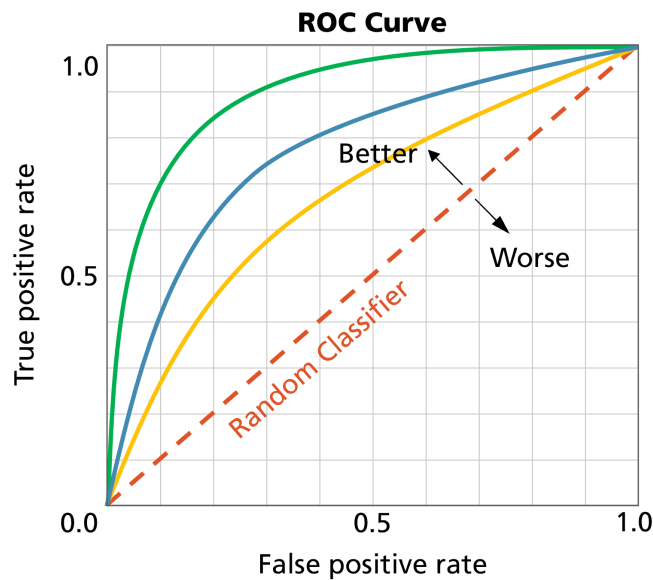


Figure 2.8: Receiver Operating Characteristic (ROC) curve.

2.5 Concluding Remarks

Recognizing changes in walking patterns as early signs of declining health and disability is crucial for promoting healthy aging and maintaining the quality of life. However, gait changes can be subtle and difficult to detect through clinical observation alone. Consequently, thorough investigations incorporating quantitative methods are necessary to identify underlying causes of gait changes. Unfortunately, the utilization of quantitative gait analysis in clinical settings remains limited due to the high cost and complexity associated with existing motion capture solutions.

While wearable solutions like inertial sensors have shown promise, their implementation in clinics or daily living still faces practical considerations. Although inertial sensors are attractive due to their compact size, low cost, portability, and low energy requirements, accurately measuring gait parameters with them remains a persistent challenge that needs to be addressed. Overcoming challenges related to errors, intelligent error-handling mechanisms, and subject-specific conditions will be critical for advancing the detection of clinically relevant gait changes using inertial sensors.

Similarly, in the field of cognitive decline and MCI screening, current methods have their limitations. Neurocognitive disorders have a significant impact on people's lives, and early detection and intervention are key to slowing down or preventing their progression. MCI, the earliest detectable manifestation of cognitive-related disorders, often goes unnoticed until it starts to interfere

with daily functioning. Thus, differentiating normal from abnormal cognitive decline represents a central question in current research.

Existing screening tools for cognitive impairment rely on paper-and-pencil tests that are time-consuming, require expert administration, and are susceptible to interpretation bias. Movement-based tests, on the other hand, provide a language-free and more universally accessible approach to studying cognitive decline. Mobile technology and wearables hold promise for MCI screening, particularly in assessing dual-task performance. Compared to manual analysis by clinicians, machine learning-based assessments can provide more objectivity and improved detection accuracy by capturing more complex relationships between variables. Nonetheless, the lack of standardized protocols and technology to objectively evaluate gait and cognitive performance hampers progress in this area. It is important to recognize that this area of study is still in its early stages and requires further development.

In summary, the integration of quantitative gait analysis and cognitive assessment through wearable technology holds great promise for improving early screening and intervention in age-related health conditions. Addressing practical considerations, standardizing protocols, and enhancing the reliability of classification methods are crucial steps toward the wider adoption of these technologies. Future opportunities lie in refining measurement techniques, exploring novel methodologies, and expanding our understanding of motor-cognitive relationships. These advancements may ultimately result in improved healthcare outcomes and an enhanced quality of life for individuals experiencing gait-related disorders and cognitive decline.

The next chapters detail our contributions to this field of study.

Part II

MOVEMENT ANALYSIS USING FOOT-WORN INERTIAL SENSORS

Chapter 3

Orientation-Invariant Spatiotemporal Gait Analysis

This chapter is based on the following publication:

V. Guimarães, I. Sousa, and M. V. Correia, “Orientation-Invariant Spatio-Temporal Gait Analysis Using Foot-Worn Inertial Sensors,” *Sensors*, vol. 21, no. 11, p. 3940, Jun. 2021, doi: 10.3390/s21113940.

In this chapter, we introduce a novel gait analysis solution based on foot-worn inertial sensors.

Methods for objective spatiotemporal gait analysis based on inertial sensors usually assume the careful alignment of the sensors on the body or they infer sensor orientation by exploring the cyclic characteristics of walking. In addition to being unrealistic to assume that the sensor can be perfectly aligned with the body, the robustness of gait analysis with respect to differences in sensor orientation has never been investigated in previous studies—potentially hindering use in clinical settings.

To address this gap in the literature, we introduce an orientation-invariant gait analysis approach and propose a method to quantitatively assess robustness to changes in sensor orientation. We test multiple approaches to obtain sensor orientation and position and validate our results in a group of young adults, using an optical motion capture system as a reference.

Overall, good agreement between systems is achieved considering an extensive set of gait metrics. Gait speed is evaluated with a relative error of -3.1 ± 9.2 cm/s, but precision improves when turning strides are excluded from the analysis, resulting in a relative error of -3.4 ± 6.9 cm/s. We demonstrate the invariance of our approach by simulating rotations of the sensor on the foot.

3.1 Introduction

Inertial sensor-based gait analysis has become highly attractive in the past years and constitutes a promising approach to assist clinical decision-making for gait-related disorders in aging [105]. However, inertial sensors are susceptible to noise, and complex algorithms are needed to reliably measure gait parameters [240].

To handle errors, studies typically exploit the cyclic nature of gait and the hypothesis that the velocity of the sensor—when placed on the foot—is zero when the foot is in contact with the ground (i.e., at some point in stance) [146]. Zero velocity intervals (ZVIs) are then used to improve the methods for estimating sensor orientation [131, 142, 241] and reconstructing displacement [118, 131, 132, 142, 242, 243]—required to evaluate spatiotemporal parameters of walking.

Even though the orientation of the sensor relative to the Earth frame can be obtained using appropriate sensor fusion methods [244, 245], the orientation of the sensor relative to the movement direction or relative to the body part, is typically unknown. In some situations, for instance, when a smartphone is used to monitor users' motion, the sensor orientation and position relative to the body may be changing over time, which requires specific methods to identify the users' motion direction with respect to the sensor. Methods based on the ellipsoidal shape of horizontal components of acceleration [246], or based on sinusoidal approximations of the acceleration data [247] have been proposed in the literature, with direct applications on pedestrian dead reckoning. When the sensor is placed on the feet, other techniques can be used. Falbriard et al. [248] propose an automatic calibration process, in which the Principal Component Analysis (PCA) is used to find the foot mediolateral axis based on the angular rates acquired during gait. In [240], the foot orientation vector is estimated using a particle filter. Although [248] and [240] claim that the methods they propose enable the use of sensors in a robust and reproducible manner, reproducibility is not assessed by the authors.

Most studies assume that at least one sensor axis is aligned with the body—typically the mediolateral axis when the sensor is placed on the feet. Based on this assumption, the mediolateral angular rate and/or forward-anterior acceleration can be directly determined and used to detect gait events [118, 142]. Sagittal foot angle can also be obtained [242], from which heel and toe clearance metrics can be calculated [133, 249]. The inertial coordinate system at the beginning of each gait cycle can be used as a base coordinate system (aligned with the mediolateral and forward-anterior axis of the body), relative to which foot orientation and trajectories can be determined [118, 243]. Besides assuming a known configuration, the extent to which algorithms depend on the precise alignment of the sensors has not yet been investigated.

Another approach for gait analysis relies on data-driven procedures. These methods require training in a relevant dataset, using features extracted from inertial sensor data [250] or, in case of using sufficiently deep architectures, regressing directly against raw sensor data [144]. In both cases, generalization is possible only if sufficient and representative data are given to the model, allowing it to learn from the data. Studies that employ data-driven approaches train their models without considering multiple sensor orientations [144, 250], so their final models may not be robust

to differences in orientation. However, assuming precise alignment of the sensors on the body seems unrealistic in practice.

In this chapter, we approach the topic of orientation-invariance explicitly. We introduce an orientation-invariant gait analysis approach relying on inertial sensors placed on the shoes and validate results in a group of young adults, using an optical motion capture system as reference. We test multiple methods from the literature to estimate sensor orientations and positions. We propose a method to quantitatively evaluate invariance to differences in orientation and demonstrate it by simulating rotations of the sensor on the foot. Experiments and results are presented in this chapter, together with a critical interpretation; findings are described considering the research problem being addressed and past work within the topic.

3.2 Materials and Methods

3.2.1 Wearable Sensors

Acceleration and angular rate were measured using two inertial measurement units (IMUs). The IMUs were developed at Fraunhofer Portugal AICOS and incorporated a 32-bit Arm Cortex M4F processor (Nordic nRF52). The device was equipped with a tri-axial gyroscope and a tri-axial accelerometer (Bosch[®] BMI160), and communicated with a computer via Bluetooth[®] Low Energy, enabling data collection at a sampling rate of 100 Hz. Sensors were placed on the shoes, near the foot instep, as shown in Figure 3.1–b.

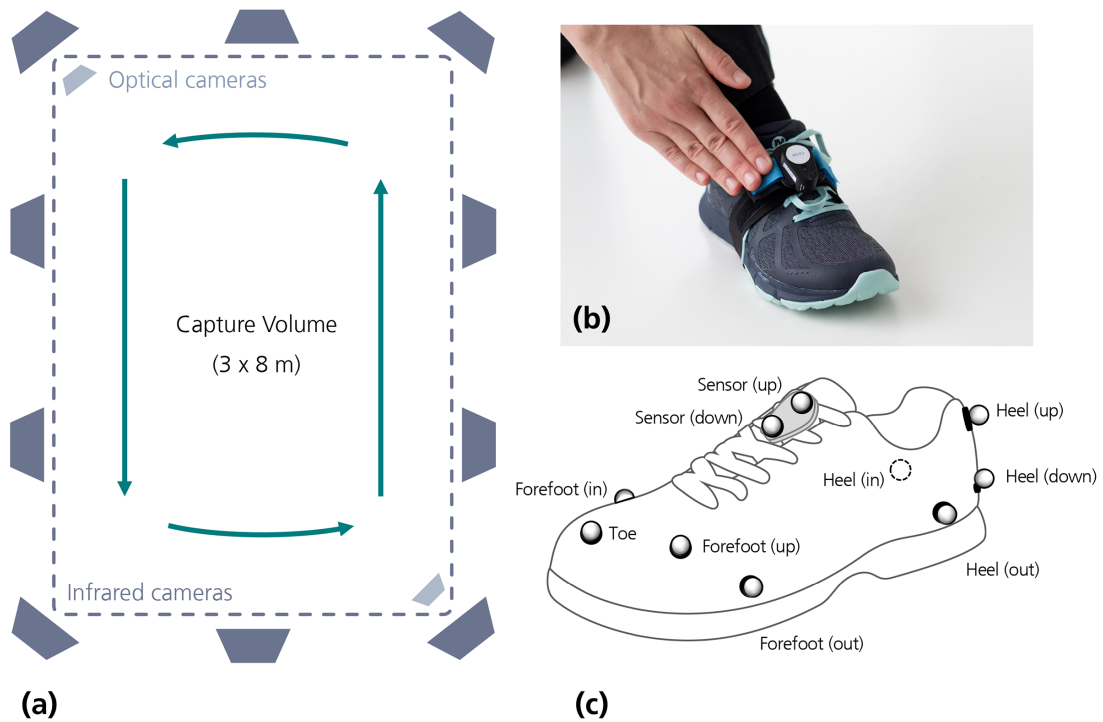


Figure 3.1: Wearable sensors and reference system: (a) Capture volume and camera installation. (b) Sensor placement. (c) Marker placement.

3.2.2 Reference System

As a reference system, we used an optical motion capture system (Vicon, Oxford Metrics). It consisted of 10 infrared cameras (Vicon Vero v2.2), plus 2 optical cameras (Vicon Vue), installed as illustrated in Figure 3.1–a, which resulted in a capture volume of around 3 per 8 m. Optical markers were placed on participants' shoes as illustrated in Figure 3.1–c.

The trajectories of the markers were captured at 100 Hz and post-processed using Vicon Nexus (Vicon Motion Systems Ltd., Version 2.10.1). Post-processing included automatic marker labeling, manual marker swapping correction, and gap filling operations, using the methods of spline, rigid body and pattern fill available in Vicon Nexus.

After post-processing, gait parameters were automatically extracted from trajectories using a Python routine. The pipeline started with the identification of steady periods to segment signals into strides, which involved the application of thresholds to the velocity of the markers. It was followed by horizontal plane correction, where the normal (vertical) vector corresponded to the second principal axis obtained using Principal Component Analysis (PCA); PCA was applied to the position vectors—obtained from the heel (down) marker—formed between two successive steady states. Before determining gait events and calculating gait parameters, trajectories were low pass filtered using a zero-lag bidirectional first-order Butterworth filter (cutoff of 20 Hz).

Initial foot contact (FC) and foot off (FO) were automatically detected, using as reference the trajectories from the markers on the heel (down) and toe. The FC instant was considered a minimum in heel vertical velocity [251], whereas the FO event was considered a maximum in vertical toe acceleration [252], as described in [253]. Temporal gait metrics—*stride*, *swing*, and *stance duration*—were calculated as defined in [242]. *Cadence* was obtained as the inverse of stride duration, converted to the units of steps per minute.

The trajectories of the heel (down), toe, and sensor centroid were used to evaluate spatial parameters, as illustrated in Figure 3.2: (i) *Stride length* (SL) was described as the linear distance obtained between two successive horizontal mid-stance sensor positions (as defined in [254]); (ii) *Turning angle* was defined as the angle (yaw) between two successive horizontal mid-stance foot vectors that were obtained from the positions of the heel and toe at mid-stance; (iii) *Swing width* (SW) was considered the maximum lateral excursion of the feet during swing, corresponding to the maximum-size vector perpendicular to the stride length direction; (iv) *Minimum toe clearance* (MTC) was obtained directly from toe trajectories, considered as the minimum peak vertical displacement during swing, to which the toe height at toe-off was subtracted. *Gait speed* was obtained by dividing stride length by its corresponding stride duration.

All parameters were reported per stride, and calculated individually for each foot side.

3.2.3 Data Collection

A convenient sample of 26 healthy young adults (average age of 29.2 ± 5.3 years, 13 males and 13 females) participated in data collection activities. After calibrating Vicon cameras and preparing the system for data acquisition, we placed the sensors and markers on the shoes, as shown

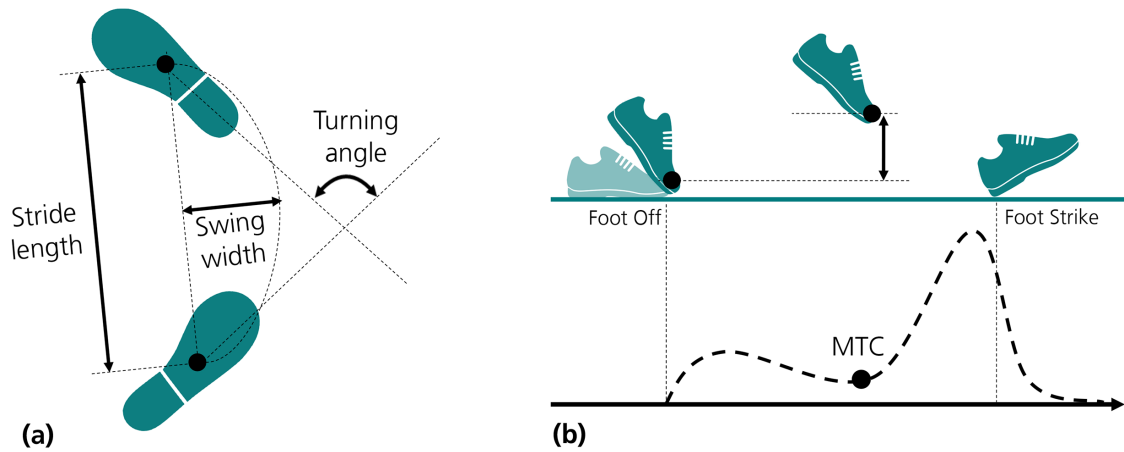


Figure 3.2: Illustration of spatial gait parameters: (a) Stride length, swing width and turning angle. (b) Minimum toe clearance (MTC).

in Figure 3.1. This was followed by a short static subject calibration trial. Afterward, we asked the participants to do three consecutive laps, during which Vicon and IMU data were collected simultaneously. In each lap, participants were requested to walk straight along the length of the capture volume (i.e., along the 8 m) and use the width of the capture volume to turn (as illustrated in Figure 3.1). Participants repeated the walking trials in both directions—clockwise and counterclockwise—walking at comfortable, slower, and faster speeds (self-selected), which resulted in 6 acquisitions per participant. All conditions were then repeated once. The study received approval from the Ethical Committee of the University of Porto (81/CEUP/2019) and all participants provided written informed consent.

3.2.4 IMU Data Processing

The following sections describe the methods for IMU data processing, inspired by typical gait analysis routines—including the stages of zero velocity detection, orientation estimation, double integration, events detection, and gait parameters estimation. The processing pipeline included the evaluation of different approaches for orientation estimation and double integration. Every stage was formulated using an orientation-invariant approach.

3.2.4.1 Zero Velocity Detection

Zero velocity intervals (ZVIs) were detected using the angular rate energy detector [255]. According to [255], angular rate provides rich information concerning the detection of ZVIs. Compared to other methods (e.g., the acceleration magnitude detector), the angular rate energy detector achieved the highest performance [255].

To calculate the energy of the angular rate magnitude, we used a sliding window with 0.15 s. The size of the window was experimentally set to ensure an appropriate energy result, i.e., not too smoothed, nor too noisy. The window size of 0.15 s resulted in a good compromise that could

adapt to all walking speeds considered. To determine ZVIs, a threshold was applied to the energy of the angular rate magnitude. The threshold was calculated based on the average of the energy, to ensure that differences in walking speed (reflected as different signal amplitudes) would be considered. The threshold was experimentally set at 1/8 of the average, to ensure that all strides in all velocities would be captured by the method.

After roughly detecting ZVIs, we proceeded with a dynamic adjustment of the intervals. For each ZVI, we did a threshold refinement search: starting with a very low threshold, we have progressively increased it (in steps of 1/20 of the ZVI energy range), until a minimum interval size was obtained. The minimum interval size was set at 0.1 s, unless an interval lower than 0.1 s was registered in the trial. Using this process, we ensured a confident detection of all strides and, additionally, that all ZVIs were refined to potentially include only zero velocity instants. The remaining intervals, i.e., the moving intervals, were used to evaluate movement in the subsequent processing stages.

3.2.4.2 Orientation Estimation

Orientation, or attitude, of the sensor relative to the global frame of reference was expressed using quaternions. A quaternion is a four-dimensional complex number that describes the orientation of a frame B relative to A , defined as $\mathbf{q} = [\cos(\theta/2) \quad -r_x \sin(\theta/2) \quad -r_y \sin(\theta/2) \quad -r_z \sin(\theta/2)]$, where θ represents the angle of rotation around axis $\mathbf{r}$, defined in frame A . In this case, frame A is the sensor frame, while frame B is the global frame of reference. The global frame of reference was defined by two perpendicular horizontal axes (x and y)—arbitrarily set—and a vertical axis, z , pointing to the sky. To obtain quaternions, we tested three methods: gyroscope integration and two complementary filters (CFs)—Madgwick and Euston.

Gyroscope Integration

The gyroscope integration method takes advantage of the knowledge of the moving and not-moving intervals. When the sensor is not moving (i.e., during a ZVI)—when all measured accelerations are due to the Earth's gravity acceleration—the accelerometer signal is used to estimate sensor inclination. The vector $\mathbf{g}_s = [\bar{a}_x, \bar{a}_y, \bar{a}_z]$, defined with the average acceleration values measured while the sensor is not moving, corresponds to the z -axis in the global frame of reference, i.e., the gravity axis. The horizontal axis— y -axis, perpendicular to gravity—is then defined arbitrarily. The initial quaternion is obtained from gravity and the horizontal vector, using the tri-axial attitude determination (TRIAD) algorithm [256].

The initial quaternion is updated each time the foot is in contact with the ground. During the intervals when the foot is moving, the quaternion $\mathbf{q}(t)$ is updated resorting to the integration of angular rate $\boldsymbol{\omega}(t)$ measured by the gyroscope, as described in [142, 241]. This is performed as defined by Equations 3.1 and 3.2:

$$\dot{\mathbf{q}}(t) = \frac{1}{2} \mathbf{q}(t - \Delta t) \otimes \mathbf{p}(\boldsymbol{\omega}(t)) \quad (3.1)$$

$$\mathbf{q}(t) = \frac{\mathbf{q}(t - \Delta t) + \dot{\mathbf{q}}(t)\Delta t}{\|\mathbf{q}(t - \Delta t) + \dot{\mathbf{q}}(t)\Delta t\|} \quad (3.2)$$

where $\dot{\mathbf{q}}(t)$ is the quaternion derivative and Δt is the sampling interval. The function $\mathbf{p}(\cdot)$ denotes the quaternion representation of a vector and the $\otimes$ operator represents the quaternion product.

Madgwick CF

In Madgwick [257], the quaternion derivative $\dot{\mathbf{q}}(t)$ used in Equation 3.2 is replaced by a corrected estimate $\hat{\dot{\mathbf{q}}}(t)$ that incorporates orientation information provided by the accelerometer. The method is based on the calculation of the gradient descent, as shown in Equation 3.3.

$$\hat{\dot{\mathbf{q}}}(t) = \dot{\mathbf{q}}(t) - \beta \frac{\Delta \boldsymbol{\epsilon}(t)}{\|\Delta \boldsymbol{\epsilon}(t)\|} \quad \text{with} \quad \Delta \boldsymbol{\epsilon}(t) = J(\boldsymbol{\epsilon}(t))^T \boldsymbol{\epsilon}(t) \quad (3.3)$$

where J denotes the Jacobian, β is the filter gain and $\boldsymbol{\epsilon}(t)$ denotes the error term, obtained by subtracting the accelerometer measurement in the sensor frame $\mathbf{a}_s(t)$ to the theoretical gravity vector in sensor coordinates $\mathbf{g}_s(t)$ (obtained by transforming gravity in global coordinates to sensor coordinates). β can be defined as $\beta = \sqrt{3/4} \bar{\omega}_{max} \pi / 180$, where $\bar{\omega}_{max}$, expressed in degrees, represents the maximum gyroscope measurement error (mean zero gyroscope measurement error).

The Madgwick filter is applied at all instants of the signal (moving and not moving intervals) considering as basis the initial quaternion determined using TRIAD, as described previously.

Euston CF

The explicit complementary filter, also known as Euston filter, was implemented as described in [241, 245]. In Euston, instead of replacing the value of the quaternion, the measured angular rate $\boldsymbol{\omega}(t)$ is replaced by a corrected angular rate signal, resulting in the following filter dynamics:

$$\hat{\dot{\mathbf{q}}}(t) = \frac{1}{2} \hat{\mathbf{q}}(t - \Delta t) \otimes \mathbf{p}(\boldsymbol{\omega}(t) + \boldsymbol{\delta}(t)) \quad (3.4)$$

in which the error term $\boldsymbol{\delta}$ is obtained following Equations 3.5 and 3.6, where the term $\mathbf{e}(t)$ describes the angular mismatch between theoretical ($\mathbf{g}_s(t)$) and measured ($\mathbf{a}_s(t)$) direction of gravity [245].

$$\mathbf{e}(t) = \frac{\mathbf{g}_s(t)}{\|\mathbf{g}_s(t)\|} \times \frac{\mathbf{a}_s(t)}{\|\mathbf{a}_s(t)\|} \quad (3.5)$$

$$\boldsymbol{\delta}(t) = k_P \mathbf{e}(t) + k_I \int \mathbf{e}(t) dt \quad (3.6)$$

The Euston filter has two adjustable parameters, the proportional gain k_P —to separate low- and high-frequency estimates of orientation—and the integrator gain k_I —to compensate for gyroscope bias [241, 245]. Similarly to the Madgwick filter, we apply Euston to all instants of the signal, considering as basis the initial quaternion determined using TRIAD [256].

3.2.4.3 Double Integration

After obtaining orientation quaternions, $\mathbf{q}(t)$, we calculate linear acceleration in global coordinates, hereinafter represented as $\mathbf{a}(t)$ for simplicity. To that purpose, we first estimate the value of the gravity vector in global coordinates, $\mathbf{g}_w = [0, 0, \bar{a}_{zv}]$, where $\bar{a}_{zv}$ is the average acceleration magnitude measured during ZVIs. Linear acceleration is then obtained as shown in Equation 3.7.

$$\mathbf{a}(t) = \mathbf{q}(t) \otimes \mathbf{p}(\mathbf{a}_s(t)) \otimes \mathbf{q}^{-1}(t) - \mathbf{g}_w \quad (3.7)$$

where $\mathbf{a}_s(t)$ represents raw acceleration, as measured by the sensor.

To obtain displacements, we integrate linear acceleration two times. On the first integration, an estimate of velocity, $\hat{\mathbf{v}}(t)$, is obtained. Integrals are computed using the Trapezoidal Rule, as shown in Equation 3.8.

$$\hat{\mathbf{v}}(t) \approx \sum_i \frac{\mathbf{a}_i + \mathbf{a}_{i-1}}{2} \Delta t \quad (3.8)$$

To bound the errors, two different methods—linear dedrift and direct and reverse integration—are employed, as we explain next. After obtaining trajectories, a novel approach to correct the final vertical position (assuming walking on a flat surface) is tested. The method rotates the trajectories so that the final height of each stride is zero. To that purpose, a rotation quaternion is calculated, using as a basis the angle with the horizontal plane at the end of the stride and the rotation vector calculated as the cross product between the vertical axis and the stride displacement vector. The rotation quaternion is used to rotate trajectories within each moving interval.

Linear Dedrift

Double integration is performed between ZVIs, on a stride-by-stride basis. To fulfill the zero-velocity assumption—on which moving intervals are bounded by zero velocity instants—a linear drift function ($\mathbf{d}_v(t)$) is estimated and subtracted from the estimated velocity, as described in [242] and shown in Equation 3.9.

$$\mathbf{v}(t) = \hat{\mathbf{v}}(t) - \mathbf{d}_v(t) \quad (3.9)$$

Trajectory $\mathbf{s}(t)$ is obtained by integrating again velocity $\mathbf{v}(t)$.

Direct and Reverse Integration

The direct and reverse integration method fuses the regular integral with a time-reversed integral so that the boundary conditions, in this case, the zero-velocity conditions, are satisfied in the initial and final values of the integral [118, 243]. The result of direct ($\mathbf{v}_{\rightarrow}(t)$) and reverse ($\mathbf{v}_{\leftarrow}(t)$) integration is combined using a sigmoid weighting function ($w(t)$), as shown in Equation 3.10.

$$\mathbf{v}(t) = (1 - w(t))\mathbf{v}_{\rightarrow}(t) + w(t)\mathbf{v}_{\leftarrow}(t) \quad (3.10)$$

The sigmoid $w(t)$, specified in Equation 3.11, is shaped using the steepness parameter, η , and the inflection point, t_i , defined between the temporal bounds t_n and t_{n+1} of each moving interval. To define t_i , a proportion α_i , between 0 and 1, of the moving interval is used.

$$w(t) = \frac{h(t) - h(t_n)}{h(t_{n+1}) - h(t_n)} \quad \text{with} \quad h(t) = \left(1 + \exp\left(-\frac{t - t_i}{\eta}\right)\right)^{-1} \quad (3.11)$$

Position is estimated by integrating velocity.

3.2.4.4 Events Detection

To detect gait events avoiding the need of determining angular rate or acceleration in body coordinates (where the alignment of the sensor on the body would need to be known [118,142,243]), we used acceleration magnitude and the vertical component of acceleration (in global coordinates).

We observed that FO events can generally be found in an acceleration magnitude perturbation before swing. To approximate this instant, we filtered acceleration magnitude using a zero-lag bidirectional 2nd order Butterworth low-pass filter with a cutoff of 7 Hz. The cutoff frequency was experimentally chosen to ensure that (i) the perturbations were smoothed resulting in a single peak value, and (ii) the resultant peak approximates the acceleration magnitude perturbation where annotated FO events are observed. The FO event was considered the first maximum peak above the average of the moving interval appearing between two ZVIs (i.e., the first value above the average surrounded by two values with lower magnitude), as illustrated in Figure 3.3.

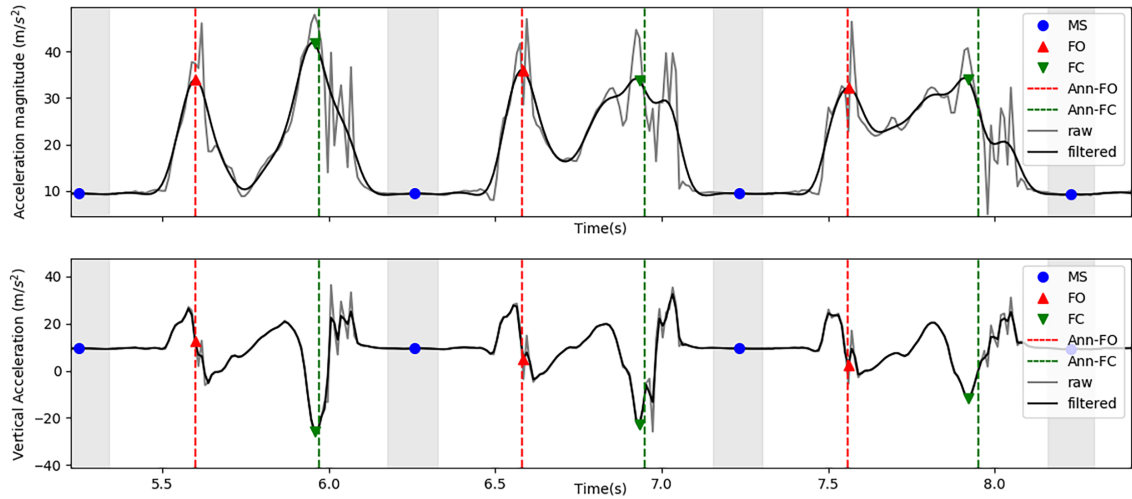


Figure 3.3: Events detection from low pass filtered acceleration magnitude and low pass filtered vertical acceleration (*FO*—Foot off; *FC*—Initial foot contact; *MS*—Mid-stance; *Ann-FO*—Annotated FO; *Ann-FC*—Annotated FC).

FCs are detected between FOs and the beginnings of the next ZVIs. FCs were considered the absolute minimum of vertical acceleration measured between these two instants (Figure 3.3). Before detecting FC events, vertical acceleration was low-pass filtered using a zero-lag bidirectional 1st order Butterworth filter with a cutoff of 30 Hz. The cutoff frequency was experimentally

chosen to ensure the attenuation of high-frequency noise that could hinder the detection of FC events.

3.2.4.5 Gait Parameters Estimation

After calculating orientation and position and determining FO and FC events, temporal and spatial parameters were estimated for each gait cycle n . Temporal parameters—stride, swing, and stance duration—were determined as defined in [242]. Cadence was obtained as the inverse of stride duration, expressed in steps per minute. Spatial parameters (illustrated in Figure 3.2) were calculated using the information of moving intervals, defined by the temporal bounds of t_n and t_{n+1} . To estimate SL and SW, we used trajectories on the horizontal plane, $\mathbf{s}_{xy}$, as determined by Equations 3.12 and 3.13, where $\vec{\mathbf{s}}_n(t)$ represents a displacement vector relative to the final stride position at t_{n+1} , obtained as $\mathbf{s}_{xy}(t_{n+1}) - \mathbf{s}_{xy}(t)$. In Equation 3.13, the symbol $\angle$ denotes the angle between two vectors.

$$SL_n = \|\vec{\mathbf{s}}_n(t_n)\| \quad (3.12)$$

$$SW_n = \max_{t \in \{FO_n:FC_{n+1}\}} \|\vec{\mathbf{s}}_n(t)\| \sin(\angle(\vec{\mathbf{s}}_n(t_n), \vec{\mathbf{s}}_n(t))) \quad (3.13)$$

Gait speed was obtained by dividing SL by its corresponding stride duration.

To calculate MTC, we used a method inspired by the work of Kanzler, C. [249]. To estimate toe trajectory, we have first estimated the distance between the sensor and the toe, r , using as a basis the angle produced by the foot at FO, $\alpha(FO_n)$, in each gait cycle n , as illustrated in Figure 3.4.

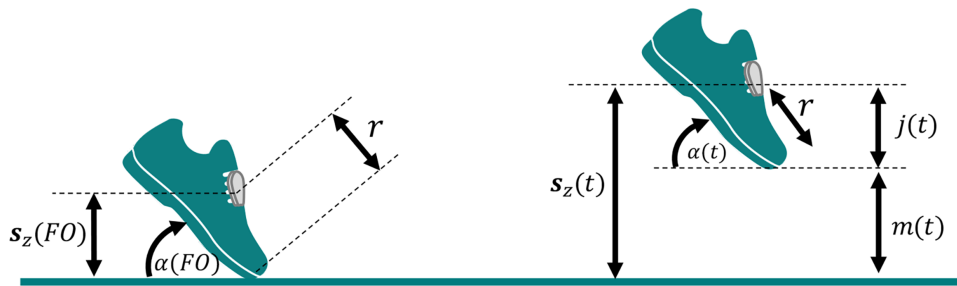


Figure 3.4: Variables involved in the calculation of toe trajectory.

To obtain the angle $\alpha(t)$, we did a series of vector transformations. First, we have converted the vertical vector $[0, 0, 1]$ to sensor coordinates, using the quaternion at the beginning of the moving interval, i.e., at the foot flat at t_n . Then, we transformed the resultant vector back to global coordinates using the quaternion at FO. This vector, $\vec{\mathbf{v}}_w$, was used to estimate the medio-lateral vector, $\vec{\mathbf{l}}_w$, using the cross product between $\vec{\mathbf{v}}_w$ and $[0, 0, 1]$. The forward vector, $\vec{\mathbf{f}}_w$, was calculated

using the cross product between $[0, 0, 1]$ and $\vec{l}_w$, which was then converted back to sensor coordinates using the quaternion at t_n . This vector, $\vec{f}_s$, parallel to the ground at foot flat and pointing forward towards the toes, was used to estimate the angle $\alpha(t)$, as depicted in Equation 3.14.

$$\alpha(t) = \angle([0, 0, 1], \vec{v}) - \alpha(t_n) \quad \text{with} \quad \vec{v} = \mathbf{q}(t) \otimes \mathbf{p}(\vec{f}_s) \otimes \mathbf{q}(t)^{-1} \quad \text{and} \quad t \in [t_n : t_{n+1}] \quad (3.14)$$

The distance between the sensor and the toe (r) was obtained by the average of the values determined in each stride n , as shown in Equation 3.15.

$$r = \frac{1}{N} \sum_{n=1}^N \frac{s_z(FO_n)}{\sin(\alpha(FO_n))} \quad (3.15)$$

where N is the total number of strides and $s_z(t)$ represents the z component of the trajectory of the sensor (i.e., its vertical displacement). MTC was considered the minimum peak vertical toe displacement measured during the swing phase of walking. This vertical displacement, $m(t)$, was estimated as shown in Equation 3.16, where $r \times \sin(\alpha(t))$ represents the vertical distance between the sensor and the toe (shown as $j(t)$ in Figure 3.4).

$$MTC_n = \min_{t \in \{FO_n: FC_{n+1}\}} m(t) \quad \text{with} \quad m(t) = \begin{cases} s_z(t) - r \times \sin(\alpha(t)) & \text{if } \alpha(t) > 0 \\ s_z(t) + r \times \sin(\alpha(t)) & \text{otherwise} \end{cases} \quad (3.16)$$

To calculate turning angles, we converted an arbitrary horizontal vector (e.g., the vector $[0, 1, 0]$) to sensor coordinates, using as basis the quaternions estimated at t_n and at t_{n+1} . The resultant vectors represented the orientation of the sensor in the horizontal plane, so that the angle between these two vectors corresponded to the turning angle.

3.2.5 Experiments

One hundred and sixty of the collected samples were post-processed using Vicon Nexus and used as a reference for IMU-based gait analysis evaluation. To avoid overfitting, we split data into development and validation sets. Approximately 30% of the samples, from uniquely randomly selected users, were included in the development set and used for algorithm design, debugging, and optimization. The remaining 70% were used for validation and orientation-invariance proof. The resulting dataset split is shown in Table 3.1.

Table 3.1: Dataset characteristics after splitting.

Metric	Development Set (7 Subjects/45 Samples)	Validation Set (19 Subjects/115 Samples)
Height (cm)	168.9 ± 7.7	172.5 ± 9.2
Gender (female/male)	4/3	9/10
Weight (kg)	67.3 ± 15.4	71.1 ± 14.7
Foot size (cm)	27.4 ± 1.6	27.8 ± 2.0

To compare gait parameters extracted from the IMU with those extracted from Vicon, systems were first synchronized. To this purpose, we used the cross-correlation between acceleration magnitude—obtained from the IMU—and the centroid of the sensor markers—obtained after deriving the centroid trajectory two times; the maximum cross-correlation was used to compensate for the time-shift between data sources, as in [131].

A tolerance of 0.1s was employed to classify FC events (and their corresponding stride gait parameters) as true positive cases. Only the strides whose FC time was consistent with those obtained with Vicon were considered for comparison. Reference strides without any corresponding IMU-derived candidate were classified as not detected.

3.2.5.1 Parameter Tuning and Algorithm Selection

The development set was used to tune parameters and select the most reliable methods for orientation estimation and double integration. As shown in Section 3.2.4.2, the possible orientation estimation methods were the Madgwick CF, the Euston CF, and the gyroscope integration, which could be used in combination with two possible double integration methods: the direct and reverse integration, and the linear dedrifting—both presented in Section 3.2.4.3. Additionally, the horizontal correction method presented in Section 3.2.4.3 could or not be employed. Each method—with the exception of gyroscope integration and linear dedrifting—included a set of parameters (also detailed in Sections 3.2.4.2 and 3.2.4.3) that needed to be optimized in view to improve the performance of the gait analysis method. Using a grid-search approach, all possible combinations of methods (i.e., orientation estimation and double integration methods) and a set of candidate parameters could be tested, where the number of resultant combinations depended on the number of parameters tested. For this reason, parameter tuning was first performed using a coarse grid of parameters—i.e., a small number of candidate parameters covering a wider range of values. The most promising combinations of parameters and methods were used to define a finer grid that considered candidate parameters defined in the neighborhood of the best parameter configurations. Parameters tested within coarse (resulting in 140 combinations) and fine grid-search (resulting in 105 combinations) are shown in Table 3.2.

To select the best combination of parameters and methods, we calculated the root mean square error (RMSE) between IMU-derived and reference gait parameters. The RMSE of each parameter was normalized by the average of the reference, and then averaged to obtain a single score per configuration. The minimum normalized RMSE defined the most appropriate set of parameters and methods.

3.2.5.2 Instrument Comparison and Validation

We compared gait parameters extracted from the IMU with those extracted from Vicon using the validation set. For each cycle, we estimated the difference between IMU-derived and reference gait parameters. Accuracy (mean of relative and absolute error) and precision (standard deviation

Table 3.2: Methods and parameters used in coarse and fine grid-search.

Type	Methods	Coarse Grid-Search	Fine Grid-Search
Orientation estimation	Madgwick CF	$\bar{\omega}_{max} = \{2, 4, 6\}$	$\bar{\omega}_{max} = \{1.0, 1.5, 2.0, 2.5, 3.0\}$
	Euston CF	$k_P = \{0.2, 0.4, 0.6\}$ $k_I = \{0, 0.05\}$	$k_P = \{0.15, 0.2, 0.25\}$ $k_I = \{0, 0.01\}$
	Gyroscope integration	Not applicable	Not applicable
Double integration	Direct and reverse integration	$\eta = \{0.1, 0.2\}$ $\alpha_i = \{0.4, 0.5, 0.6\}$	$\eta = \{0.15, 0.2, 0.25\}$ $\alpha_i = 0.55, 0.6, 0.65$
	Linear dedrift	Not applicable	Not applicable
	Horizontal correction	{True, False}	{True}

of relative and absolute error) were reported for each parameter. Agreement between the two instruments was assessed using 95% limits of agreement, as introduced by Bland Altman [258]. Data were assessed for normal distribution using Shapiro–Wilk tests, to decide for the use of parametric or non-parametric tests. Correlation between instruments was calculated using the correlation coefficients of Pearson (r_p)—in case of normal distribution—or Spearman (r_s)—when data could not be assumed to be normally distributed. We have also reported RMSE and equivalence tests using an equivalence zone of $\pm 5\%$ of the average of the metric. Equivalence tests were based on Paired T -test (T)—for parametric—or Wilcoxon signed-rank test (W)—for non-parametric.

To validate results in a scenario where only straight walking is considered for gait assessment (as required to assess several gait disorders [65]), we repeated validation tests without including turns. For this purpose, a turning stride was considered a stride where the turning angle (as measured by the reference system) was above 20 degrees (as in [132]).

A significance level (p -value) of 5% was used to evaluate the results.

3.2.5.3 Orientation Invariance

To test for orientation invariance, we simulated multiple rotations of the IMU on the shoes. For that purpose, we sampled uniform random rotations (quaternions), as suggested by Shoemaker, K. [259], and used those quaternions to synthetically rotate raw inertial sensor data. Random orientation quaternions, $\mathbf{q}_{rnd}$, were generated using:

$$\mathbf{q}_{rnd} = \begin{bmatrix} \sqrt{1-u_1} \sin(2\pi u_2) & \sqrt{1-u_1} \cos(2\pi u_2) & \sqrt{u_1} \sin(2\pi u_3) & \sqrt{u_1} \cos(2\pi u_3) \end{bmatrix} \quad (3.17)$$

where $u_1, u_2, u_3 \in [0, 1]$ are randomly generated numbers [259].

To evaluate the performance of the system when IMUs were placed at random rotations, we compared gait parameters extracted from the original sensor orientation with those extracted from

a rotated version of the sensor. To quantify differences, we calculated the Root Mean Square Deviation (RMSD), correlation (using Pearson-parametric-or Spearman-non-parametric), and equivalence tests using a stricter equivalence zone of $\pm 1\%$ of the average of the metric. Equivalence tests were based on Paired T -test (T)—for parametric—or Wilcoxon signed-rank test (W)—for non-parametric. To choose an appropriate test, samples were first tested for normal distribution using Shapiro–Wilk; non-parametric tests were chosen in case of non-normal distribution. A significance level (p -value) of 5% was used to evaluate the results.

3.3 Results

3.3.1 Algorithm Selection and Parameter Tuning

The most promising combinations of parameters and methods were defined by the results of the coarse grid search. We selected the best-performing combinations considering the minimum normalized RMSE achieved by each candidate method. The best combinations in coarse grid-search were: (i) the Madgwick CF ($\tilde{\omega}_{max} = 2.0$) combined with direct and reverse integration ($\eta = 0.2$, $\alpha_i = 0.6$) or with linear dedrifting methods; (ii) the Euston CF ($k_p = 0.2$, $k_I = 0$) combined with direct and reverse integration ($\eta = 0.2$, $\alpha_i = 0.6$); and (iii) the gyroscope integration combined with linear dedrifting, all with active horizontal correction. The second optimization was performed using a finer grid of parameters (shown in Table 3.2), defined in the vicinity of the best-performing combinations. The resulting best combination of methods and parameters was the Euston CF ($k_p^{opt} = 0.15$, $k_I^{opt} = 0$) with direct and reverse integration ($\eta^{opt} = 0.25$, $\alpha_i^{opt} = 0.55$) and active horizontal correction. The RMSE values obtained in the development set using this parameter configuration are shown in Table 3.3.

Table 3.3: Performance on the development set, using the best combination of methods and parameters ($n = 2788$ strides).

Metric	IMU	VICON	Rel. Error	Abs. Error	RMSE
Stride dur. (s)	1.23 (0.27)	1.23 (0.27)	0.00 (0.03)	0.02 (0.03)	0.03
Swing dur. (s)	0.41 (0.08)	0.40 (0.07)	0.01 (0.04)	0.03 (0.03)	0.04
Stance dur. (s)	0.82 (0.21)	0.83 (0.21)	−0.01 (0.03)	0.02 (0.02)	0.03
Cad. (st/min)	101.8 (21.0)	101.7 (20.9)	0.1 (2.6)	1.7 (1.9)	2.6
SL (cm)	112.4 (25.2)	114.9 (26.0)	−2.5 (7.3)	5.3 (5.6)	7.7
Speed (cm/s)	97.8 (35.9)	100.2 (37.9)	−2.3 (6.9)	4.8 (5.4)	7.3
SW (cm)	8.9 (8.1)	9.2 (8.5)	−0.3 (1.3)	0.9 (1.0)	1.3
MTC (cm)	1.7 (0.8)	1.7 (0.6)	0.1 (0.8)	0.6 (0.5)	0.8
Turn angle (°)	36.0 (40.3)	35.3 (40.3)	0.7 (1.4)	1.0 (1.2)	1.6

Shown are mean values (standard deviation) and RMSE. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

3.3.2 Validation

The best combination of methods and parameters was used to generate results for validation. In the analysis, 7015 strides out of 7142 (i.e., approximately 98.2%) were included, meaning that only

127 strides (i.e., 1.8%) were classified as not detected. FC events were detected with an average relative error of -0.01 ± 0.02 s and limits of agreement of -0.06 and 0.04 s. FO events were detected with an average relative error of -0.01 ± 0.05 s and limits of agreement of -0.11 and 0.08 s. FC and FO events were detected with the same average relative error, but the dispersion of the errors in FO detection was higher, as evidenced by the standard deviation and limits of agreement.

The comparison of IMU-derived and reference gait parameters, obtained in the validation set, is shown in Table 3.4.

Table 3.4: Performance on the validation set, including turning strides ($n = 7015$ strides).

Metric	IMU	VICON	Rel. Error	Abs. Error	Lim. Agr.	RMSE	Corr. [†]	Equival. [‡]
Stride dur. (s)	1.22 (0.23)	1.23 (0.24)	0.00 (0.05)	0.02 (0.05)	$[-0.11, 0.11]$	0.05	0.99	± 0.06 (0.0)
Swing dur. (s)	0.41 (0.08)	0.41 (0.06)	0.01 (0.05)	0.03 (0.04)	$[-0.08, 0.09]$	0.05	0.87	± 0.02 (0.0)
Stance dur. (s)	0.81 (0.18)	0.82 (0.19)	-0.01 (0.05)	0.03 (0.05)	$[-0.11, 0.10]$	0.05	0.98	± 0.04 (0.0)
Cad. (st/min)	101.2 (17.5)	101.1 (17.4)	0.1 (3.1)	1.8 (2.5)	$[-6.0, 6.1]$	3.1	0.99	± 5.05 (0.0)
SL (cm)	120.8 (25.1)	124.2 (26.5)	-3.5 (9.7)	6.4 (8.0)	$[-22.5, 15.6]$	10.3	0.94	± 6.21 (0.0)
Speed (cm/s)	103.6 (33.9)	106.7 (36.3)	-3.1 (9.2)	5.8 (7.8)	$[-21.1, 15.0]$	9.7	0.98	± 5.33 (0.0)
SW (cm)	9.9 (9.2)	10.2 (10.1)	-0.4 (4.4)	1.5 (4.2)	$[-9.0, 8.3]$	4.4	0.93	± 0.51 (0.0)
MTC (cm)	1.7 (1.0)	1.9 (0.7)	-0.2 (0.8)	0.6 (0.6)	$[-1.9, 1.4]$	0.9	0.55	± 0.10 (1.0)
Turn angle (°)	37.3 (41.0)	36.4 (41.0)	0.9 (8.6)	2.3 (8.4)	$[-16.0, 17.8]$	8.7	0.97	± 1.82 (0.0)

Shown are mean values (standard deviation), limits of agreement, RMSE, correlation and equivalence interval (p -value). [†] All correlations were based on Spearman and have $p < 0.01$. [‡] Equivalence tests were based on Wilcoxon signed-rank test. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

According to the classification proposed in [260], high (i.e., between 70 and 90) to very high (i.e., above 90) correlation was obtained in all variables, except MTC that had a moderate correlation ($r_s = 0.55$). Equivalence tests revealed all metrics to be practically equivalent (with $p < 0.01$), except MTC, considering equivalence intervals corresponding to 5% of the average of the metric. Average absolute errors in stride duration of 0.02 ± 0.05 s were obtained, which represents an average error of two samples considering the sampling rate of 100 Hz. SL had an average relative error of -3.5 ± 9.7 cm (Table 3.4).

Results excluding turns are shown in Table 3.5, where about 2% of the reference strides (i.e., 78 strides) are considered as not detected, resulting in the analysis of 3785 strides. Correlations between variables are high (i.e., between 70 and 90) or very high (i.e., above 90), except for MTC and turning angles where Spearman correlations are 0.55 and 0.68, respectively, and classified as moderate. Accordingly, equivalence tests reveal all metrics to be practically equivalent (with $p < 0.01$), except for MTC and turning angles. Average relative errors of 0.00 ± 0.04 s and -3.9 ± 6.2 cm were obtained for stride duration and SL, respectively (Table 3.5).

3.3.3 Orientation Invariance

The comparison of gait parameters extracted using the original sensor orientation and using simulated rotations is shown in Table 3.6. As can be observed, all parameters, except the turning angle, present the same values when the sensor is rotated, i.e., all parameters have an RMSD of

Table 3.5: Performance on the validation set, excluding turning strides ($n = 3785$ strides).

Metric	IMU	VICON	Rel. Error	Abs. Error	Lim. Agr.	RMSE	Corr. [†]	Equival. [‡]
Stride dur. (s)	1.22 (0.24)	1.22 (0.24)	0.00 (0.04)	0.02 (0.04)	$[-0.08, 0.08]$	0.04	0.99	± 0.06 (0.0)
Swing dur. (s)	0.41 (0.07)	0.40 (0.06)	0.00 (0.04)	0.02 (0.04)	$[-0.08, 0.09]$	0.04	0.89	± 0.02 (0.0)
Stance dur. (s)	0.81 (0.19)	0.82 (0.19)	0.00 (0.04)	0.02 (0.03)	$[-0.09, 0.08]$	0.04	0.98	± 0.04 (0.0)
Cad. (st/min)	101.8 (17.6)	101.5 (17.4)	0.3 (2.5)	1.5 (2.1)	$[-4.7, 5.2]$	2.5	0.99	± 5.08 (0.0)
SL (cm)	129.8 (18.3)	133.7 (20.7)	-3.9 (6.2)	4.6 (5.8)	$[-16.1, 8.2]$	7.4	0.98	± 6.69 (0.0)
Speed (cm/s)	112.0 (31.9)	115.3 (35.1)	-3.4 (6.9)	4.4 (6.3)	$[-17.0, 10.2]$	7.7	0.99	± 5.77 (0.0)
SW (cm)	4.9 (3.2)	4.7 (2.2)	0.1 (2.8)	0.9 (2.7)	$[-5.4, 5.6]$	2.8	0.84	± 0.24 (0.0)
MTC (cm)	1.4 (0.7)	1.8 (0.6)	-0.4 (0.7)	0.6 (0.5)	$[-1.7, 0.9]$	0.8	0.55	± 0.09 (1.0)
Turn angle ($^{\circ}$)	6.4 (7.0)	5.1 (4.8)	1.2 (5.1)	1.5 (5.1)	$[-8.8, 11.3]$	5.3	0.68	± 0.26 (1.0)

Shown are mean values (standard deviation), limits of agreement, RMSE, correlation and equivalence interval (p -value). [†] All correlations were based on Spearman and have $p < 0.01$. [‡] Equivalence tests were based on Wilcoxon signed-rank test. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

0.0, except the turning angle that has an RMSD of 1.5° . The correlation between turning angles extracted using the original orientation and random rotations of the sensor is very high ($r_s = 0.99$). The performance of the system when data are synthetically rotated remains practically equivalent (equivalence tests have $p < 0.01$ for all gait metrics, considering an equivalence interval of 1% of the average of the metric).

Table 3.6: Orientation invariance results ($n = 7138$ strides).

Metric	Original Orientation	Random Rotations	RMSD	Correlation [†]	Equivalence [‡]
Stride dur. (s)	1.23 (0.26)	1.23 (0.26)	0.0	$r_p = 1.0$	± 0.01 (0.0)
Swing dur. (s)	0.41 (0.08)	0.41 (0.08)	0.0	$r_p = 1.0$	± 0.00 (0.0)
Stance dur. (s)	0.82 (0.20)	0.82 (0.20)	0.0	$r_p = 1.0$	± 0.01 (0.0)
Cad. (st/min)	100.7 (17.9)	100.7 (17.9)	0.0	$r_p = 1.0$	± 1.01 (0.0)
SL (cm)	120.4 (25.5)	120.4 (25.5)	0.0	$r_p = 1.0$	± 1.20 (0.0)
Speed (cm/s)	102.9 (34.3)	102.9 (34.3)	0.0	$r_p = 1.0$	± 1.03 (0.0)
SW (cm)	9.8 (9.2)	9.8 (9.2)	0.0	$r_p = 1.0$	± 0.10 (0.0)
MTC (cm)	1.7 (1.0)	1.7 (1.0)	0.0	$r_p = 1.0$	± 0.02 (0.0)
Turn angle ($^{\circ}$)	37.4 (41.0)	37.4 (41.0)	1.5	$r_s = 0.99$	± 0.37 (0.0)

Shown are mean values (standard deviation), RMSD, Correlation—with Spearman or Pearson—and equivalence interval (p -value). [†] All correlations have $p < 0.01$. [‡] All equivalence tests are based on Paired T -test, except for the turning angle, which is based on Wilcoxon signed-rank test. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

Figure 3.5 allows further (visual) inspection of the generated rotations and their relationship with the differences in measured turning angles (the only metric that presented some differences when the sensor was synthetically rotated). As can be observed, rotations are uniformly distributed across the whole 3D space. The differences in turning angles are randomly distributed through the space, showing no particular tendency towards a specific region.

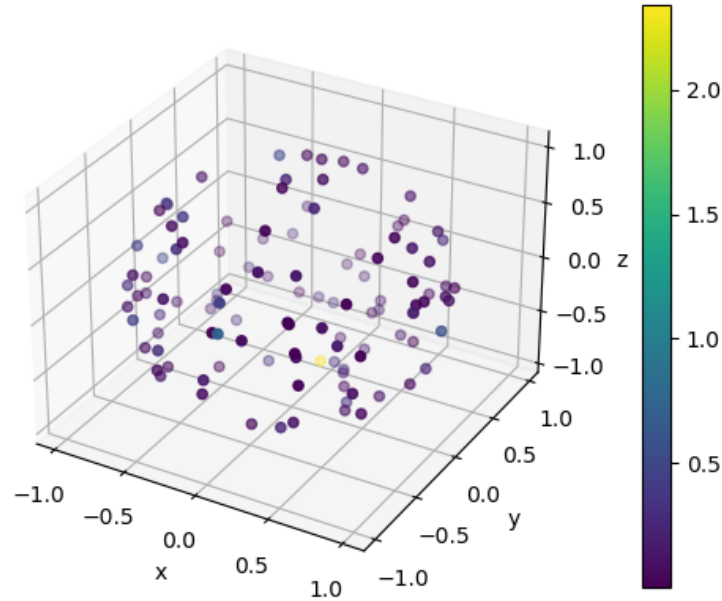


Figure 3.5: Synthetically generated rotations and their relationship with differences in measured turning angles. Colors represent the absolute difference in turning angles, presented in degrees.

3.4 Discussion

In this chapter, we proposed an orientation-invariant gait analysis approach: sensor alignment on the foot is unknown and can be anything. We based our approach on previously studied methods, where some adaptations were introduced in view to maintaining all methods independent of differences in sensor orientation.

For this purpose, orientation-invariant signals were used. To detect ZVIs, we relied on the energy of angular rate, as in [255]. To detect gait events (FC and FO), instead of using mediolateral angular rate and/or forward acceleration (as proposed by [118, 142] or [243]), we used features from acceleration magnitude and from the vertical component of acceleration, obtained in global coordinates.

Instead of representing trajectories using the body coordinate system (as proposed by [118, 243]), we evaluated trajectories in global coordinates, from which SL and SW could be estimated. However, to calculate MTC, we estimated a mediolateral and a forward axis, which were defined with basis on the mediolateral rotation produced by the foot when moving from foot-flat to FO. These axes were required to estimate the angle $\alpha(t)$ —as shown in Equation (3.14)—and could be used to obtain sensor data in body coordinates—similar to the use of PCA, as proposed by Falbriard et al. [248].

3.4.1 Algorithms Selection

The proposed approach was tested in a group of young adults, without any visible gait disorder, using an optical motion capture system as a reference. We tested multiple orientation estimations and double integration approaches. Based on the results achieved on the development set, we selected the Euston CF ($k_p^{opt} = 0.15$, $k_f^{opt} = 0$), combined with direct and reverse integration ($\eta^{opt} = 0.25$, $\alpha_i^{opt} = 0.55$) with active horizontal correction. With an η^{opt} of 0.25, the sigmoid used in direct and reverse integration (Equation 3.11) approximates the linear shape. Yet, the results achieved with the linear and reverse integration were better than the results achieved with the linear dedrifting method. According to [245] and [241], the $k_f^{opt} = 0$ used in Euston CF is an appropriate choice due to the short integration times (below 5–10 min) and the slow dynamics of the movements involved. The horizontal correction mechanism allowed for a compensation of the final vertical position achieved in each stride, assuming walking on a flat surface. Although it improved the results in our study, the method can only be applied when walking on a flat surface, and cannot be generalized to walking on inclined surfaces or stairs.

In [241], the same orientation estimation and double integration methods were benchmarked, using a dataset comprising 20 healthy subjects (between 16 and 80 years old), all using the same shoe model and walking straight at multiple speeds in a path with 10 m. According to the authors, the best-performing orientation estimation method (evaluated by the angles obtained in each axis) was Madgwick CF ($\bar{\omega}_{max} = 3.04$), followed by gyroscope integration and Euston CF ($k_p = 0.0046$, $k_f = 0$); but the performance of the three methods did not differ much. For double integration, the direct and reverse integration ($\eta = 0.08$, $\alpha_i = 0.6$) was the best performing method, which is also according to our results. The optimal set of parameters differed from ours, which may be explained by differences in datasets, namely due to different sensor characteristics, subjects, and protocols. While we applied orientation estimation methods at all instants of the signal, in [241] acceleration data were only used when the magnitude of the acceleration approximated the value of the gravity (i.e., the sensor is roughly static). Moreover, the criteria for evaluating the methods' performance differed from ours. In [241], orientation estimation methods were optimized first based on the measured angles; double integration methods were optimized after selecting the best orientation estimation method, using as criteria the error distribution of estimated velocities and clearance parameters [241]. In our study, we considered that orientation estimation methods have an impact on the performance of double integration methods, which, altogether, have an impact on all gait metrics considered. Therefore, optimizing for a single metric could penalize the performance of the other metrics. For this reason, we opted for the joint selection of methods and the tuning of parameters based on the overall results achieved in all metrics considered.

The results achieved on the development set with the optimized set of methods and parameters is shown in Table 3.3. Even though optimization was performed in this set, some of the metrics differ by some units from the reference system. Table 3.4 shows the results on the validation set. As expected, the central tendency and the dispersion of the errors are lower on the development set where algorithm selection and parameter tuning occurred. The generalization of the method

can only be effectively evaluated using the validation set, which included data from unique users never seen during the optimization process.

3.4.2 Comparison with Other Studies

Using the sensor on foot instep, Mariani et al. [261] tested gait event detection using multiple candidate features (i.e., minimum, maximum, or zero-crossing) and multiple signals—including signals that are independent of IMU orientation, namely acceleration magnitude and the derivative of angular rate magnitude. According to their results, the best candidate features are observed in the acceleration magnitude signal [261], close to the instants we propose to detect FC and FO in this work. On average, FO and FC events were detected 0.01 s (i.e., one sample) before the annotations, but the dispersion of the errors was higher for FO detection. While FCs can generally be well perceived on acceleration signals (they represent an impact of the foot on the ground), TO events may be less evident, which may justify the results. Nevertheless, event detection errors are, in general, low, causing a small impact on estimated temporal parameters.

Our event detection method resulted in absolute errors between 0.02 s and 0.03 s in temporal parameters, with precision between 0.03 s and 0.05 s (as shown in Tables 3.4 and 3.5). These errors represent average differences of just two or three samples at 100 Hz, not differing much whether we include turns or not, which demonstrates the robustness of the method. In [142], relative errors of 0.00 ± 0.07 s, -0.01 ± 0.04 s, and 0.01 ± 0.07 s are obtained for stride, swing, and stance duration in a group of geriatric inpatients, where events are determined based on the analysis of mediolateral angular rate and forward acceleration, assuming a fixed alignment of the sensor on the shoes. In [243], sensors are placed on the shanks, obtaining relative errors of 0.00 ± 0.02 s in stride duration in a group of young adults. Although the performance on temporal parameters is more or less consistent between studies, it is worth mentioning that while studies differ in subjects and protocols applied, the procedures to estimate reference gait events are also different, which hinders the comparison of results. Moreover, these studies are conducted in a laboratory setting, ensuring the precise alignment of the sensors on the body. For this reason, they may not represent the results that would be achieved in a more realistic clinical setting.

In our study, relative errors of -3.5 ± 9.7 cm and -3.1 ± 9.2 cm/s were obtained for SL and gait speed, respectively. When we exclude turns from the analysis, the precision is improved, resulting in relative errors of -3.9 ± 6.2 cm for SL and of -3.4 ± 6.9 cm/s for gait speed. Additionally, RMSE values are lowered and correlations increase (see Tables 3.4 and 3.5). These results show that increased errors in SL (reflected also as increased errors in gait speed) may be related to differences imposed by gait patterns while turning, so the evaluation of these metrics in straight walking seems more reliable. Results would possibly improve if parameter optimization was performed using straight walking exclusively, however, at the cost of making the method less robust.

In [132], relative errors of 1.3 ± 3.0 cm and of 2.8 ± 2.4 cm/s were obtained for SL and speed, respectively, in a group of patients with Parkinson's Disease and age-matched elderly subjects, after discarding turning, initiation and termination cycles. In [142], relative errors of -0.3 ± 8.4

cm were obtained for SL in a group of geriatric inpatients walking straight. Using the same dataset, Hannink et al. [144] achieved relative errors of 0.0 ± 5.4 cm, using a deep convolutional neural network. In [243], SL and gait speed had a performance of 5.4 ± 3.1 cm and 3.4 ± 3.9 cm/s, considering only data from young adults obtained while walking straight. In a group of young and elderly volunteers, Mariani et al. [131] achieved a performance of 1.5 ± 6.8 cm for SL and 1.4 ± 5.6 cm/s for speed, using a protocol that included assessment with U-turn and 8-turn. As we can see from the literature, results are highly heterogeneous and comparison of performance in a fair and robust way is not possible due to the different protocols, subjects, and reference systems employed. Yet, considering that these metrics had very high correlations with the reference system in our study, we can consider our approach appropriate to assess movement performance. This is especially true when we consider straight walking tests, where precision is improved.

Although the precision of SW is improved when turns are excluded (relative error of -0.4 ± 4.4 cm versus 0.1 ± 2.8 cm when turns are excluded) and RMSE is lowered (4.4 cm versus 2.8 cm), the correlation between IMU-based and reference-based SW is worse when turns are excluded. We also observe that, when we exclude turns, the average and standard deviation of the metric decreases, meaning that curves may be associated with increased SW necessary to describe the trajectory of the turn. When we exclude turns, the values of SW are more consistent (closer to the mean) and, as such, harder to correlate, which may justify the results. The same observation is also valid for turning angles where we can see that even though precision is improved (relative error of $0.9 \pm 8.6^\circ$ versus $1.2 \pm 5.1^\circ$), the correlation decreases when we exclude turns (see Tables 3.4 and 3.5). In [132], relative errors of 0.15 ± 2.13 cm and $0.12 \pm 3.59^\circ$ are obtained for SW and turning angle, considering only straight and steady walking.

MTC is one of the metrics with the highest errors, considering their relation to the average of the metric. Relative errors of -0.2 ± 0.8 cm and -0.4 ± 0.7 cm are obtained for MTC when turns are included and excluded, respectively; we observe a small improvement in precision and RMSE when turns are excluded, which again, highlights the challenging conditions possibly imposed by the turns.

In [133], a method is proposed to determine toe trajectories that requires the size of the shoe as an input; using this method, authors achieved relative errors of 1.3 ± 0.9 cm in MTC, in a group of healthy adults walking straight, where only steady walking was included in the analysis. In [249], prior information about shoe dimensions was not required, and relative errors of 1.7 ± 0.7 cm were obtained in a group of young, mid-age and old subjects, where conditions for straight or turning walking are not specified. Although the reported relative errors are similar to our results, higher correlations (0.91) are documented by the authors [249]. However, to achieve these results, Kanzler et al. [249] employ a correction that adjusts the amplitude of toe clearance trajectories—possibly penalizing the generalization of the method. Moreover, the authors discuss that the assumption of the rigid shoe model may not be realistic due to the bending of the shoe at FO, which may also constitute a source of errors in our approach. These errors, combined with a poor estimation of sensor-to-toe distance and errors imposed by the reference—due to a toe marker that is not placed precisely at the tiptoe—can possibly justify the results.

All gait metrics extracted using the IMU present moderate, high, or very high correlations with the reference system. Metrics are also practically equivalent, except for MTC and for turning angle when turns are excluded (Tables 3.4 and 3.5). Based on these results, we can state that, overall, there is a good agreement between both systems, which denotes the potential of the solution. Sensor wireless capabilities, combined with a flexible alignment on the foot—fostered by the orientation-invariant approach—make the proposed solution promising for use at clinics and ambulatory settings.

3.4.3 Limitations and Future Work

Although the reference system used in this study is currently considered the gold standard for gait analysis [118, 131, 132, 243, 249], possible errors can be introduced by the reference. For instance, due to errors in marker labeling or gap-filling operations, the trajectories of the markers may not always reliably replicate the actual trajectories performed. Additionally, the detection of events from trajectories may also display errors that can impact temporal parameters used as reference [252].

Despite our efforts to make the process completely invariant to differences in sensor orientation, we can observe some differences in turning angles when we synthetically rotate sensor data (see Table 3.6). These differences have no apparent correlation with the generated rotations, as can be visually confirmed in Figure 3.5. Moreover, rotations were uniformly sampled to ensure a uniform distribution along all possible 3D transformations (Figure 3.5), which avoided bias on observations—i.e., ensuring that the method is truly invariant to changes in sensor orientation, no matter how it is oriented relative to the feet. The differences in turning angle, not observed in any other metric extracted, may be explained by the sensor fusion process employed to estimate sensor orientation relative to the global frame. While the estimation of the vertical axis employs corrections based on measured acceleration values, in the horizontal plane no correction mechanisms can be employed, and only integration of angular rates is used to estimate changes in orientation (or heading) over time. This may lead to cumulative error propagation due to the integration process, which may differ depending on sensor orientation: different sensor orientations may lead to distinct error propagation profiles. Given the random nature of the noise properties of gyroscope measurements [262], no particular tendency for an increased error towards a specific rotation can be observed. All gait parameters extracted using synthetically rotated data remained practically equivalent (see Table 3.6), which demonstrates the invariance of the method to differences in orientation. Based on these results, we can state that our method does not require careful alignment of the sensor on the foot, which may increase trust and potentially simplify the data acquisition process in the context of clinics.

Future work should address spatiotemporal gait analysis in specific groups (e.g., older adults or groups with a specific pathology), so that validity and robustness of the method to different walking pattern characteristics may be assessed, demonstrating its possible application in real scenarios.

3.5 Conclusion

Foot-worn inertial sensors were used to evaluate a comprehensive set of spatiotemporal gait metrics in a group of young adults. To avoid restrictions on sensor alignment, we proposed an orientation-invariant gait analysis approach and assessed its performance using an optical motion capture system as a reference. Overall, good agreement between both systems was achieved in our study, demonstrating the robustness and reliability of the proposed approach. Additionally, we demonstrated the invariance of the method by simulating the rotations of the sensor on the foot. Taking advantage of this feature, and considering the wireless capabilities of the sensor, we postulate that the proposed solution is highly attractive for use in clinics and ambulatory settings. Its flexibility, combined with low errors achieved in the evaluation of gait parameters, may leverage trust and potentially simplify the data acquisition process. The solution should be evaluated with people with gait-related disorders, so that it may support clinical decision-making in real scenarios.

Chapter 4

A Deep Learning Approach for Foot Trajectory Estimation

This chapter is based on the following publication:

V. Guimarães, I. Sousa, and M. V. Correia, “A Deep Learning Approach for Foot Trajectory Estimation in Gait Analysis Using Inertial Sensors,” *Sensors*, vol. 21, no. 22, p. 7517, Nov. 2021, doi: 10.3390/s21227517.

An instrumented gait analysis resorting to inertial sensors allows the complete evaluation of spatiotemporal gait parameters, offering an alternative to laboratory-based assessments. To estimate gait parameters, foot trajectories are typically obtained by integrating acceleration two times. However, to deal with cumulative integration errors, additional error-handling strategies are required.

In this chapter, we propose an alternative approach based on a deep recurrent neural network to estimate heel and toe trajectories. We propose a coordinate frame transformation for stride trajectories that eliminates the dependency from previous strides and external inputs. Predicted trajectories are used to estimate an extensive set of spatiotemporal gait parameters. We evaluate the results in a dataset comprising foot-worn inertial sensor data acquired from a group of young adults, using an optical motion capture system as a reference.

Heel and toe trajectories are predicted with low errors, in line with reference trajectories. A good agreement is also achieved between the reference and estimated gait parameters, in particular when turning strides are excluded from the analysis. The performance of the method is shown to be robust to imperfect sensor-foot alignment conditions.

4.1 Introduction

Inertial sensors placed on the feet allow the complete evaluation of spatiotemporal gait parameters, being currently accepted as one of the best positions to evaluate gait [138]. To evaluate spatiotemporal gait parameters, the foot trajectory is typically obtained by integrating acceleration two times [143, 241]. This process requires not only a good estimation of sensor orientation (to isolate linear acceleration from the acceleration due to gravity), but also a good strategy to deal with cumulative double integration errors. To this purpose, techniques based on a zero-velocity drift correction (applied to the moments when the foot is in contact with the ground) are commonly employed [131, 132, 142, 241]. The complexity of the required error handling strategies—often designed and optimized for a specific patient group [131, 132, 142, 241]—still make the estimation of foot trajectories a challenging task.

Using some specific trigonometric relationships, previous studies could estimate the vertical trajectories of the heel and toe from inertial sensor data acquired at the dorsum of the foot [133, 249]. In [133], shoe size was also used as an input, while in [249] heel and toe clearance metrics were obtained without previous knowledge of shoe dimensions. The estimation of heel and toe clearance metrics depends on the reliable estimation of foot trajectories [133, 249]. Moreover, both studies assumed a rigid shoe model, which may not be realistic and introduce additional errors, e.g., due to the bending of the shoe at foot off [249].

To overcome the limitations of the conventional gait analysis and trajectory estimation methods, we propose a deep learning approach. We cast the gait analysis problem as a sequential learning problem that allows us to estimate heel and toe trajectories, while walking, on a stride-by-stride basis, using information from inertial sensors placed on the shoes. Deep learning has previously been applied to the context of gait analysis, but, to the best of our knowledge, this is the first study using deep learning techniques to estimate three-dimensional (3D) heel and toe trajectories from inertial sensor data; therefore, emulating the trajectories' data obtained from heel and toe markers in gold standard optical motion capture systems. Using these techniques, we overpass the need of designing and optimizing conventional approaches based on orientation estimation and double integration. Moreover, heel and toe trajectories are estimated in 3D, overcoming the limitations of previous heel and toe clearance estimation methods that could only evaluate the trajectories of the vertical component [133, 249]. From heel and toe trajectories, an extensive set of spatiotemporal gait parameters can be obtained [132, 143, 253, 254].

Previous works in gait analysis used deep learning to regress gait parameters directly from inertial sensor data. In [144], the stride length was obtained from inertial sensor data using a deep convolutional neural network. In [263], a similar architecture was used to estimate several parameters of interest, including the stride length, width, foot angle, heel contact time, and toe contact time. Even though these studies offered promising results, the proposed inference models could only predict gait parameters that were included in the training stage [144, 263]. To integrate additional parameters, further training (and possibly a change in network architecture) would be required, which constitutes a limitation of these approaches. In contrast, the prediction of heel and

toe trajectories allows the estimation of many spatiotemporal gait parameters of interest, which constitutes a more powerful approach.

Previous studies in odometry proposed the use of deep learning strategies to estimate the trajectories of a moving body. In the context of odometry, (naive) approaches based on double integration usually lead to critical errors due to integration drift, as zero-velocity updates are not always possible [264, 265]. Studies have used deep neural networks to predict changes in pose, allowing the reconstruction of trajectories in 2D or in 3D [264, 265]. In [266], and in [267], a deep neural network could estimate changes in distance and/or heading, enough to reconstruct the trajectories of a pedestrian. By predicting changes in pose, or in distance and heading, these studies rely on trajectory reconstruction mechanisms (incrementally composed) that, over time, would accumulate errors [264, 265, 267]. This formulation is, as such, inadequate for the context of gait analysis.

To eliminate the dependency from previous strides and from external information (e.g., shoe length), we propose a new representation for heel and toe trajectories that is obtained for each stride. This formulation avoids the accumulation of errors due to trajectory reconstruction mechanisms, that, under this formulation, are not required. We propose, optimize, and evaluate a deep recurrent neural network to predict heel and toe trajectories from inertial sensor data, and use the predicted trajectories to extract and evaluate spatiotemporal gait parameters. We evaluate the results in a dataset comprising inertial sensors and reference data acquired from a group of young adults. We compare the results obtained using the deep learning approach with a conventional approach relying on orientation estimation and double integration techniques. Finally, we demonstrate that our model can generalize, being robust to different orientations of the sensor on the feet. This feature is important to ensure that the method is robust to imperfect sensor–foot alignment conditions, which may support its application in real scenarios outside the lab.

This chapter describes and discusses our findings in relation to previous works in related fields.

4.2 Materials and Methods

4.2.1 Dataset

The dataset thoroughly described in the previous Chapter 3, Section 3.2, was used in this study. It consisted of acceleration and angular rate measured by two inertial measurement units (IMUs), and marker trajectories acquired by an optical motion capture system (Vicon, Oxford Metrics). IMU sensors were placed on the dorsum of the shoes, as shown in Figure 4.1–a, tightly secured with an elastic band to reduce motion artifacts. Optical markers were placed on the shoes, using the marker set illustrated in Figure 4.1–b, which included one marker near the heel and another marker near the toe (highlighted in Figure 4.1–b). Markers' trajectories were captured by 10 infrared cameras (Vicon Vero v2.2), resulting in a capture volume of around 3 per 8 m. Data were acquired at a sampling rate of 100 Hz, the rate most commonly used in gait analysis [138].

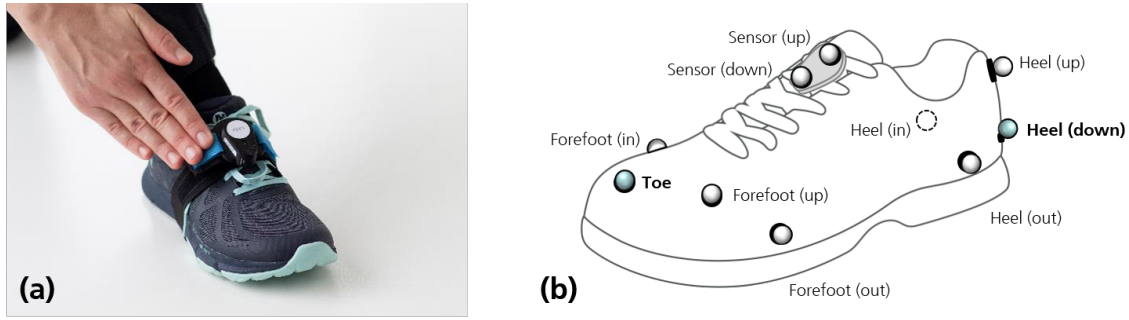


Figure 4.1: Placement of the inertial sensor (a) and markers (b) on the shoes.

The dataset comprised walking data from 26 healthy young adults (average age of 29.2 ± 5.3 years, 13 females and 13 males), each using their own shoes. Participants performed repeated walking trials under different conditions, including different speeds (comfortable, slower, and faster self-selected speeds) and different directions (clockwise and counterclockwise). Each walking trial included three consecutive laps over the lab’s capture volume, each lap comprising straight walking (along the length of the capture volume) and turns (along the width of the capture volume). Self-selected speeds ranged from 0.41 to 2.01 m/s (average 1.09 ± 0.31 m/s).

The study received approval from the Ethical Committee of the University of Porto (81/CE-UP/2019) and all participants provided written informed consent.

4.2.2 Reference Trajectories and Gait Parameters

Marker trajectories were used to detect gait events and parameters. Before detecting them, we estimated the steady periods of walking, i.e., the moments when the foot was not moving and was in contact with the ground. These periods were identified using a threshold-based approach—applied to the velocity of the centroid of heel and toe trajectories—, as described in [253]. These periods were used to estimate the vector normal to the horizontal plane of movement—resorting to Principal Component Analysis (PCA)—, which was used to calibrate the lab’s horizontal plane, as described in Section 3.2.2 (Chapter 3).

Reference gait events and parameters were automatically extracted from marker trajectories using a Python routine. Foot contact (FC) and foot off (FO) events were detected from local minima and maxima occurring on vertical heel velocity and vertical toe acceleration, using the method described in [253]. Before detecting gait events, trajectories were low pass filtered using a zero-lag bidirectional first-order Butterworth filter (cutoff of 20 Hz), as described in Section 3.2.2 (Chapter 3).

Temporal gait parameters—stride, stance, and swing duration—were obtained from FC and FO events. Cadence was calculated as the inverse of stride duration, in units of steps per minute. Spatial gait parameters—stride length (SL), swing width (SW), and minimum toe clearance (MTC)—were calculated as illustrated in Figure 4.2, using the trajectories of the heel and toe. Gait speed was obtained by dividing stride length by its corresponding stride duration.

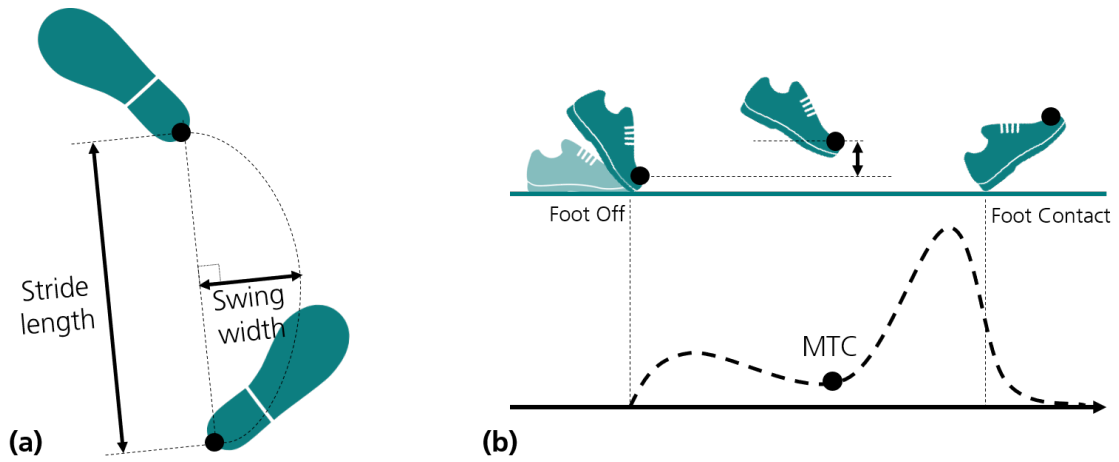


Figure 4.2: Definition of spatial gait parameters based on heel and toe trajectories: (a) Stride length and swing width. (b) Minimum toe clearance (MTC).

4.2.3 Data Preparation

Inertial sensor data were synchronized with reference trajectories using the maximum cross-correlation between acceleration magnitude—obtained from the IMU—and the centroid of heel and toe trajectories, as described in Section 3.2.5 (Chapter 3).

Before feeding data to the neural network, a series of preprocessing steps were conducted, which included segmentation of continuous recordings, coordinate transformations, padding to a fixed length, and normalization. Preprocessing steps were applied to the right and left feet data without any distinction.

Spatial gait parameters, such as stride length, are typically defined within mid-stance events, i.e., between the midpoints of steady periods. For this reason, mid-stance events (detected from the reference trajectories) were used to segment the data. In [144], the best results for stride length estimation using a neural network were achieved when strides were defined between mid-stance events. Using this formulation, the start and end points of each segment (or stride) corresponded to instants when the foot was completely in contact with the ground, providing a stable basis for spatial metrics calculation and coordinates transformation—as described next.

Considering that states such as attitude or position cannot be directly observed from inertial sensor data, they need to be derived from previous states and, for this reason, data segments are not independent [265]. To deal with these constraints, some authors proposed to estimate a change in state (delta) rather than the state of the system [264, 265]. In this work, we proposed to represent trajectories within a stride using a new coordinate system that did not depend on the previous stride. The new coordinate system was obtained after subtracting the origin, followed by rotating the global coordinate system (lab) so that the main movement direction (formed by the vector connecting the start and end points of the trajectory in the stride, in the horizontal plane) corresponded to the x-axis in the new coordinate system. The vertical direction aligned with the lab's vertical direction (z-axis), and the y-axis was perpendicular to the x and z axes. A new

coordinate system was obtained for each stride, and for heel and toe trajectories independently, as illustrated in Figure 4.3.

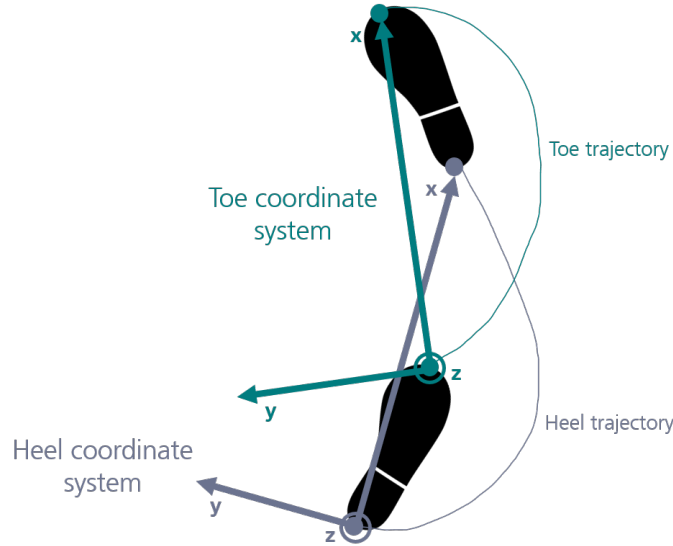


Figure 4.3: Coordinate systems of the heel and toe stride trajectories.

The trajectories in this new coordinate system can be obtained from inertial sensor data resorting to double integration and orientation estimation techniques, without considering any inputs from previous strides [268]. While gravity provides a reference to the vertical component of trajectories (z-axis), horizontal components can be obtained (by double integration) using any convention (i.e., pointing to any horizontal direction), and later transformed so that the x-axis can be derived from the vector connecting the start and end points of the trajectory in the stride [268]. This formulation ensures that strides are independent. Moreover, as heel and toe coordinate systems were independently defined, no prior knowledge concerning the relationship between heel and toe markers (e.g., their distance) was required under this formulation. Previous studies demonstrate that vertical heel and toe trajectories can be obtained from inertial sensor data acquired at the dorsum of the foot, without any additional input [249, 268], which supports the theoretical background of the proposed deep learning approach. Gait parameters could be directly extracted from heel and toe trajectories under this new formulation, as explained in detail in Section 4.2.7.

Rotation was applied to inertial sensor data sequences for data augmentation purposes. Rotations were randomly sampled from a uniform distribution of quaternions, using Equation 3.17 from Chapter 3, as proposed in [259]. Five randomly sampled artificial rotations were added to each data segment so that different rotations were applied to each data segment and the number of sequences was increased sixfold. Generated rotations were uniformly distributed across the whole 3D space. With this process, we not only increased the total number of sequences in the dataset, but also introduced variations that may support a better generalization performance of the model, robust to different sensor orientations or sensor placement errors. Data augmentation through rotation is a common strategy applied in this domain [269–271].

To ensure that all sequences (inertial sensor data and trajectories) had the same size, they were padded with their initial value to obtain a fixed size of 256 samples—as proposed in [144].

The dataset was then split into three distinct partitions: training, validation, and test. Approximately 20% of the strides, from unique randomly selected participants, were included in the test set. The remaining 80% (60% for training, and 20% for validation—also from unique randomly selected participants) were used for the optimization of hyperparameters and model fitting (Table 4.1).

Table 4.1: Dataset splitting.

Partition	Number of Subjects	Number of Strides
Training set (60%)	14	31,470
Validation set (20%)	6	13,872
Test set (20%)	6	12,474

Accelerations, angular rates, and trajectories were normalized using their respective maximum values, computed along all strides in the training set.

The operations concerning data preparation, network training, and model performance evaluation are summarized in Figure 4.4. Network training and model performance evaluation are presented in detail in the next sections.

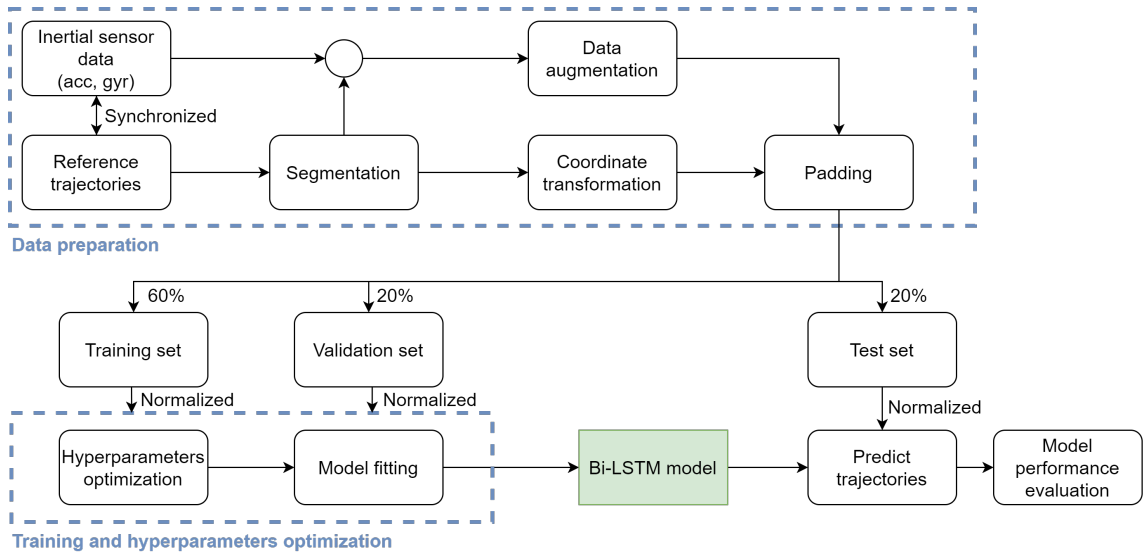


Figure 4.4: Data analysis process diagram: data preparation, network training, and model performance evaluation.

4.2.4 Network Architecture

A stacked bidirectional long short-term memory (LSTM) recurrent network was used in this study, comprised of two bidirectional LSTM layers. LSTM networks are an extension of the vanilla recurrent neural network (RNN) architecture, designed to explore temporal dependencies within sequential data. The LSTM cell has a more complex structure and was shown to better handle

long-term dependencies within the data [272]. LSTM networks have been proposed to solve problems dealing with wearable inertial sensor data, e.g., in the area of odometry [264,265], pedestrian dead reckoning [266], kinematics [270,273–275], attitude estimation [271,276], fall risk assessment [269], activity recognition [277,278], among others.

We specifically used LSTM layers in a bidirectional architecture, i.e., the network included forward and backward LSTM layers that were simultaneously trained, so that both time directions (past and future) were considered and contributed to the outcome [272]. The bidirectional architecture improved model performance in problems dealing with inertial sensor data, e.g., in [265], or in [279]. In the gait analysis domain, double integration with linear dedrifting or direct and reverse integration are typically applied to bound integration errors, considering that the foot should have a velocity of zero in the initial and final instants of the stride [118,241,242]. A bidirectional architecture was proposed here to deal with these domain-specific constraints. The outputs of the forward and backward layers were combined using the sum.

The first bidirectional LSTM layer received the input sequences, i.e., acceleration and angular rate data, with dimension $(N_{strides} \times 256 \times 6)$, where $N_{strides}$ corresponds to the number of strides, 256 is the length of the sequence, and 6 is the number of axes (3 axes for acceleration data plus 3 axes for angular rate). The second bidirectional LSTM layer output heel and toe trajectories arranged in a multidimensional sequence with the same dimension $(N_{strides} \times 256 \times 6)$, where 6 represents the number of axes (3 axes for toe trajectories plus 3 axes for heel trajectories). Right and left inertial sensor data were fed to the network without any distinction.

4.2.5 Training and Hyperparameters Optimization

The neural network was trained to model the relation between inertial sensor data and heel and toe trajectories within a single stride. As a loss function, we used the mean squared error (*MSE*), as depicted in Equation 4.1.

$$MSE = \frac{1}{N} \sum_{n=1}^T (y_n - \hat{y}_n)^2 \quad (4.1)$$

where y_n and $\hat{y}_n$ are the flattened reference and estimated sequences, and N is the number of elements in a sequence (i.e., 256 samples multiplied by 6 axes).

For the training, we used the Adam optimizer, one of the most popular gradient-based optimization methods for stochastic learning [280]. Random subsets of the training set (called mini-batches) were shown to the optimizer in one iteration of the training loop (or epoch) to speed up the learning phase (stochastic learning). The data were re-shuffled at every epoch, to avoid bias errors due to training data order. The mini-batch error was summarized with the average of the individual MSE calculated for each sequence in the batch.

To achieve the best results, we optimized hyperparameters using an automated hyperband search [281], as provided by Keras-tuner [282]. In this process, we also explored the addition of a dropout layer after the first bidirectional LSTM layer, that would randomly drop a given number of nodes during training to prevent overfitting [283].

Hyperparameters comprised the number of units in the first bidirectional LSTM layer, the dropout rate (where a zero rate implied no dropout), the learning rate in the Adam optimizer, and the size of the mini-batches. The configuration space considered in the hyperband search is shown in Table 4.2.

Table 4.2: Hyperparameters optimization.

Parameter	Configuration Space	Optimal Configuration
Number of Units	32–160	160
Dropout	0–0.30	0.1
Learning Rate	1×10^{-6} – 1×10^{-2}	1×10^{-3}
Batch Size	100–400	100

The best hyperparameters were selected with basis on the validation loss, obtained from the sequences in the validation set. The hyperband was configured to run for a maximum of 350 epochs (expectedly higher enough to converge—according to some previous tests), using a reduction factor of three, and two hyperband iterations [281, 282]. We also configured the hyperband search to perform three executions per trial—and summarize the trial with the average validation loss—to account for random influences (e.g., due to dropout, training batch selection, etc.). An early stop criterion was applied to the training process, so that the training would stop if no improvement in the validation loss was verified during the next 50 epochs. The optimal configuration obtained using the hyperband search is shown in Table 4.2, and the final network architecture is shown in Figure 4.5.

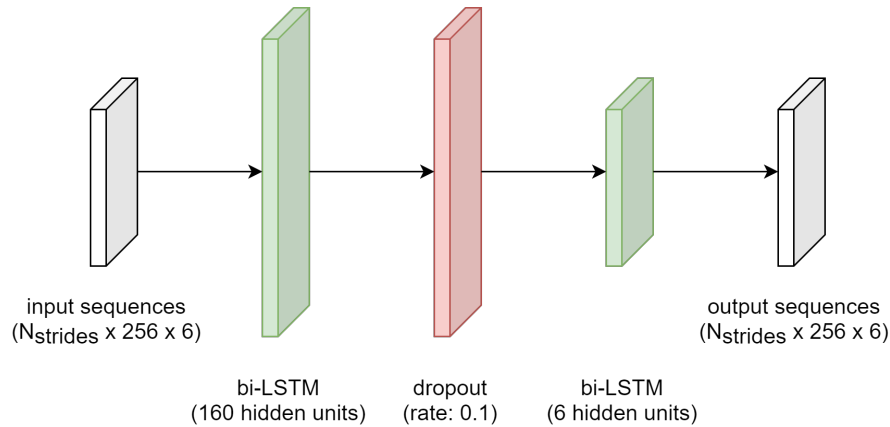


Figure 4.5: Neural network architecture.

The best set of hyperparameters was used to train the final model, using the training set, a maximum number of 350 epochs, and early stopping. The model with the best validation loss throughout the training was chosen as the final one for the testing.

The networks were trained using Python (<https://www.python.org/>, accessed on 7 July 2021) 3.7.9, Tensorflow (<https://www.tensorflow.org/>, accessed on 7 July 2021) 2.2.0 and Keras (<https://keras.io/>, accessed on 7 July 2021) 2.3.0, using the NVIDIA® V100 GPU with 16 GB of memory.

4.2.6 Model Performance Evaluation

While the validation set was used to evaluate models during hyperparameters tuning and model fitting, the test set was used to determine the generalization performance of the model.

Because of the normalization applied during data preparation (Section 4.2.3), the model predictions (heel and toe trajectories) obtained from the test samples needed to be rescaled back to the original (physical) units, using the inverse operation, i.e., using the scaling parameters calculated from the training set. Contrarily to the training process, during testing, the full architecture of the network was used without dropping any connections.

To report the performance of the model we used MSE, mean absolute error (MAE), root mean squared error (RMSE), and average Euclidean distance. These metrics have been commonly used in the literature to report the performance of estimated trajectories [264, 265, 267].

4.2.7 Gait Parameters from Predictions

Gait parameters were extracted from heel and toe trajectories, which, in turn, were obtained from inertial sensor data using the deep learning model. The full pipeline required inertial sensor data to be segmented, and each segment normalized with regard to sequence length and scale (using the scaling parameters calculated from the training set) before generating model predictions. The full data analysis process concerning the estimation of gait parameters from deep learning predictions is illustrated in Figure 4.6.

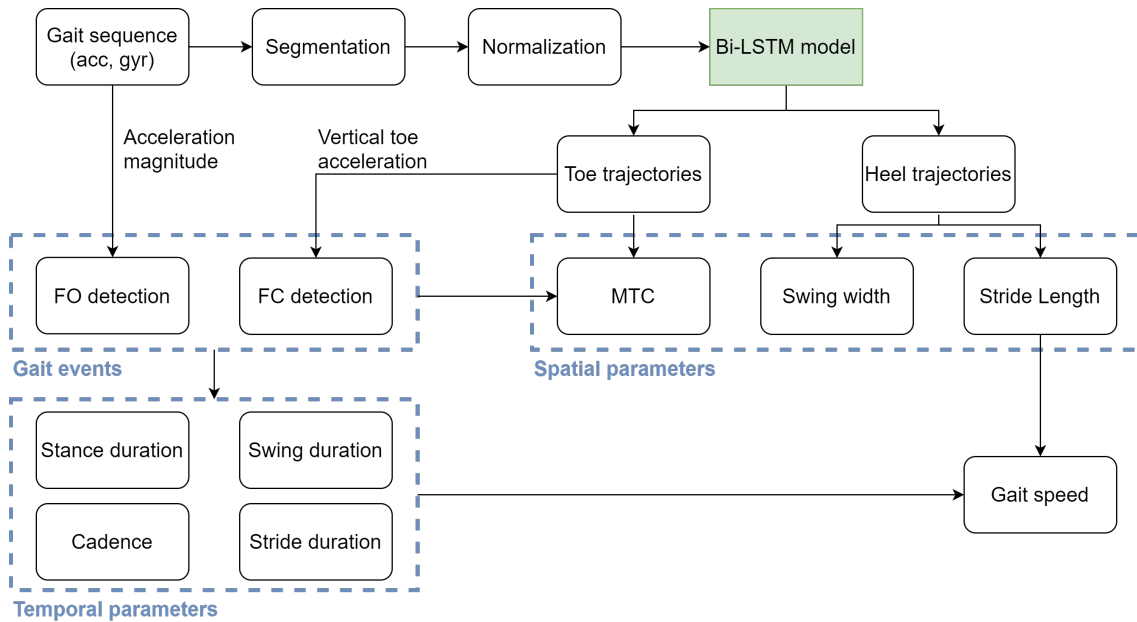


Figure 4.6: Data analysis process diagram: estimating gait parameters from predicted heel and toe trajectories.

To segment inertial sensor data, steady periods were first detected using the angular rate energy detector [255]. A sliding window with 0.15 s was used to calculate the energy of the angular rate

magnitude, as in Section 3.2.4.1 (Chapter 3). The threshold was calculated based on the average of the energy, as described in Section 3.2.4.1 (Chapter 3). Mid-stance events were used to segment continuous inertial sensor data recordings.

Generated segments were then padded with their initial values to obtain a fixed size of 256 samples. Using the scaling parameters calculated from the training set, segments were then normalized and used as input to the neural network. Predicted trajectories were converted back to the original units using the scaling parameters of the training set, and expressed in meters.

FO and FC events were detected as described in Section 3.2.4.4 (Chapter 3). The FO event was considered the first maximum after a steady interval, obtained from the acceleration magnitude (calculated from raw inertial sensor data). The FC event was considered an absolute minimum between FO and the next steady interval, obtained from the vertical acceleration. To obtain this signal, predicted toe trajectories were derived two times, revealing a signal pattern similar to the one used in Section 3.2.4.4 (Chapter 3) to extract FC events (in Section 3.2.4.4 vertical acceleration was calculated from inertial sensor data after the estimation of sensor orientation). Detected events and corresponding signals are shown in Figure 4.7.

Temporal gait parameters, i.e., stride, swing, and stance duration, were determined from FO and FC events. To estimate stride length, we calculated the maximum displacement achieved in the main direction of movement, i.e., the x-axis, in predicted heel trajectories. Swing width was directly obtained from the y component of predicted heel trajectories, considered the maximum absolute value obtained in this direction. MTC was estimated directly from the vertical component (z-axis) of the toe trajectories; it was obtained from a minimum between FO and the next FC event, as illustrated in Figure 4.2. Finally, gait speed was obtained by dividing stride length by its corresponding stride duration.

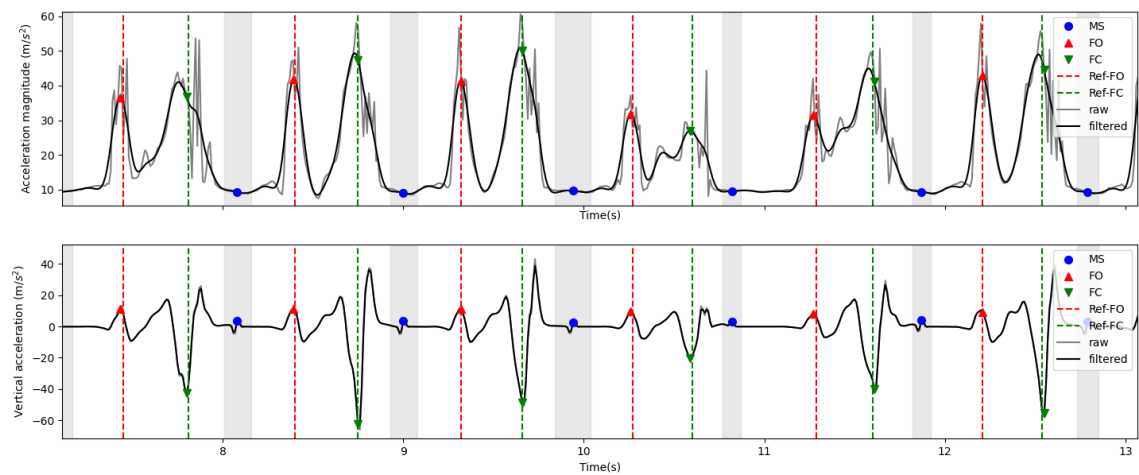


Figure 4.7: Events detection from acceleration magnitude (obtained from the inertial sensor data) and vertical acceleration (obtained from predicted toe trajectories). FO—foot off; FC—foot contact; MS—mid-stance; Ref-FO—reference FO; Ref-FC—reference FC.

4.2.8 Instrument Comparison and Validation

This section describes the metrics used to compare gait parameters extracted from predicted trajectories (obtained from the inertial sensor data) and reference gait parameters (obtained from reference trajectories). All evaluation metrics were calculated using the test set.

FC events, and their corresponding stride gait parameters, were classified as true positive cases if they occurred within a tolerance of 0.1 s relative to the reference. Only these cases were considered for comparison with the reference system. Reference strides without any corresponding IMU-derived candidate were classified as not detected.

IMU-derived and reference gait parameters were compared on a stride-by-stride basis, using the metrics of accuracy (mean of relative and absolute differences), precision (standard deviation of relative and absolute differences), and RMSE. We also reported the 95% limits of agreement, as introduced by Bland Altman [258]. The correlation between instruments was calculated using Pearson's (r_p)—in case of normal distribution—or Spearman's (r_s)—when data could not be assumed to be normally distributed. We also reported equivalence tests using an equivalence zone of $\pm 5\%$ of the average of the metric. Equivalence tests were based on paired T -test (T)—for parametric—or Wilcoxon signed-rank test (W)—for non-parametric. To choose the appropriate test, data were first tested for normal distribution using Shapiro–Wilk test.

As straight walking tests are most commonly used to assess gait disorders [65], we repeated the analysis considering the scenario where only straight walking strides were included. As in [132], turning strides were defined as strides with a turning angle above 20 degrees.

Finally, to test if the network's inference capabilities (as well as the resulting gait parameters) were robust to differences in sensor orientation, we applied the same method as described in Section 3.2.5.3 (Chapter 3). To that purpose, we simulated multiple rotations of the IMU on the shoes using Equation 3.17 (Chapter 3). The resulting quaternions were uniformly distributed across all 3D space [259] and were used to synthetically rotate raw inertial sensor data in a walking trial. We compared gait metrics extracted from the original sensor orientation with those extracted from a rotated version of the sensor. To quantify differences, we reported the Root Mean Square Deviation (RMSD), correlation, and equivalence tests, as described above.

A significance level (p -value) of 5% was used to evaluate the results.

4.2.9 Comparison with the Conventional Gait Analysis Approach

The conventional gait analysis approach relies on orientation estimation and double integration methods, from which an estimate of 3D sensor trajectories can be obtained [131,132,142,143,241]. To obtain orientation relative to a fixed frame of reference, we integrated gyroscope data and used the moments when the foot was in contact with the ground to obtain an initial estimate of sensor inclination relative to the vertical (based on gravity measured by the accelerometer). This method was previously described and validated for foot-worn inertial sensors [131, 132]. Linear acceleration (i.e., acceleration excluding gravity) was obtained from the calculated orientation quaternions. Double integration of linear acceleration, expressed in the fixed frame of reference,

allowed the estimation of sensor trajectories. To bound the errors, linear dedrifted was applied between zero-velocity intervals, as described in [131, 242]. Sensor trajectories were expressed on a stride-by-stride basis, using the method described in Section 4.2.3. Gait events were obtained as described in Section 4.2.7, i.e., using acceleration magnitude to detect FO events and vertical sensor acceleration to detect FC events. Stride length and swing width were obtained from the estimated sensor trajectories. MTC was estimated assuming a rigid shoe model and using trigonometric relationships as described in [249, 268].

Sensor trajectories, estimated using the conventional approach, were compared with reference sensor trajectories obtained from the centroid of the two markers placed on the sensor (see Figure 4.1). Trajectories and gait parameters were evaluated using the mean and standard deviation of the absolute differences.

4.3 Results

4.3.1 Network Training and Trajectories Estimation

The optimal results of the hyperparameter search are shown in Table 4.2. The optimal configuration used a batch size of 100 samples, a learning rate of 1×10^{-3} , and a dropout rate of 10%. For the first bidirectional LSTM layer, the best configuration was achieved with 160 units. Two hyperparameters—number of units in the first layer and batch size—were optimally defined at the upper and lower boundaries of the search space. The hyperparameter optimization process took approximately 1 day and 6 h.

Figure 4.8 shows the mean squared error (loss) achieved on the training and validation sets while training the model with the selected hyperparameters. The training loss decreased throughout the training, accompanied by a decrease in the validation loss. The best validation loss was achieved in epoch 118 due to early stopping; the training stopped after 50 epochs, at 168.

The deep learning model was used to predict heel and toe trajectories from inertial sensor data. Figure 4.9 shows an example of predicted and reference trajectories, obtained from a sample in the test set. Each graph shows a different component of the trajectories: on the first graph, the x component is shown, representing the main movement direction (stride length); the second graph shows the y component, or, the width of the strides; the bottom graph shows the vertical component of the trajectories (z-axis), or heel and toe clearance.

In all the components (x, y, and z), the shapes of the predicted trajectories were generally in agreement with reference trajectories. Moreover, the temporal alignment of predicted sequences was in line with the reference (Figure 4.9).

The performance of the neural network is presented quantitatively in Table 4.3, using the metrics of MSE, MAE, RMSE, and the Euclidean distance. The performance on the validation set was higher than the one achieved on the test set for both heel and toe trajectories, for all metrics evaluated. Heel and toe trajectories were predicted with an equivalent performance both in validation and test partitions.

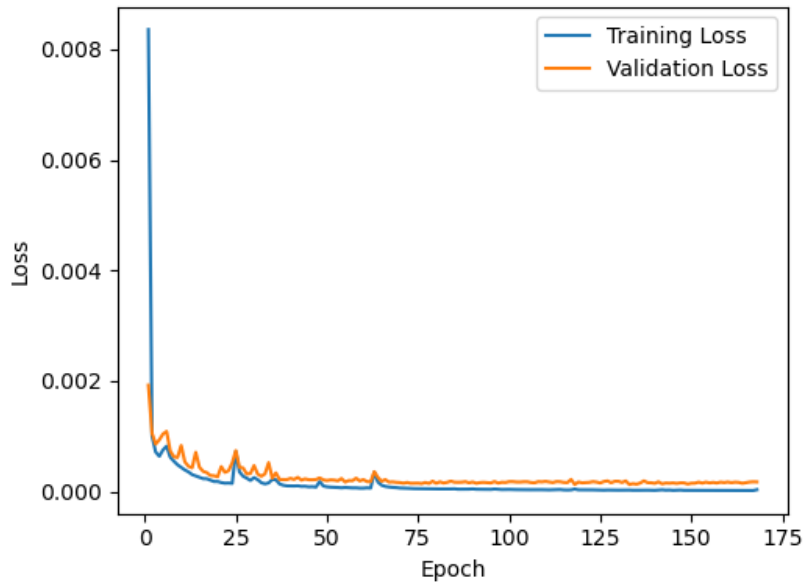


Figure 4.8: Mean squared error (loss) achieved while training the model with the best selection of hyperparameters. The loss was calculated from normalized trajectories.

4.3.2 Gait Parameters Using the Deep Learning Approach

Based on FC events (obtained from vertical acceleration derived from predicted toe trajectories), 2068 strides out of 2151 reference strides (i.e., approximately 96.1%) were considered true positive cases, i.e., only 83 strides (3.9%) were classified as not detected. FC events were detected with a relative error (i.e., relative difference between detected and annotated event times) of -0.01 ± 0.02 s and an absolute error (i.e., absolute difference between detected and annotated event times) of 0.02 ± 0.02 s. FO events (obtained from raw acceleration magnitude) were detected with a relative error of -0.01 ± 0.03 s and an absolute error of 0.02 ± 0.02 s. Thus, on average, FC and FO events were detected before the occurrence of the reference event.

The resulting performance on temporal and spatial gait parameters, obtained on the test set, is shown in Table 4.4.

Using the classification scheme proposed in [260], all metrics presented a high (between 0.70 and 0.90) or very high (above 0.90) correlation with the reference, except MTC which had a negligible correlation (below 0.30). The gait speed was estimated with a relative error of -2.2 ± 10.5 cm/s and an absolute error of 5.8 ± 9.0 cm/s. All metrics were practically equivalent (with $p < 0.01$), except SW and MTC.

Table 4.5 summarizes the performance on the test set when turns were excluded from the analysis.

RMSE values improved for almost all metrics (except for stride and swing duration, where the RMSE value maintained) when turns were excluded from the analysis. Absolute errors (accuracy and precision) and relative errors (precision) also improved for most of the parameters. When turns were excluded, all metrics presented a very high (above 0.90) correlation, except SW—which had a high correlation (above 0.70)—and MTC—which maintained a negligible correlation (below

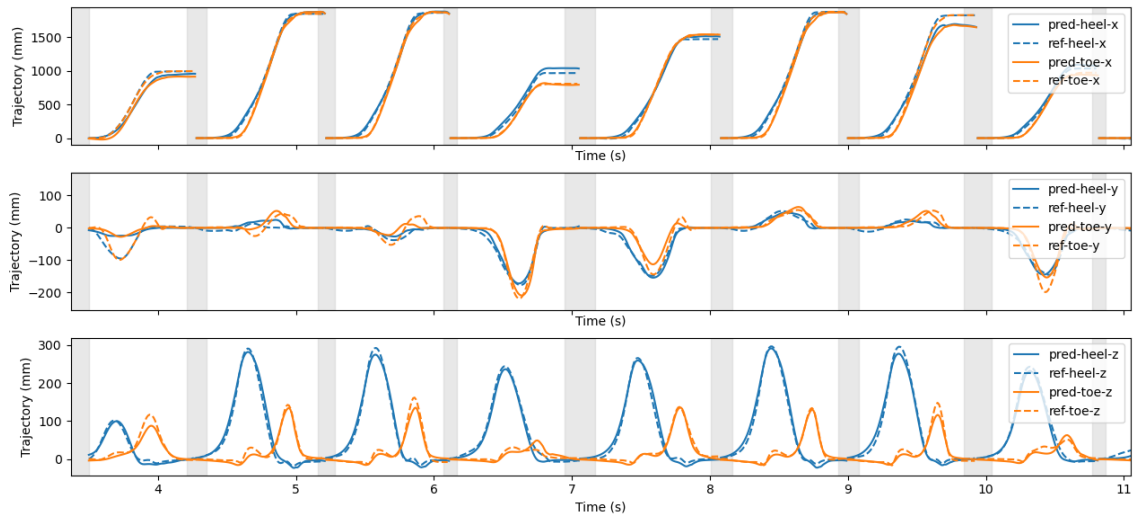


Figure 4.9: An example of predicted (full lines) and reference (dashed lines) trajectories, obtained from a sample in the test set. Heel and toe trajectories are shown in different colors.

0.30). All metrics were considered practically equivalent (with $p < 0.01$), except SW and MTC. Average relative and absolute errors of -2.2 ± 7.0 cm/s and 3.9 ± 6.2 cm/s were obtained for the gait speed.

4.3.3 Robustness to Changes in Orientation

Table 4.6 summarizes the results of the comparison between gait parameters extracted using original IMU data and simulated rotations of the sensor.

The correlation between variables was very high for all metrics (above 0.90), except for MTC, where the correlations were high (above 0.70). When data were synthetically rotated, all parameters remained practically equivalent, with $p < 0.01$.

4.3.4 Comparison with the Conventional Gait Analysis Approach

Table 4.7 summarizes the comparison of the performance achieved with the deep learning approach and the conventional approach described in Section 4.2.9. As mentioned in Section 4.2.9, the conventional approach could only estimate 3D sensor trajectories, while the deep learning approach was able to predict 3D heel and toe trajectories. Mean absolute errors were lower for heel and toe trajectories—predicted using the deep learning model—than for sensor trajectories—estimated using the conventional approach—, whether turns were included or not. The resulting gait parameters were also estimated with lower mean absolute error and standard deviation when using the deep learning approach. The differences in performance were more apparent on spatial gait parameters—stride length, SW, and MTC—and gait speed than on temporal gait parameters. When turns were excluded from the analysis, the trajectories and gait parameters were estimated with a lower error. In both scenarios (i.e., including and excluding turns), better results were achieved when the deep learning approach was employed.

Table 4.3: Performance of the neural network on validation and test sets.

Metric	Validation Set		Test Set	
	Heel Traj.	Toe Traj.	Heel Traj.	Toe Traj.
MSE (mm ²)	538.0	534.6	1043.7	1087.7
MAE (mm)	7.6	7.0	8.6	8.2
RMSE (mm)	14.2	13.3	17.2	16.8
Eucl. Dist. (mm)	17.9	16.6	20.2	19.5

4.4 Discussion

In this study, we proposed a novel IMU-based gait analysis solution based on a deep recurrent neural network. The model was used to estimate foot trajectories, being proposed as an alternative to the traditional orientation estimation and double-integration approaches found in the literature [131–133, 142, 241, 249, 268]. A stacked bidirectional LSTM network was used to estimate heel and toe trajectories, from which clinically relevant gait parameters could be extracted. Dissimilar to the gold standard instrumented gait analysis conducted in the lab, the proposed method relies on the use of foot-worn wearable sensors that can be applied outside the lab, e.g., in a clinical setting. To better support operation under an imperfect sensor placement (with regard to its orientation on the shoes), we proposed a data augmentation scheme, and evaluated the robustness of the model to changes in orientation. Results were evaluated in a dataset comprising foot-worn inertial sensor data and reference data acquired from a group of young adults.

4.4.1 Problem Formulation

The network architecture used in this study was inspired by past works in related fields, namely, in [265], [274] or [275]. In kinematics, in particular, LSTM networks were proposed to predict multidimensional sequences of kinematic variables—e.g., joint angles and moments—from multidimensional sequences of inertial sensor data—acceleration and angular rotation [274, 275]. Using a similar approach, we proposed to predict multidimensional trajectories from multidimensional IMU data. Final results were reported from data on the test set that included samples (and subjects) never seen by the model during the training process. These results, thus, represent the generalization performance of the model.

To optimize the results, hyperparameter selection (using the hyperband search) was conducted prior to model fitting, and the validation set was used to monitor training performance (Figure 4.8). The optimal results of the hyperparameter search, in particular, the two hyperparameters—number of units on the first LSTM layer and batch size—were equal to the upper or lower boundaries of the search space (Table 4.2). Although this may indicate that better configurations could have been obtained by searching among wider configuration spaces, due to training time constraints, the boundaries of the search were not further decreased or increased, as it would result in a significant increase in training time. Still, trajectories were predicted with a low error compared to

Table 4.4: Performance on the test set, including turns ($n = 2068$ strides).

Parameter	IMU	VICON	Rel. Error	Abs. Error	Lim. Agr.	RMSE	Corr. [†]	Equival. [‡]
Stride dur. (s)	1.15 (0.20)	1.15 (0.20)	0.00 (0.04)	0.02 (0.04)	[−0.08, 0.08]	0.04	0.99	±0.06 (0.0)
Swing dur. (s)	0.40 (0.07)	0.39 (0.06)	0.00 (0.04)	0.02 (0.03)	[−0.07, 0.08]	0.04	0.88	±0.02 (0.0)
Stance dur. (s)	0.75 (0.15)	0.76 (0.16)	−0.01 (0.03)	0.02 (0.02)	[−0.06, 0.05]	0.03	0.98	±0.04 (0.0)
Cad. (st/min)	107.5 (17.4)	107.5 (17.8)	0.1 (4.7)	1.8 (4.4)	[−9.2, 9.4]	4.7	0.99	±5.37 (0.0)
SL (cm)	126.7 (27.3)	129.4 (28.4)	−2.6 (10.0)	5.8 (8.6)	[−22.3, 17.0]	10.4	0.94	±6.47 (0.0)
Speed (cm/s)	115.9 (39.3)	118.0 (40.0)	−2.2 (10.5)	5.8 (9.0)	[−22.7, 18.4]	10.7	0.97	±5.90 (0.0)
SW (cm)	9.4 (10.3)	10.4 (10.8)	−1.0 (5.8)	2.6 (5.3)	[−12.3, 10.3]	5.9	0.88	±0.52 (1.0)
MTC (cm)	1.7 (0.4)	1.9 (0.8)	−0.2 (0.8)	0.6 (0.6)	[−1.9, 1.5]	0.9	0.20	±0.10 (0.3)

Shown are mean values (standard deviation), limits of agreement, RMSE, correlation, and equivalence interval (p -value). [†] All correlations were based on Spearman and have $p < 0.01$. [‡] Equivalence tests were based on Wilcoxon signed-rank test. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

the reference, already denoting satisfactory results and a good trade-off between speed and performance.

To avoid overfitting and a reduced generalization of the model, the training was stopped when the validation loss could not improve any further. The best validation loss was achieved before reaching the maximum number of epochs (i.e., before reaching 350 epochs), which confirmed the adequacy of this number. A dropout regularization was also introduced to reduce overfitting during training. The performance of the model during training (Figure 4.8) indicated that both measures were effective in preventing overfitting, resulting in learning curves (training and validation losses) with a similar pattern.

We proposed a novel representation for heel and toe trajectories to ensure that sequences were independent and did not rely on previous sequence information nor calibration procedures. The new formulation also represented heel and toe trajectories independently from each other; therefore, avoiding the need to input additional information, such as the relationship between them (e.g., distance and/or angle).

Previous studies from the areas of odometry [264, 265] and pedestrian dead reckoning [267] approached the concern of obtaining inter-segment independence by proposing a network output that estimates a change in the state rather than the actual state of the system. The results were reported in terms of positioning errors of the reconstructed trajectories (e.g., [264] reports an RMSE of 48.2 cm, evaluated in excerpts with 20 s of a single person walking; [267] reports an average Euclidean distance of 1.47 m, evaluated in paths with an average length of 90.3 m). Considering that reconstructed trajectories were obtained from changes in the state (given an initial state of the system), estimated trajectory errors were cumulative [264, 265], and, for this reason, these approaches were not suitable to our problem concerning the stride-by-stride gait analysis.

On the other hand, as we considered a formulation that was inter-segment independent, it was not possible for us to reconstruct the full trajectory of the walking trial, as the relationship between segments was missing from our predictions. This is not, however, a concern of the gait analysis, and should not be considered a limitation of the study. Rather, as errors did not accumulate along the stride nor stride to stride (as can be observed in Figure 4.9), errors affected each stride (and

Table 4.5: Performance on the test set, excluding turns ($n = 1108$ strides).

Parameter	IMU	VICON	Rel. Error	Abs. Error	Lim. Agr.	RMSE	Corr. [†]	Equival. [‡]
Stride dur. (s)	1.14 (0.21)	1.15 (0.21)	0.00 (0.04)	0.01 (0.04)	[−0.08, 0.08]	0.04	0.99	±0.06 (0.0)
Swing dur. (s)	0.39 (0.07)	0.39 (0.05)	0.00 (0.04)	0.02 (0.04)	[−0.08, 0.08]	0.04	0.92	±0.02 (0.0)
Stance dur. (s)	0.75 (0.16)	0.76 (0.16)	−0.01 (0.02)	0.02 (0.02)	[−0.05, 0.04]	0.02	0.99	±0.04 (0.0)
Cad. (st/min)	107.9 (17.5)	107.7 (17.4)	0.2 (2.8)	1.3 (2.5)	[−5.3, 5.6]	2.8	0.99	±5.39 (0.0)
SL (cm)	136.1 (23.3)	138.8 (24.6)	−2.6 (5.4)	3.8 (4.7)	[−13.3, 8.0]	6.1	0.98	±6.94 (0.0)
Speed (cm/s)	124.8 (38.2)	127.1 (39.3)	−2.2 (7.0)	3.9 (6.2)	[−15.9, 11.4]	7.3	0.99	±6.35 (0.0)
SW (cm)	2.7 (2.2)	3.5 (2.1)	−0.8 (1.8)	1.3 (1.5)	[−4.3, 2.7]	2.0	0.76	±0.17 (1.0)
MTC (cm)	1.6 (0.3)	1.8 (0.7)	−0.2 (0.8)	0.5 (0.6)	[−1.7, 1.3]	0.8	0.08	±0.09 (0.2)

Shown are mean values (standard deviation), limits of agreement, RMSE, correlation, and equivalence interval (p -value). [†] All correlations were based on Spearman and have $p < 0.05$. [‡] Equivalence tests were based on Wilcoxon signed-rank test. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

respective gait parameters) independently.

Our formulation was designed to ensure an end-to-end approach capable of predicting heel and toe trajectories from inertial sensor data, and, from the trajectories, extract clinically relevant gait parameters. However, the proposed formulation did not allow us to estimate turning angles from the trajectories, which constitutes a limitation of our approach. Although straight walking tests are more commonly employed to assess older adults' gait performance [65], in specific patient populations, e.g., in patients with Parkinson's Disease, turning deficits are very pronounced and should be part of the gait assessment [284]. To approach these specific patient populations, turning angles (and other specific biomechanical features from turning [284]) could be estimated separately, for instance, using another neural network, or, alternatively, integrated into the current neural network using a multi-task approach [285]. The estimation of the turning angles was considered out of the scope of the current research.

4.4.2 Deep Learning-Based Gait Analysis Performance

A good agreement between IMU-derived and reference gait parameters was achieved in our study. Temporal parameters, as well as the stride length and speed, were evaluated with high or very high correlations, and demonstrated equivalence with the reference method (Tables 4.4 and 4.5).

The swing width and MTC presented the lowest agreement with reference parameters (as evidenced by correlation coefficients and equivalence tests). Both parameters presented a low average value, and the swing width, in particular, was further lowered when turns were excluded. In both metrics, small errors had a higher impact on performance due to the narrow range of the parameters. According to [286], the use of inertial sensors to estimate MTC remains a challenging task, due to this effect.

The performance of the estimated gait parameters was, in general, improved when turns were excluded from the analysis, which implied that evaluating turns constituted a more challenging task for the network. Turns may present a more complex pattern—stride trajectories, accelerations, and angular rates differ from straight walking—due to feet adaptations while turning [287], and were, as such, more difficult to evaluate. A similar observation was also determined in the

Table 4.6: Random orientation simulation results ($n = 2139$ strides).

Parameter	Original Orientation	Random Orientation	RMSD	Correlation [†]	Equivalence [‡]
Stride dur. (s)	1.17 (0.27)	1.17 (0.27)	0.03	$r_s = 0.99$	± 0.06 (0.0)
Swing dur. (s)	0.40 (0.07)	0.40 (0.06)	0.03	$r_s = 0.92$	± 0.02 (0.0)
Stance dur. (s)	0.77 (0.23)	0.77 (0.23)	0.02	$r_s = 0.99$	± 0.04 (0.0)
Cad. (st/min)	106.4 (18.5)	106.4 (18.4)	2.5	$r_s = 0.99$	± 5.32 (0.0)
SL (cm)	126.1 (27.4)	126.8 (26.3)	5.7	$r_s = 0.98$	± 6.34 (0.0)
Speed (cm/s)	114.3 (40.0)	114.7 (38.5)	6.5	$r_s = 0.99$	± 5.73 (0.0)
SW (cm)	9.3 (10.3)	9.5 (10.2)	2.3	$r_s = 0.94$	± 0.47 (0.0)
MTC (cm)	1.7 (0.4)	1.6 (0.4)	0.3	$r_s = 0.78$	± 0.08 (0.0)

Shown are mean values (standard deviation), RMSD, Correlation—with Spearman or Pearson—, and equivalence interval (p -value). [†] All correlations have $p < 0.01$. [‡] All equivalence tests are based on Wilcoxon signed-rank test. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

previous Chapter 3; while this was true for almost all parameters, MTC seemed to have been less influenced by this effect, as errors remained in the same order. The average of the metric lowered when turns were excluded, but the differences were very small, which may imply little effect of turning on MTC in our results. Previous studies found significant differences in foot clearance while walking straight and turning [131,288], which contradicts our results. The detection of temporal parameters (i.e., gait events) seemed to be robust to the variability in strides due to turning. Possibly, optimizing and training a deep neural network that would only include straight walking strides could optimize the results for straight, steady walking; however, by including less variability in the data, also the generalization of the method could possibly drop. Results showed that the proposed method may be adequate to evaluate straight walking tests, with important applications in geriatrics and health-related areas.

Supported by our experiments reported in Table 4.6, the performance of our method seemed to not depend on the precise alignment of the sensors on the feet. The proposed data augmentation scheme allowed the model to capture the relationships between inertial sensor data and stride trajectories regardless of the orientation of the sensor on the foot. Due to the intrinsic errors from the predictions, the resulting RMSD values in Table 4.6 were not equal to zero, although all metrics were considered practically equivalent. This is an important requirement for solutions designed to operate outside the lab, in real settings, where the careful alignment of the sensor on the foot may not be attained for practical reasons. This feature may potentially simplify the data acquisition process, while increasing trust in the context of the clinical gait analysis.

4.4.3 Comparison with the Conventional Gait Analysis Approach

A simple conventional gait analysis approach—resorting to orientation estimation and double integration—was implemented in this study for a comparison with the proposed deep learning-based gait analysis solution. The conventional approach resorts to the double integration of linear acceleration—expressed in a fixed frame of reference—obtained from raw acceleration after subtracting the gravity component. The orientation of the sensor relative to the vertical—obtained

Table 4.7: Comparison of the performance achieved with the deep learning approach and the conventional approach.

Parameter	Including Turns		Excluding Turns	
	Deep Learning Approach	Conventional Approach	Deep Learning Approach	Conventional Approach
Trajectories (mm)	8.7 (14.7) (heel) 8.7 (13.2) (toe)	9.8 (12.7) (sensor)	7.4 (16.3) (heel) 7.5 (14.8) (toe)	8.6 (13.0) (sensor)
Stride dur. (s)	0.02 (0.04)	0.02 (0.05)	0.01 (0.04)	0.02 (0.05)
Swing dur. (s)	0.02 (0.03)	0.03 (0.04)	0.02 (0.04)	0.02 (0.05)
Stance dur. (s)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
Cad. (st/min)	1.8 (4.4)	1.9 (4.8)	1.3 (2.5)	1.4 (2.9)
SL (cm)	5.8 (8.6)	10.1 (10.9)	3.8 (4.7)	8.1 (8.1)
Speed (cm/s)	5.8 (9.0)	9.4 (11.2)	3.9 (6.2)	7.6 (9.6)
SW (cm)	2.6 (5.3)	3.5 (6.2)	1.3 (1.5)	1.9 (4.2)
MTC (cm)	0.6 (0.6)	1.5 (1.0)	0.5 (0.6)	1.4 (0.8)

Shown are mean absolute errors (standard deviation), reported for the samples in the test set. SL: Stride Length, SW: Swing width, MTC: Minimum toe clearance.

after integrating angular rate—allows the estimation of the gravity component during movement. As both methods depend mostly on integration (double integration of linear acceleration to obtain sensor trajectories and integration of angular rate to obtain sensor orientation), errors were cumulative along the stride [131, 132, 142, 143, 241]. Even though linear dedrifted was applied to mitigate some of the integration errors [131, 242], the conventional approach revealed a worse performance than the deep learning approach, which was evidenced by the results presented in Table 4.7.

As sensor trajectories were determined with a higher error—using the conventional approach—than the heel and toe trajectories—obtained using the deep learning approach—, the resulting gait parameters were also obtained with a higher error. Spatial gait parameters, in particular, were affected by the errors in estimated foot trajectories, resulting in higher errors when the conventional approach was applied. Similarly to the deep learning-based approach, the performance of the metrics improved when turns were excluded from the analysis. As turns presented a more complex pattern, it is very likely that trajectory estimation errors were higher during the turns due to the cumulative behavior of the errors in the conventional approach.

While the proposed deep learning approach could estimate 3D heel and toe trajectories, the conventional approach estimated 3D sensor trajectories. To obtain MTC, trigonometric relationships (assuming a rigid shoe model) were employed by the conventional approach, which allowed the estimation of the vertical component of the toe trajectories [133, 249]. The resulting values accumulated errors from the processes of estimating the sensor trajectory and estimating the vertical toe trajectory. The proposed deep learning model was able to predict 3D toe trajectories directly from inertial sensor data, resulting in lower errors for the MTC estimation.

4.4.4 Framing our Method within the State-of-the-Art

To detect gait events, in our study, we used the method described in Chapter 3, proposed to ensure robustness to differences in sensor orientation on the feet. Our study used the same dataset described in Chapter 3, but, while in Chapter 3 we estimated FC events from vertical acceleration—obtained after estimating sensor orientation—, in this Chapter we calculated vertical acceleration directly from predicted toe trajectories. Although the source of the signal to extract FC events differed from Chapter 3, the performance on gait event detection was very similar, which demonstrated the robustness of our method in predicting toe trajectories.

The resulting temporal metrics—stride, swing, and stance duration—had average absolute errors on the order of the 0.02 s (i.e., two samples at 100 Hz), with standard deviations ranging from 0.02 s to 0.04 s. These results were in line with the state-of-the-art [142, 143]. Using a deep learning approach to directly regress gait parameters from inertial sensor data, [263] reports lower precision on the estimation of temporal parameters (± 0.07 s for stride, ± 0.05 s for swing, and ± 0.07 for stance duration). Although we also relied on the outcomes of a neural network to detect FC events, the results achieved by our method demonstrated more consistent performance.

The performance on the estimation of spatial gait parameters was also according to past works in the field, although, compared to some works, lower precision was achieved in some of the parameters. Using the same dataset, and although reporting results for a different set of subjects, [268] reports relative errors of -3.5 ± 9.7 cm for the stride length and -3.1 ± 9.2 cm/s for gait speed (i.e., less accuracy but better precision) using an orientation-invariant, double integration-based, approach. Better accuracy and precision (relative errors of 1.5 ± 6.8 cm and 1.4 ± 5.6 cm/s) were also obtained in [131] with a double-integration approach, using a dataset with young and elderly volunteers performing U-turns and eight-turns.

With a deep convolutional neural network, [144] reports a relative error of 0.01 ± 5.37 cm for stride length. In [263], using a similar approach, stride length was estimated with a relative error of -0.15 ± 6.09 cm. Although these results seemed more promising than the results we presented in this work, we need to consider that the dataset used in both studies referred to a group of geriatric inpatients walking straight. In our results, when we excluded turns from the analysis, the precision improved, resulting in -2.6 ± 5.4 cm for the stride length and -2.2 ± 7.0 cm/s for speed, which were already in line with previous works. Using a double integration approach, [142] reports relative errors of -0.3 ± 8.4 cm (stride length) in a group of geriatric inpatients walking straight. In [143], mean relative errors of -5.4 ± 3.1 cm and -3.4 ± 3.9 cm/s (i.e., better precision) were obtained for stride length and gait speed in a group of young adults walking straight.

MTC, the parameter with the lowest performance in this study, achieved results on the level of previous works. In [133, 249], a method based on double integration and a trigonometric relationship was applied to determine heel and toe trajectories from IMU data acquired at the dorsum of the feet. The method proposed by [133] required, additionally, the shoe size as an input. These studies reported a relative error of 1.3 ± 0.9 cm [133] and 1.7 ± 0.7 cm [249] for MTC. Our results were, thus, in line with the state-of-the-art, although—as previously discussed—errors remained

very high compared to the low values of the parameter.

Most of these studies from the literature depend on the precise alignment of the sensors on the body, e.g., [131, 142, 143]. Moreover, by being conducted in laboratory settings, results reported by most of the studies were not capturing the imperfections of a poor sensor-foot alignment and, as such, may not represent the actual results that would be achieved in a more realistic setting. Deep regression models, as proposed by [144, 263], also did not capture possible errors in sensor orientation, as the training set is limited to a single orientation. As discussed previously, our method relaxed this assumption by integrating adequate data augmentation techniques, which resulted in a gait analysis solution that was, in principle, robust to changes in sensor orientation. Nevertheless, additional tests may be required to confirm the robustness of our method in practical conditions.

As we could verify from the above discussion, deep learning models do not necessarily lead to better results compared to traditional approaches resorting to double integration (e.g., see [144] versus [143]). However, we should mention that most of the studies resorting to double integration developed their algorithms and reported their results using the same dataset [131, 132, 142, 143, 249], and, for this reason, they were not assessing the generalization performance of the method and may have been too optimistic. Contrarily, studies employing deep learning approaches are very careful with data splitting, so that part of the data is never seen by the model during training and can be used to assess its generalization performance [144, 263]. Although double-integration and zero-velocity update methods rely on geometric and physical reasoning—and, for that reason, they are not so much dependent on the data—, this should be referred to as a possible limitation of these studies, whose results must be interpreted with caution.

More generally, results from the different studies should be compared and interpreted carefully, considering the different properties of the datasets employed. Different protocols (e.g., walking paths, walking speeds, sensors characteristics, etc.), subjects (e.g., young or older adults), and reference systems (e.g., optical motion capture systems or pressure sensitive walkways) may have an impact on the characteristics of the datasets that limit their comparability. Results are highly heterogeneous among the different studies; thus, a fair and robust comparison of performance is not always possible.

Compared to deep regression models [144, 263] (proposed to extract gait metrics directly from inertial sensor data), our approach seemed to provide better precision for temporal metrics, and equivalent results for stride length. To regress gait parameters directly from inertial sensor data, deep regression models need to implicitly learn how to map foot trajectories from inertial sensor data. In our approach, this relation was addressed explicitly, which may constitute an advantage. On the other hand, the MSE loss used for training only captured a summary of the prediction performance, not focusing on the most important instants, or events, used to extract gait parameters later on. This may explain why our approach did not display a clear advantage concerning gait parameters estimation. As an additional advantage, our approach allowed not only the estimation of gait parameters, but also the evaluation of heel and toe trajectories on a stride-by-stride basis, e.g., resorting to the visualization and analysis of lateral profiles (as in [143, 240]). Additionally, heel

and toe trajectories enabled the extraction of additional gait parameters (e.g., heel clearance metrics) that, unless considered during the training of deep regression models, could not be extracted using these methods.

4.4.5 Limitations

Due to their black-box nature, deep learning models are often difficult to explain and, for that reason, they usually lack trust in how well they will generalize [289]. Dissimilar to more traditional gait analysis approaches found in the literature—that rely on geometric and physical reasoning—data-driven deep neural networks are implicitly dependent on the data used for training, which should, ideally, capture as much variability as possible, to allow a good generalization of the model [290].

This study assessed the performance of the deep learning model on a considerably high number of strides (after data augmentation); however, the number of participants included in training and test sets was relatively low. To better assess the generalization of the model, a cross-validation setup could have been applied, but, due to training time constraints, we opted for a single training-test split. Additional participants would need to be included to assess the generalization performance of the model in a more representative set. Additionally, only young adults were included in the study. Although the system was designed envisioning the assessment of gait performance in older adults, this group was not included in this preliminary study, which constitutes a limitation.

As for the reference system, in this study, we used an optical motion capture system. Although it is currently considered the gold standard for a gait analysis, possible errors can also be introduced by the reference. Errors due to the wrong labeling of markers or inaccuracies on gap-filling operations (which rely on manual operation) are common [291, 292], and may lead to wrongly defined stride trajectories. The fact that many strides are included in the training set may mitigate the effects of training with inaccurate trajectories, yet, these effects should not be neglected.

4.5 Conclusion

This chapter proposed a novel approach for foot stride trajectory estimation in gait analysis. A deep recurrent neural network was proposed and optimized to allow the estimation of heel and toe trajectories on a stride-by-stride basis. To support this formulation, stride trajectories were represented in a new coordinate system that ensured inter-stride independence. Gait parameters were extracted from predicted trajectories and evaluated on a dataset comprising foot-worn inertial sensor data and reference walking data acquired from a group of young adults.

We demonstrated that bidirectional LSTM networks can be applied to the context of gait analysis, in a sequence-to-sequence setup, to obtain heel and toe trajectories, using as an input the sequences from raw inertial sensor data. Our results showed that clinical gait parameters could be extracted from predicted stride trajectories, achieving an overall good agreement with the reference system. The proposed solution was also robust to imperfect sensor-foot alignment conditions, which may support its application in real scenarios.

Future work should investigate the performance of the method in additional datasets (with different persons) for a better evaluation of the generalization performance. It should also address specific groups, e.g., older adults with mobility limitations and/or cognitive impairment, for which the validity and robustness of the method should be assessed. The estimation of turning angles should be explored in future work to allow for a more detailed analysis of walking in specific patient populations. Additional gait parameters, e.g., heel clearance, may also be included. Future work should finally focus on demonstrating the clinical value of the proposed gait analysis solution.

Although there is certainly room for improving the proposed method, results achieved at the current stage of work are already encouraging. The prediction of foot trajectories from wearable solutions is highly relevant to assist clinical decision-making in gait-related disorders, offering an interesting alternative for the assessment of gait in clinical settings or ambulatory conditions.

Chapter 5

Stepping Performance Measures for Multidirectional Steps Evaluation

This chapter is based on the following publications:

V. Guimarães, I. Sousa and M. V. Correia, “Detection and classification of multidirectional steps for motor-cognitive training in older adults using shoe-mounted inertial sensors,” *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 6926-6929, doi: 10.1109/EMBC.2019.8856851.

V. Guimarães, I. Sousa, and M. V. Correia, “Detection of Foot Motions for Interaction With Exergames Using Shoe-Mounted Inertial Sensors,” *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 5, pp. 895–904, Oct. 2023.

Interactive games have the potential to mitigate or prevent gait impairments and cognitive decline in older adults. Although inertial sensors have been successfully applied to exergaming in the past, the problem of detecting foot motions to interact with stepping exergames is still largely understudied.

In this chapter, we describe the development of a novel real-time step detection and direction classification approach to be used in the evaluation of multidirectional steps and interaction while playing motor-cognitive games. Two shoe-mounted inertial sensors were used to capture foot motions, which were treated interchangeably after the application of a novel foot sagittal reflection method. A single multi-class classifier was developed to distinguish front, back, side, and center steps originating from any of these positions. Since some of these steps exhibit similar displacement patterns, the previous step position was also considered as an input to the classifier. The method was tested in a group of young and older adults achieving an accuracy of 93.1%. Performance remained consistent throughout the acquisition time due to the introduction of a novel calibration approach designed to handle sensor orientation drift over time. This study provided the first insights into the potential of inertial sensors to detect foot motions required to interact with stepping exergames. Experimental results support their application in a real scenario.

5.1 Introduction

Rapid and well-directed steps are required to perform activities of daily living. In older adults, accidental falls during everyday activities are commonly attributed to an impaired stepping ability [293]. Studies show that both volitional and reactive stepping interventions can prevent falls in older adults, leading to a reduction of around 50% in the number of falls [294], which is crucial to prevent care dependency and reduce the costs of health and long-term care [8].

Stepping interventions are usually provided by interactive video games: the so-called stepping exergames [295, 296]. They use floor platforms with stepping targets designed to detect steps directions: the player stands at the center of the platform and performs steps (without displacing the body) by reaching pressure-sensitive areas located at the front, back, right, and left [295, 297, 298]. In addition to imposing restrictions on movements (e.g., concerning stepping amplitude), these platforms have some space requirements that limit their widespread adoption. In contrast, due to their compact size, portability, and low energy requirements, inertial sensors are well suited to measure human movement outside the lab [119, 120]. Although inertial sensors have been successfully applied to exergaming in the past [299], very few studies approached the topic of detecting multidirectional steps for interaction with stepping exergames using inertial sensors [300].

Instrumented shoes offer a very privileged position to evaluate foot motions, and have been primarily used for pedestrian tracking and gait analysis applications [143, 301]. In pedestrian dead-reckoning applications—where the tracking of walking trajectories is performed by estimating relative movements—the estimation of steps direction can be performed based on walking models [247] or using the Principal Component Analysis of accelerations [302, 303]. These approaches are usually needed because the orientation of the sensor (e.g., the smartphone) may change relative to the body [302, 303]. When the sensor is placed on the foot, and assuming that sensor orientation relative to the foot is fixed and will not change over time, it is possible to estimate the displacements by integrating acceleration. Due to sensor noise, techniques such as the Zero Velocity Update need to be employed to limit integration errors [143, 301]. The assumption of zero velocity when the foot is in contact with the ground serves as an initial condition for the step-wise integration of acceleration [143, 301]. To estimate sensor orientation relative to a fixed coordinate system, methods such as Kalman filters or Complementary Filters are typically employed, which combine information from accelerometers and gyroscopes [119, 301]. Threshold-based techniques—based on short-term statistics of accelerations, angular rates, or the combination of both—are used to detect steps and the moments when the foot is in contact with the ground [255, 304].

Although these techniques allow step detection and estimation of trajectories and displacement, the classification of stepping directions (i.e., front, side, or back) for interaction with exergames has yet to be explored. A previous work has tackled this problem in the past [300], inspired by previous works on pedestrian dead-reckoning and gait analysis. In [300], authors employed a machine learning approach that allowed classification of each step direction using features extracted in the initial stage of the step, achieving accuracies above 97%. A different classifier was employed to detect each of the considered directions. Features such as mean and

variance of acceleration and angular rates, signal magnitude area, and position change, were considered in the classification process. They detected steps to the front, side, and back, but did not consider the detection of steps to the center [300], which should happen each time the person returns the foot to the center after performing a step. Moreover, they did not consider the possible scenario of stepping directly from one position to another without returning to the center (e.g., stepping directly from the side to the front) or stepping on the spot, for instance, to step again in the same direction without returning the foot to the center. Yet, all these directions may occur while stepping and are required for natural interaction with stepping exergames [295, 297, 298]. Considering that inertial sensors can only measure relative quantities and that some steps present similar displacements, the inclusion of additional directions/displacements would bring increased complexity to the classification problem.

To address these constraints and allow more natural interaction with stepping exergames, we developed a new method for evaluating and classifying multidirectional steps. The main goal of this work was thus to develop a robust method to detect and classify steps occurring in all possible directions—including stepping on spot or moving directly from one position to another without returning to the center—relying on the analysis of foot-worn inertial sensor data. A single multi-class machine learning classifier was used to classify the direction of the steps being performed by either the left or the right foot. Foot motions of both feet were treated interchangeably after the application of a novel procedure for reflecting data across the foot sagittal plane. Since only data from gyroscopes and accelerometers were used, and to better handle orientation estimation drifts due to the integration of angular rate, we proposed a novel approach to periodically recalibrate the system. To better handle similarities between steps, the final position of the previous step was included as input to the classifier. The developments presented in this work considered the requirements needed for future application in stepping exergames, namely, the ability to detect all possible steps directions, and process each data sample as it is received to allow real-time interaction experience. The method was developed and tested using young and older adults' data collected in a laboratory setting.

This chapter is structured as follows: in Section 5.2 we provide a comprehensive description of the methods, including the data acquisition process, the signal processing and feature extraction operations, and finally, the classification experiments; in Section 5.3 we present the results; the discussion of the results, limitations, and future work are presented in Section 5.4; conclusions are drawn in Section 5.5.

5.2 Materials and Methods

5.2.1 Dataset Acquisition

We acquired data from young and older adults, comprising steps in multiple directions. Participants were positioned at the center of the room, facing a fixed direction (to simulate interaction with exergames), and performed the steps without changing their body position (like illustrated

in Figure 5.1a). No specific instructions were given concerning the characteristics of the movements, except that each step should be performed in an intended direction (front, side, or back) by contacting the foot with the ground to mark that specific direction.

The data were collected in two distinct stages. In the first stage, participants were older adults, and they were instructed to perform steps in multiple directions following a randomized sequence of presentation. It included i) stepping to the front, to the back, or to the side and returning the feet to the center, or ii) stepping directly to the front, back, or side without returning to the center. Additional data were collected in the second stage, where the order of the steps was fixed to cover all possible combinations (displacements) in the same proportions. Due to the Covid-19 pandemic, in this stage we could only acquire data from young adults.

Inertial sensors were placed near the instep of the shoes, with the x-axis pointing to the ankles, the z-axis pointing to the top, and the y-axis pointing to the right side of the body (as illustrated in Figure 5.1b). To reduce motion artifacts, sensors were tightly secured on the shoes using a Velcro strap. The same wearable device described in Chapters 3 and 4, equipped with a 32-bit Arm Cortex M4F processor (Nordic nRF52), a tri-axial gyroscope, and a tri-axial accelerometer (Bosch® BMI160), was used in this study. Sensors were wirelessly connected to the computer using Bluetooth® Low Energy. The connection was performed via browser using the Web Bluetooth API [305]. For that purpose, a simple web application was developed which allowed connection to the selected devices, initiation and finalization of a recording, and subsequent storage of the collected data in a text file. Acceleration and angular rate were collected at a sampling rate of 100 Hz, which is commonly used in gait analysis [143, 240, 241]. Before making steps, participants were asked to maintain their feet still, in contact with the ground, for at least 3 seconds (calibration period). Possible foot displacements are illustrated in Figure 5.1c, which also included stepping on spot (e.g., with the foot at the front, mark again this position with a step) and stepping directly from the back to the front or vice-versa.

Video recordings were acquired simultaneously and were used to manually annotate steps (time and direction, by visual inspection). The camera was mounted on a tripod and positioned either at the front or at the back of the participant to capture all his/her step movements. Inertial sensor data and video recordings were manually synchronized so that signal peaks would be aligned with their respective foot contacts.

Seniors were recruited from the city of Porto, through Fraunhofer's Colaborar network (<http://colaborar.fraunhofer.pt/>, accessed on 3 June 2022). They lived independently, were older than 65 years old and had no visible gait problems. Young adults were collaborators from Fraunhofer Portugal AICOS. Participants were excluded if they had any mobility or neurological disorder preventing them from stepping or understanding the risks and benefits of the study. The study received ethical approval from the Ethical Committee of the University of Porto (reference: 81-CEUP-2019), and all participants provided their informed consent.

Overall, 42 participants (18 young adults and 24 older adults) contributed to the acquisition of step data. Older adults (13 female, 11 male) averaged 71.5 ± 4.5 years old and contributed to the acquisition of 4485 steps, whereas young adults (10 female, 8 male) averaged 30.0 ± 5.6 years old

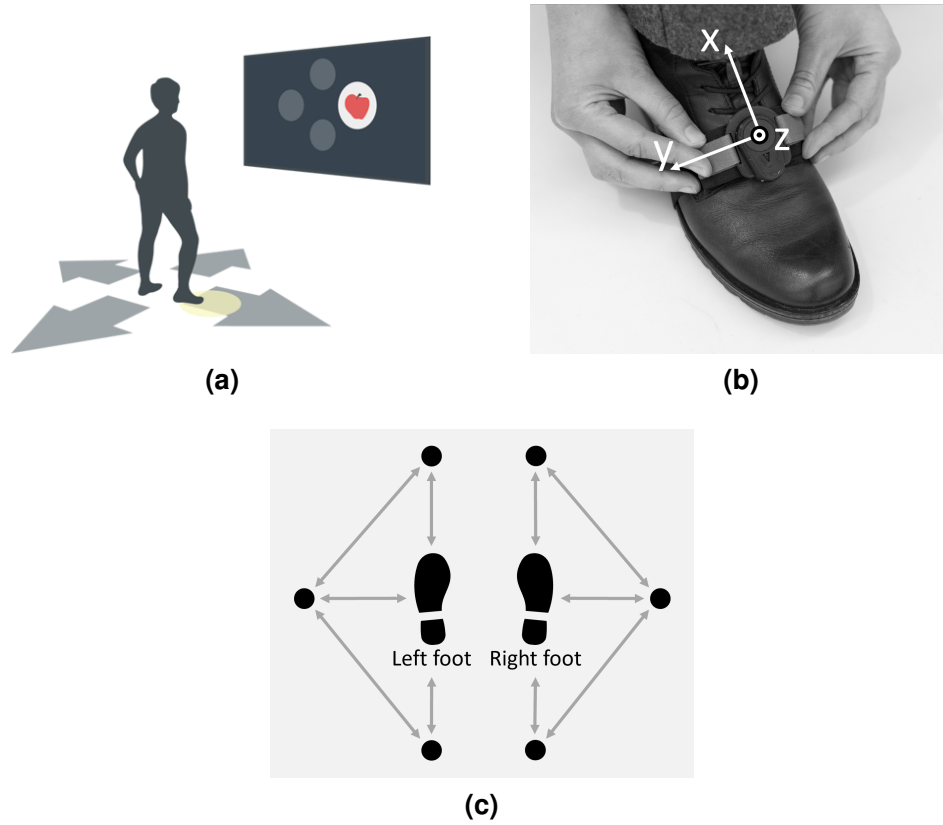


Figure 5.1: Shoe-mounted inertial sensors and steps directions for interaction with games. (a) Proposed interaction. (b) Sensor position and coordinate frame. (c) Possible displacements of the feet while executing multidirectional steps.

and contributed to the acquisition of 2094 steps.

5.2.2 Sensor Data Processing

Collected sensor data were processed to obtain isolated steps, each characterized by a set of features to be used in the supervised learning task. Sensor data processing included the stages of estimating sensor orientation, obtaining calibration data, segmenting steps, transforming data, and, finally, calculating a set of relevant features. Each of these stages is sequentially described in this section.

5.2.2.1 Sensor Orientation Estimation

A global reference frame was defined to allow uniform data representation irrespective of sensor orientation. The orientation of the sensor relative to the global frame was represented by the quaternion $q(t)$ and was obtained for all time instants t resorting to a quaternion-based derivation of the Passive Complementary Filter [306, 307]. This filter used acceleration data to estimate sensor inclination and angular rate to estimate relative changes in sensor orientation. The global

frame was thus defined by a vertical axis (gravity), and two arbitrary perpendicular axes defined in the horizontal plane.

5.2.2.2 Calibration

The calibration period was used to estimate the baseline orientation of the sensor—the orientation of the sensor when the subject remained still with both feet apart. During this period, acceleration magnitude and quaternions were collected and averaged resulting in the definition of static acceleration magnitude (a_{static}) and a normalized static quaternion ($\mathbf{q}_{static}$). The quaternion $\mathbf{q}_{static}$ represented baseline sensor orientation (i.e., the orientation of the feet toward his/her facing direction, which, in a game context would represent the orientation of the feet when the subject is facing the exergame display). a_{static} represented the measured value of the acceleration of gravity. The gravity vector, in global (or earth) coordinates, was thus determined by $\vec{\mathbf{g}}_{cal}^e = [0, 0, a_{static}]$. The calibration gravity vector was then expressed in sensor coordinates using (5.1):

$$\vec{\mathbf{g}}_{cal}^s = \mathbf{q}_{static}^{-1} \otimes \mathbf{p}(\vec{\mathbf{g}}_{cal}^e) \otimes \mathbf{q}_{static} \quad (5.1)$$

where $\vec{\mathbf{g}}_{cal}^s$ represents the calibration gravity vector in sensor coordinates, the operator $\otimes$ represents quaternion multiplication, $\mathbf{p}(\cdot)$ is the quaternion representation of a vector, and $\mathbf{q}_{static}^{-1}$ is the conjugate of $\mathbf{q}_{static}$.

The calibration period was detected resorting to the analysis of the standard deviation of angular rate magnitude, computed in sliding windows of 1-second length. Calibration time finished when the standard deviation of the angular rate magnitude was maintained below 0.2 rad/s (empirically defined) for at least 2.5 seconds.

5.2.2.3 Step Segmentation

The step segmentation procedure was defined in view to obtaining isolated steps—from foot-off (when the foot leaves the ground) to foot-contact (when the foot hits the ground)—in nearly real-time. To avoid delays in step detection, no pre-processing steps, like low-pass filtering, were applied [308]. The segmentation of the steps was performed by resorting to a threshold-based state machine, as illustrated in Figure 5.2. The segmentation process was inspired by the threshold-based techniques typically employed in pedestrian tracking and gait analysis applications [255, 304].

The state machine received acceleration and angular rate data streams and processed each sample as it was received. The detection of a step was performed based on the impacts measured when the foot hits the ground, i.e., when the slope of the acceleration magnitude went below a pre-defined threshold due to the sudden deceleration caused by the impact on the ground. Before being able to detect the impacts, the state machine needed to detect movement start (i.e., the foot leaving the ground), which was performed based on the values of acceleration and angular rate magnitudes. The movement was considered to start when the acceleration magnitude and/or angular rate magnitude deviated from the resting state, i.e., the acceleration magnitude deviated

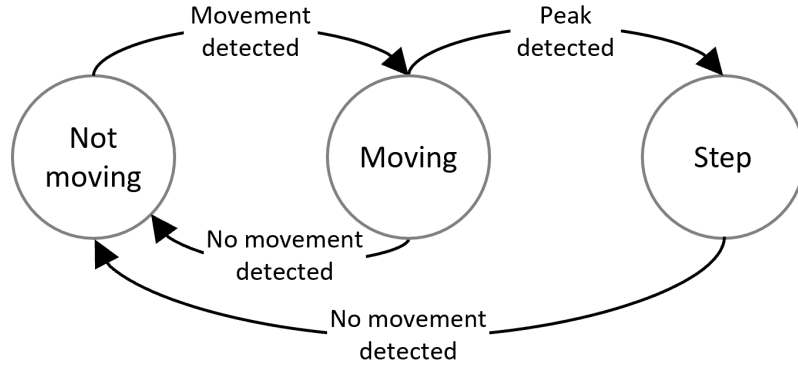


Figure 5.2: Step segmentation state diagram.

from the value of a_{static} and/or the angular rate magnitude deviated from zero, both with a pre-defined tolerance value, set experimentally. Data collected from the start of the movement until an impact was detected were considered a stepping movement. Collected values—raw acceleration values, angular rates, and quaternions—were used as input for data transformation.

5.2.2.4 Data Transformation

Quaternions $\mathbf{q}(t)$ collected within a step were used to separate the components of gravity, $\vec{\mathbf{g}}^s(t)$, and linear acceleration, $\vec{\mathbf{a}}_{linear}^s(t)$, from raw acceleration values, $\vec{\mathbf{a}}_{raw}(t)$, using the following equations (5.2) and (5.3):

$$\vec{\mathbf{g}}^s(t) = \mathbf{q}^{-1}(t) \otimes \mathbf{p}(\vec{\mathbf{g}}_{cal}^e) \otimes \mathbf{q}(t) \quad (5.2)$$

$$\vec{\mathbf{a}}_{linear}^s = \vec{\mathbf{a}}_{raw}(t) - \vec{\mathbf{g}}^s(t) \quad (5.3)$$

The calibration gravity vector $\vec{\mathbf{g}}_{cal}^e$ was then subtracted from $\vec{\mathbf{g}}^s(t)$ to eliminate the effects of sensor tilt caused by different foot and shoe styles. After this subtraction, $\vec{\mathbf{g}}^s(t)$ represented sensor tilt (relative to the vertical direction) as if the feet were flat and, before moving, parallel to the ground.

Linear acceleration was integrated two times to obtain linear displacement. Angular rate was integrated one time to obtain angular displacement. To ensure that all displacements were expressed in an appropriate coordinate frame—ensuring comparability between participants—we expressed linear acceleration and angular rate in a new coordinate frame before integrating. Instead of expressing the vectors using the current sensor coordinate system, we used a horizontalized version of the baseline sensor orientation defined by $\mathbf{q}_{static}$. This new coordinate system was used instead of the global coordinate system because users could be facing any direction in the global system. The new coordinate system was thus defined by the vertical and two horizontal axes matching the facing direction of the user as defined by $\mathbf{q}_{static}$. To obtain these vectors we used successive quaternions transformations, as depicted in equations (5.4) and (5.5):

$$\vec{v}^e(t) = \mathbf{q}(t) \otimes \mathbf{p}(\vec{v}_{in}^s(t)) \otimes \mathbf{q}^{-1}(t) \quad (5.4)$$

$$\vec{v}^{out}(t) = \mathbf{q}_{static}^{-1} \otimes \mathbf{p}([v_x^e(t), v_y^e(t), 0]) \otimes \mathbf{q}_{static} \quad (5.5)$$

where $\vec{v}_{in}^s(t)$ represents an input vector (either linear acceleration or angular rate, expressed in sensor coordinates), $\vec{v}^e(t)$ represents the vector in the global frame, and $\vec{v}^{out}(t)$ represents the output vector expressed in baseline sensor coordinates as defined by $\mathbf{q}_{static}$ after removing the vertical component of $\vec{v}^e(t)$, $v_z^e(t)$. This new coordinate system eliminates the effects of having participants facing different directions in the global coordinate system, ensuring a common representation.

To perform the transformations in equations (5.4) and (5.5), we used $\mathbf{q}_{static}$, the orientation of the foot during calibration. This representation was considered more intuitive and consistent as it allowed relative displacements to be expressed in a common way in different acquisitions and participants. We also introduced a periodic re-calculation of $\mathbf{q}_{static}$, which was updated every time the sensor was evaluated as static (not moving) and the angle between calibration gravity ($\vec{g}_{cal}^s$) and current gravity ($\vec{g}^s(t)$) was less than 10 degrees (empirically defined). This value was considered low enough to capture when the foot was in full contact with the ground, resembling its inclination during calibration.

Drift correction methods in pedestrian tracking applications usually need to rely on external data sources (e.g., the magnetometer) [309]. Unlike pedestrian tracking applications, in this application we were not interested in tracking the heading of the person, but, contrarily, to periodically reset the coordinate system so that it matches the facing direction of the user. The calibration process proposed in this work ensured that the accumulation of drift (on the horizontal plane) due to the integration of angular rate within sensor fusion would be mitigated (no further correction could be applied due to the inexistence of magnetic field data). Further, if the person starts doing steps facing another direction, the system can adapt.

Integrations were performed using the Trapezoidal Rule and only happened within the step segment to avoid a large accumulation of errors due to integration drift [143, 301].

5.2.2.5 Left Foot Reflection

Right and left feet were treated independently up until this point. To avoid having two different classifiers (i.e., one for each foot), and also to increase the amount of data available for training, left foot data were manipulated so that the representation of its movement would be equivalent to the movements on the right foot. Left foot data (represented in sensor coordinates) were reflected across the foot sagittal plane, x-z, as illustrated in Figure 5.3, using Equation 5.6.

$$\vec{v}_{out} = \mathbf{q}_r \otimes \mathbf{p}(\vec{v}_{in}) \otimes \mathbf{q}_r \quad (5.6)$$

where $\mathbf{q}_r$ denotes the reflection quaternion, built with an angle of rotation of 0 rad and the vector $[0, 1, 0]$, normal to the reflection plane. After reflection, angular measurement vectors (i.e. angular

velocity and rotation angles) are inverted, so that they match the opposite side.

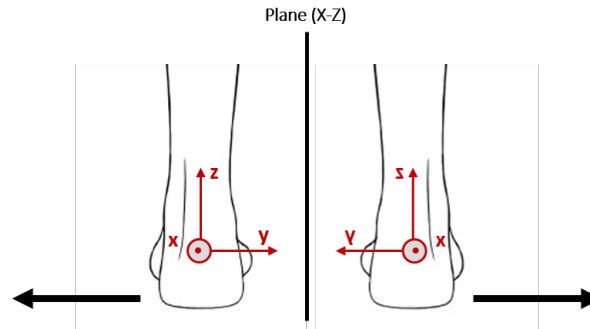


Figure 5.3: Illustration of the reflection in plane x-z.

5.2.2.6 Feature Calculation

We calculated features in two equally sized, consecutive, sub-regions of the step, which allowed the capturing of signal shape characteristics without resorting to complex template matching mechanisms that could prevent real-time operation [310]. The size of the sub-regions varied with the duration of the steps. Features were extracted after applying the data transformation operations described previously. Peak values, normalized peak values, average values, and normalized average values of the x, y, and z axes were extracted from calibrated gravity, linear acceleration, angular rate, and angular displacement within the two sub-regions. Peak values and normalized peak values of linear displacement were also extracted on the second sub-region (closer to the instant when the foot contacts with the ground). We also included the final position of the current step (vector resulting from the double integration, original and normalized) and the final position of the previous step (only horizontal components). A total of 110 features were extracted from each step.

5.2.3 Step Classification

Features extracted from the steps were used to fit a machine learning model for step classification. The corresponding annotation of each step was determined considering a time tolerance of 0.25s. Each step was labeled with one of the possible directions: “Front”, “Side”, “Back”, or “Center”. The dataset was then split into training and test partitions, using a 25:75 ratio and ensuring that unique users were included in each partition. The test set included approximately 25% of the steps, selected randomly from 11 users (6 young adults and 5 older adults); the remaining 75% were included in the training set, selected from the remaining 31 users (12 young adults and 19 older adults). This splitting scheme was chosen to ensure sufficient representation of unique users (young and older adults) in each data partition. The resulting number of steps in each partition is shown in Table 5.1.

Python 2.7 and scikit-learn 0.20.2 were used to train and evaluate machine learning models. Hyperparameter tuning was performed in the training set using grid search cross-validation (5

Table 5.1: Number of steps in training and test sets.

Partition	Front	Side	Back	Centre	Total
Training set					
Older adults	744	729	718	1323	3514
Young adults	343	343	332	331	1349
Total	1087	1072	1050	1654	4863
Test set					
Older adults	194	205	189	383	971
Young adults	183	190	188	184	745
Total	377	395	377	567	1716

folds). Several classifiers were tested, namely, a Logistic Regression model with multiple L1 regularization strengths, a Decision Tree classifier with multiple splitting criteria, a Support Vector Machine with multiple kernels and regularization strengths, and a Gaussian Naïve Bayes. These models were chosen due to their simplicity, potentially compatible with real-time classification. After finding the best set of hyperparameters, models were retrained using the whole training data. Models were then used to classify the test set and evaluate the corresponding classification performance.

5.2.4 Evaluation

The performance of the step segmentation was evaluated by the number of steps that could be detected, calculated as a percentage of the total number of annotated steps. A step was considered as detected if the instant of its impact did not exceed 0.25 s from the annotated time.

Step classification performance was evaluated using accuracy, precision, recall, and F1 score, defined in Equations 2.2, 2.5, 2.3, and 2.6 from Chapter 2, calculated for each class. We reported the average of the metrics weighted by the number of true instances for each class. These metrics were evaluated on the test set, which was never seen by the model during training. The accuracy of the classification was also reported for the validation sets within cross-validation (reported as an average $\pm$ standard deviation).

Confusion matrices were reported for the test set and for each foot side separately. The differences in the distributions of the classification accuracy of the left and right steps were assessed using the (Pearson) chi-square test (χ^2) from SciPy 1.7.1, with p-values lower than 0.05 indicating statistical significance.

The accuracy of the classification was also analyzed with regard to the annotated displacements (e.g., stepping to the front coming from the center, stepping to the front coming from the side, etc.), and reported separately for young and older adults' data. The corresponding confusion matrix was reported using the test set data.

Finally, we reported classification performance after deactivating the mechanism of the periodic re-calculation of $\mathbf{q}_{static}$, to evaluate its impact on classification performance. We assessed misclassification errors over time considering the classifications obtained in all samples of the test set, with and without the periodic re-calculation of $\mathbf{q}_{static}$. We grouped timestamps in intervals

of 20 seconds and counted the number of errors that occurred in each time bin. Misclassification errors were reported as a percentage over time.

5.3 Results

5.3.1 Step Segmentation

The proposed step segmentation approach was able to detect 92.6% of the annotated steps (i.e., 6579 steps). Only 7.4% of the annotated steps (528 steps) were not detected. Most of the steps missed were steps to the center (Table 5.2).

Table 5.2: Number of steps missed.

Metric	Front	Side	Back	Center	Total
Steps missed	53	75	67	333	528
Percentage	3.5%	4.9%	4.5%	13.0%	7.4%

An exemplary step segmentation result, including the corresponding states of the state machine, is shown in Figure 5.4. The two peaks corresponding to foot-off and foot-strike in each step [300] can be observed in Figure 5.4.

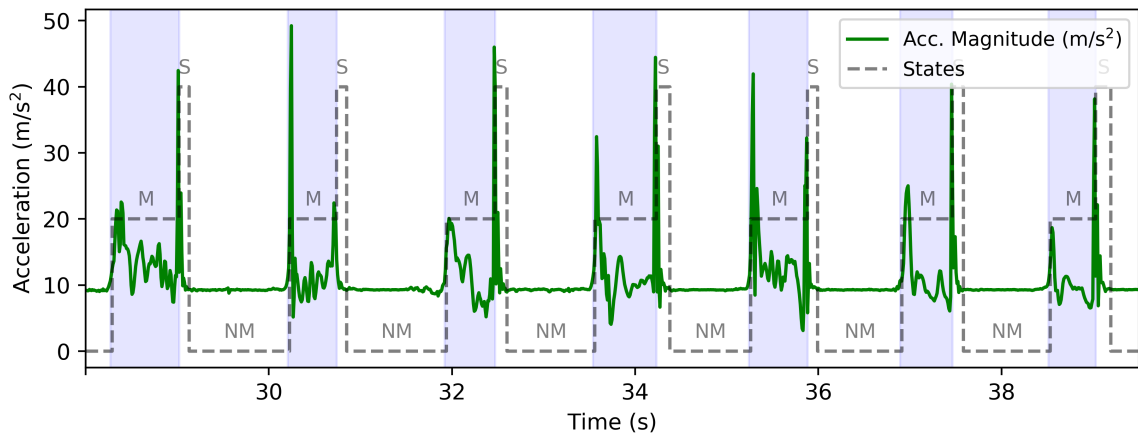


Figure 5.4: Exemplary step segmentation result (States: M-Moving, NM-Not Moving, and S-Step).

5.3.2 Step Classification

Table 5.3 summarizes classification performance obtained with Naïve Bayes, Decision Tree, Support Vector Machine, and Logistic Regression models. The best performance was achieved with the Logistic Regression model, with classification accuracies of $92.3 \pm 2.8\%$ in validation sets (within cross-validation) and 93.1% in the test set. Moreover, the Logistic Regression model achieved a precision of 93.1% and a recall and F1-score of 93.0% in the test set. In older adults' data, the classification accuracy was 95.7%. Young adults had an accuracy of 89.2%.

Table 5.3: Classification performance.

Classifier	Accuracy (%)	
	Validation sets	Test set
Naïve Bayes	79.9 ± 3.6	77.3
Decision Tree	86.0 ± 3.3	87.2
Support Vector Machine	91.5 ± 3.1	91.8
Logistic Regression	92.3 ± 2.8	93.1

In Figure 5.5 we present the confusion matrices obtained with the Logistic Regression model for all steps and for the right and left steps separately. The lowest overall performance was obtained in the classification of front steps (accuracy of 91.2%), with 7.7% of the steps being misclassified as steps to the center. Additionally, 3.7% of the backward steps were misclassified as center steps and 7.2% of the steps to the center were misclassified as forward or backward. 4.6% of the side steps were misclassified as backward. The best overall performance was obtained on the classification of backward steps (accuracy of 95.8%). No statistically significant differences were found in the distributions of the classification accuracies of left and right footsteps ($\chi^2(3) = 0.297, p = 0.961$).

		All Steps				Right Foot				Left Foot			
		Predicted Values				Predicted Values				Predicted Values			
		F	S	B	C	F	S	B	C	F	S	B	C
True Values	F	344	4	0	29	174	2	0	12	170	2	0	17
	S	6	370	18	1	4	180	14	0	2	190	4	1
	B	1	1	361	14	0	1	181	2	1	0	180	12
	C	24	5	17	521	16	5	11	259	8	0	6	262
Acc (%)		91.2	93.7	95.8	91.9	92.6	90.9	98.4	89.0	89.9	96.4	93.3	94.9

Figure 5.5: Confusion matrices for the classification of steps using the Logistic Regression model (F-Front, S-Side, B-Back, C-Center).

The analysis of accuracy values with regards to the annotated displacements, shown in Figure 5.6, revealed that most of the misclassification errors obtained with the Logistic Regression model occurred between steps potentially sharing similar displacement patterns as was the case of steps to the front coming from the front (misclassified as center), steps to the front coming from the back (misclassified as center steps), steps to the side coming from the front (misclassified as backward), or center steps coming from the back (misclassified as forward). The best performances were achieved on the classification of steps coming or returning to the center, which were also the displacements with the highest representation in the dataset due to older adults' data. The lowest performance was registered in the classification of center steps coming from the center (63.6%).

Displacements were evenly distributed in young adults' data.

		Predicted Values								# Steps		Acc. (%)
		F		S		B		C				
True values	F-F	10	33	0	1	0	0	1	10	11	44	78.2
	S-F	13	45	0	0	0	0	3	3	16	48	90.6
	B-F	16	38	0	0	0	0	4	7	20	45	83.1
	C-F	144	45	2	1	0	0	1	0	147	46	97.9
	F-S	0	1	19	34	0	12	0	0	19	47	80.3
	S-S	1	1	10	42	2	3	1	0	14	46	86.7
	B-S	2	1	17	47	0	0	0	0	19	48	95.5
	C-S	0	0	153	48	0	1	0	0	153	49	99.5
	F-B	0	0	0	0	17	46	2	0	19	46	96.9
	S-B	0	0	0	0	16	47	2	0	18	47	96.9
	B-B	0	1	1	0	7	42	2	4	10	47	86.0
	C-B	0	0	0	0	139	47	3	1	142	48	97.9
	F-C	0	0	0	0	2	6	114	41	116	47	95.1
	S-C	0	0	0	0	0	2	137	43	137	45	98.9
	B-C	11	7	0	2	0	0	119	39	130	48	88.8
	C-C	0	6	0	3	0	7	0	28	0	44	63.6

Older adults

Young adults

Figure 5.6: Confusion matrix and accuracy values obtained with the Logistic Regression model for the detection of steps with regards to the annotated displacements, presented for young and older adults' data (F-Front, S-Side, B-Back, C-Center, where e.g., "S-F" means "stepping to the front coming from the side").

Figure 5.7 shows misclassification errors over time reported for all samples in the test set using the Logistic Regression model. The maximum recording time in our test set was around 3 minutes, although most samples lasted less—reflected as a decreasing number of total steps over time. The percentage of misclassified steps had some fluctuations, but no trend towards an increase over time could be observed. When the periodic calibration mechanism was deactivated, the percentage of steps wrongly classified began to increase over time. Classification accuracy decreased to 91.2%.

5.4 Discussion

This study used shoe-mounted inertial sensors to measure and classify the movements of the feet. It provided the first insights into step detection performance and informed its future application in the context of exergaming. This work proposed to detect all possible step motions, including main directions (front, side, or back), stepping to the center, stepping on the spot, or moving the feet directly from one position to the next without returning to the center. For that, we used a single multi-class machine learning model. Three novel mechanisms were proposed and introduced in the system: i) a novel foot sagittal reflection method has been introduced to allow the analysis of both feet without any distinction; ii) the last step position was considered as an input to the classifier to better discriminate similar step displacements; and iii) a method was proposed to

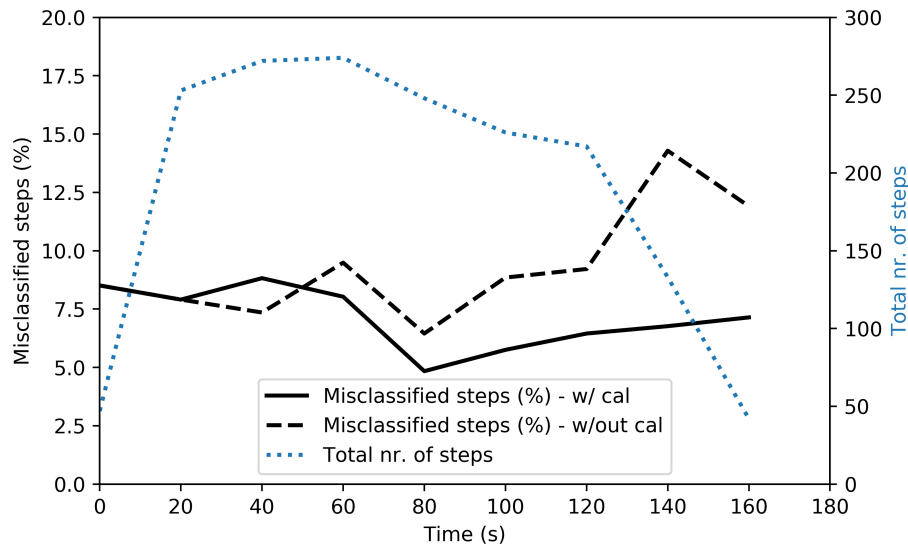


Figure 5.7: Misclassified steps over time, including performance after deactivating the periodic calibration mechanism (dashed line). Values are reported for all samples in the test set, using the Logistic Regression model.

periodically recalibrate sensor orientation estimation. The evaluation of the method was performed in a laboratory context, with data from young and older adults, achieving a detection rate of 92.6%, and classification accuracy of 93.1%. The best classification results were obtained with a Logistic Regression model.

5.4.1 Step Detection and Classification Performance

The classification performance of left and right footsteps did not differ significantly, with both sides presenting similar accuracies. A single classifier could handle the classification of both body sides due to the data reflection process applied to the left foot data. This process allowed not only a reduction in the number of classes but also the fitting of a single multi-class classifier using an additional amount of data (left and right footsteps).

To better distinguish steps, we have evaluated not only the displacements but also sensor inclination, angular rate, and angular displacements. We expressed the data in sensor baseline orientation coordinates so that the system would adapt to different sensor orientations, irrespective of the body-facing direction. We noticed that participants performed their steps resorting to multiple strategies, for instance, placing the whole foot on the ground, or using only the tiptoes (no specific instruction was given concerning the characteristics of the steps). Backward steps, for instance, were usually performed using the tiptoes, which made them a bit more distinct (the highest accuracy of 95.8% was achieved on backward steps).

To capture differences in movement patterns, we extracted the features from two sub-regions of the signal. To help distinguish steps with similar displacements, we included the last detected position (x and y coordinates) as a feature for the classification. Moreover, we introduced a mechanism for the periodic re-calibration of baseline sensor orientation, to better deal with sensor orientation

drift over time. This mechanism should also allow adapting to changes in body-facing direction (not tested). Although still some misclassification errors could be observed, the performance of the classifier remained high when the additional step directions were included. The calibration mechanism allowed classification performance to remain stable throughout acquisition time.

The lowest performance was achieved on the detection of steps to the center coming from the center (accuracy of 63.6%). Although center steps have been included in the classification task, they are usually not needed to interact with exergames [295, 297, 298]. Most of the times, though, the person needs to return the feet to the center, and, if an impact is detected, the step needs to be properly classified. In our dataset, most of the steps to the center were performed with low impact and could not be captured by the step segmentation method, resulting in higher rates of missed detection (Table 5.2). Lowering detection thresholds would allow capturing more steps, however, this could lead to the detection of false events due to residual movements of the feet. In a game context, the feedback provided by the game should allow users to adjust the impact of their steps so that they are detected only when there is an actual intention to mark a certain position.

Most step displacements had accuracies above 75%, with over 85% of the displacements presenting accuracies above 80% (Figure 5.6). The main step directions with origin at the center, i.e., stepping to the front, to the back, or to the side coming from the center—particularly suitable for the interaction with exergames [295, 297, 298]—had classification accuracies above 97%. Despite some misclassifications, this performance is comparable to a previous study and might be sufficient to interact with exergames [300]. Moreover, misclassifications can be easily perceived as such while playing exergames, which should allow participants to adapt their movements to ensure their accurate detection. The performance obtained on the test set was similar to that obtained on the validation sets within cross-validation, being thus indicative of the generalization performance of the method.

Overall, the classification performance was higher in older adults than in young adults, which should be explained by the higher representation of the best-classified displacements (i.e., stepping directions with origin at the center). Also, due to the higher number of steps provided by older adults in the training set, the resulting classifier should be better adjusted to the detection of older adults' stepping patterns. A post-analysis revealed that older adults performed steps with lower length (25.2 ± 6.9 cm vs 41.6 ± 17.4 cm), shorter duration (0.45 ± 0.09 s vs 0.52 ± 0.10 s), and lower speed (57.99 ± 14.91 cm/s vs 80.63 ± 28.71 cm/s) when compared to young adults. These differences in stepping speed are consistent with the expected declines in gait speed while aging [311]. The differences between groups were statistically significant for the three variables according to an Independent Samples T-test ($p < 0.001$). Considering that measurement errors in inertial sensor data are affected by movement speed [312, 313], the distinct characteristics of the steps may also justify the differences in classification performance between groups. Nevertheless, the solution should be applied to the context of exergaming in older adults, so it is our best interest to maximize performance in this group.

5.4.2 Comparison with Other Studies

As in [300], we used signal processing techniques typical of other domains like gait analysis and pedestrian dead reckoning: sensor fusion to obtain sensor orientation [119, 301], threshold-based step and zero velocity detection to segment steps [255, 304], and sensor data integration to obtain linear and angular displacements [143, 301].

Although the results obtained in this study seem worse than those obtained in [300] (i.e., an accuracy of 93.1% versus 97.0%), it is worth noting that in this study we did not limit the analysis to the main step directions—from the center to the right, left, front, or back—as happened in [300]. These step directions present distinct displacements, the reason for which they can be more easily recognized. In our study, accuracies above 97% were achieved in the classification of these directions. Zhou et al. [300] neglected the movement of bringing the foot back to the center after a step or stepping directly from one position to the next without returning to the center.

In [300] multiple classifiers were trained to discriminate each step direction, but authors did not explain how classifications were combined to provide a single classification result—overall accuracy was also not reported. In contrast, we used a single multi-class classifier, capable of identifying both left and right footsteps. To ensure real-time applicability, [300] classified the steps using only data from the beginning of each step (fixed size). However, this short period may not always represent the intended direction of movement. In contrast, we segmented and performed feature extraction and classification on the whole stepping movement (from foot-off to foot-contact), so that the whole movement and intended step direction are captured by the system.

5.4.3 Limitations

The proposed method could detect the main directions of movement, however, in the context of exergames also the diagonals can be interesting [293]. Introducing diagonals increases the complexity of the problem as they can trigger similar displacements. To address this issue, more accurate detection of trajectories, possibly with the inclusion of additional sensors [314] and/or the continuous monitoring of the relative position between the two feet, could be a wise way forward. Since we are not monitoring the trunk-facing direction and only relative movement can be obtained from inertial sensor data (except for the sensor inclination that can be obtained from gravity), misclassifications may occur due to the wrong interpretation of displacements. For instance, if the right foot is rotated outward, a step to the right may be mistaken as a forward step because, as seen from its own coordinate system, the foot appears to be moving forward. Moreover, sensor noise and movement complexity may both negatively influence foot motion detection accuracy. If consecutive steps are performed without any short absence of movement between them, the step segmentation approach may fail to properly segment them. Although this is unlikely to happen in normal exergame use (especially for older adults [293, 295]), we need to be aware of these limitations. Again, considering the inclusion of additional sensors [314] or employing more complex template matching methods could help to address these limitations.

Although the classification task was formulated so that classes (front, side, back, and center) were balanced, each of these step directions can be performed with distinct characteristics, either due to distinct step origins or distinct stepping strategies (e.g., using the whole foot or the tip-toes), leading to an increased variability within classes. Stepping displacements in our training set presented a distribution similar to the one shown in Figure 5.6 for the test set, with a larger representation of steps coming or returning to the center as a result of the random order of steps provided by older adults. We expect these displacements to be the most frequent while interacting with exergames [295, 297, 298]. Nevertheless, the fact that displacements were unbalanced in our dataset may have hampered the learning task and negatively affect classification performance. Moreover, stepping annotations were performed manually through visual inspection, and therefore they may not always be very rigorous. As annotations were not verified, it is possible that they include some errors that affect not only the training but also the generalization performance evaluation. Steps were recorded in the lab and may not fully reproduce the characteristics of the steps while naturally interacting with exergames. Additionally, the generalization performance cannot be extrapolated to other groups such as older adults with mobility limitations and/or cognitive impairment. Further research is required to test the feasibility and/or improve the solution for these groups.

5.4.4 Future Work

To evaluate the impact of the unbalanced distributions of step displacements, additional data would be needed, either collected from the participants or synthesized [314]. Alternatively, to solve the supervised learning task with a small number of labeled data, generative adversarial networks could be explored, as suggested in [315]. Other deep learning models could be employed to avoid manual feature extraction and allow more complex modeling of the problem [314]. We could also formulate the classification task as a displacement classification task with 16 classes, so that each class would have a more unique representation.

We should also evaluate the performance of the method while playing exergames, to assess step detection and classification performance in a realistic setting. Practical experiments should be designed to evaluate step detection latency while playing exergames, which should result from a combination of factors, including Bluetooth communication latency and processing delay. To allow interaction with exergames, the system needs to identify steps directions in nearly real-time, with delays up to 100 ms being usually accepted [316]. Finally, we should evaluate the performance of the method with older adults with mobility or cognitive impairments, potential target users of stepping exergames, as their stepping characteristics may differ from healthy older adults. Step characteristics (e.g., step length, step duration, step height) measured by inertial sensors—without imposing any movement restriction—may be used to objectively evaluate stepping performance in older adults. These parameters are already extracted using our proposed approach. Moreover, the performance of the method should be investigated in other application scenarios, for instance, to monitor steps directions and walking patterns while performing daily activities (e.g., to detect corrective steps performed while walking to avoid falls). As suggested in [317], the

current solution could benefit from including additional sensor data, for instance, electromyography (EMG) data to measure muscle activity. By better understanding which muscles are activated, not only classification results could be improved but also steps characteristics could be evaluated in a more holistic manner. Multi-sensor data fusion technology, relying on machine learning and wearable technology, has achieved very promising results in several applications of human motion analysis [314]. The inclusion of additional sensors could be explored in the future.

5.5 Conclusion

The problem of detecting steps directions while interacting with exergames resorting to inertial sensors is still largely understudied. In this work, we developed and evaluated the potential of using shoe-mounted inertial sensors to detect and classify multidirectional steps, envisioning its application in the context of exergaming. This work relied on machine learning techniques to discriminate steps to the front, back, side, or center arising from any of these positions. To better distinguish steps directions with similar displacements, we used the previous step position as a feature of the classification task. We have also proposed a novel calibration method to deal with sensor orientation drift over time and a novel foot sagittal reflection method to be able to analyze right and left foot data without any distinction. We achieved a classification accuracy of 93.1% that remained consistent throughout the acquisition time. Experimental results support the use of inertial sensors as an alternative to floor platforms to estimate step direction in the context of exergames. Moreover, inertial sensors allow the evaluation of additional movement characteristics, like step duration, length, or speed, that may provide important information concerning individuals' performance. Further work is required to evaluate step detection latency and assess the performance of the method while playing stepping exergames.

Part III

TOWARD THE DETECTION OF COGNITIVE IMPAIRMENT

Chapter 6

Gameplay Metrics for Motor-Cognitive Assessment: a Correlational Study

This chapter is based on the following publications:

V. Guimarães, I. Sousa, E. D. de Bruin, J. Pais, and M. V. Correia, “Using shoe-mounted inertial sensors and stepping exergames to assess the motor-cognitive status of older adults: A correlational study,” *Digital Health*, vol. 9, p. 205520762311670, Jan. 2023, doi: 10.1177/20552076231167001.

V. Guimarães et al., “An Exergame Solution for Personalized Multicomponent Training in Older Adults,” *Applied Sciences*, vol. 11, no. 17, p. 7986, Aug. 2021, doi: 10.3390/app11177986.

Stepping exergames designed to stimulate physical and cognitive skills can provide important information concerning individuals’ performance. In this study, we investigated the potential of stepping and gameplay metrics to assess the motor-cognitive status of older adults.

Stepping and gameplay metrics were recorded in a longitudinal study involving 13 older adults with mobility limitations. Game parameters included games’ scores and reaction times. Stepping parameters included length, height, speed, and duration, measured by inertial sensors placed on the shoes while interacting with the exergames. Parameters measured on the first gameplay were correlated against standard cognitive and mobility assessments, including the Montreal Cognitive Assessment (MoCA), gait speed, and the Short Physical Performance Battery (SPPB). Based on MoCA scores, patients were then stratified into two groups: cognitively impaired and healthy controls. The differences between the two groups were visually inspected, considering their within-game progression over the training period.

Stepping and gameplay metrics had moderate-to-strong correlations with cognitive and mobility performance indicators: faster, longer, and higher steps were associated with better mobility scores; better cognitive games’ scores and reaction times, and longer and faster steps were associated with better cognitive performance. The preliminary visual analysis revealed that the group

with cognitive impairment required more time to advance to the next difficulty level, also presenting slower reaction times and stepping speeds when compared to the healthy control group.

Stepping exergames may be useful for assessing the cognitive and motor status of older adults, potentially allowing assessments to be more frequent, affordable, and enjoyable. Further research is required to confirm results in the long term using a larger and more diverse sample.

6.1 Introduction

Interactive games constitute a very promising avenue for the unobtrusive screening of cognitive ability. For being played primarily as a form of entertainment, games allow assessments to be more frequent, affordable, and enjoyable [33, 318]. Many commercial off-the-shelf games incorporate cognitive challenges such as memorizing sequences, making decisions, recognizing patterns, or analyzing complex information. Despite not being primarily designed for this purpose, Mandryk and Birk [319] support that even existing games can be used as biomarkers of cognitive decline, as they generate rich information about players' performance in multiple cognitive challenges. Other games are being purposefully developed for the screening of cognitive disorders. As an example, Tong et al. [28] proposed a serious game featuring the Go/No-Go task, designed for tablets, to measure inhibition ability in older adults in a hospital emergency department. The performance of the game correlated significantly with global cognitive performance (as assessed by MoCA scores), being also much easier to administer and fun [28]. The scores of another mobile game designed for tablets and based on augmented reality with manipulation of physical objects showed high concurrent validity with the MoCA test in non-demented older adults [320]. In another study, two serious games played with the support of custom-designed buttons revealed significantly lower reaction times and accuracy rates in older adults with cognitive impairment [321].

Games combining challenging tasks with physical activity, referred to as exergames, have been proposed to improve physical and cognitive parameters in older adults [295, 297]. Although their effects are traditionally assessed using external outcome measures after gameplay [295, 297], parameters extracted directly from the exergames may be used to monitor both the cognitive and the physical state of older adults [322, 323]. For example, parameters extracted from a balance training exergame employing a three-dimensional moveable plate presented moderate-to-high correlations with established motor and cognitive tests in older adults with mild-to-moderate dementia [322]. Similar results were obtained for an exergame training program applied to cognitively intact older adults that used a stepping platform to detect multidirectional steps [323]. Stepping exergames require the execution of volitional and inhibitory stepping movements, involving the sensory system, information processing skills (such as fast decision-making, dual tasking, and inhibitory response), and neuromuscular control, which are needed to prevent loss of balance and avoid falls in daily living [295, 324]. Litz et al. [323] showed that game scores and reaction times extracted from a stepping exergame are associated with motor-cognitive performance in older adults with limited functional status, yielding moderate-to-high correlations with cognitive tests, measures of lower extremity function and dynamic balance.

In the scope of the project VITAAL (AAL-2017-066), funded by the European Commission through the Active Assisted Living Program, an exergame was developed to stimulate physical and cognitive skills in older adults by providing a multicomponent, simultaneous, training approach [325]. VITAAL included exergames for balance and motor-cognitive training focused on attention and executive functions [325]. The interaction with the exergame was performed by resorting to multidirectional steps that were detected by inertial sensors mounted on the shoes, using the methods described in Chapter 5. Simultaneously, these inertial sensors allowed the extraction of several steps characteristics, including, their length, height, and duration, as described in Chapter 5.

This study aimed at investigating the potential value of VITAAL exergames as an unobtrusive tool for the assessment of motor-cognitive status in older adults. We investigated the role of stepping characteristics and game metrics extracted during gameplay in a correlational study that compared these metrics with standard assessments of cognition and mobility. In this respect, our hypotheses were that (i) stepping characteristics would significantly correlate with cognition and mobility performance indicators, (ii) cognitive games' metrics would correlate with cognitive performance, and (iii) within-game progression would present differences between older adults with and without cognitive impairment.

6.2 Materials and Methods

A recent work developed a stepping exergame that used inertial sensors to detect multidirectional steps (using the methods described in Chapter 5) [325]. Game data were collected in a longitudinal study and analyzed using a mixed design, comprising a cross-sectional correlational study and a longitudinal analysis. This section describes the methods used to investigate the potential of this exergame as a tool to assess the motor-cognitive status of older adults.

6.2.1 Participants and Procedures

This study provides a secondary analysis of data collected under the scope of a Randomized Controlled Pilot Trial (RCT), originally designed to assess the feasibility of the exergame as an intervention for patients with mobility impairment (ClinicalTrials.gov identifier: NCT04587895). The RCT was conducted in a physiotherapy clinic, Physio SPArtoS, located in Interlaken, Bern, Switzerland. The inclusion and exclusion criteria are presented in Table 6.1.

Participants recruited to the RCT were randomly allocated to the control group or to the intervention group. This study has only analyzed the data from the intervention group, who played the exergame three sessions per week (each session lasting about 30 minutes), during a maximum of 12 weeks (no longer than two weeks of break were allowed). All participants that played the exergame at least once were included in this study. In each session, participants played an individually calculated amount of strength, cognitive, and balance training, that remained the same over the 12-week intervention period. The exergame provided training progression by automatically adjusting the difficulty of movements, tasks, and/or speed to the participants' performance [325].

Table 6.1: Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
• Aged 60+ years old • Living independently, in a residency dwelling, or with care • Able to stand straight for at least 10 minutes without aids (to be able to play the exergame) • Visual acuity with correction sufficient to work on a TV screen • Short Physical Performance Battery (SPPB) test score below 10 as an indication of impaired mobility [326, 327]	• Mobility impairments that didn't allow to play the exergame • Severe cognitive impairment (MoCA score below the 1st percentile according to Thomann et al. [328]) • Acute or unstable chronic diseases (e.g., recent cardiac infarction, uncontrolled high blood pressure or cardiovascular disease, uncontrolled diabetes) • Orthopedic or neurological diseases that inhibited exergame training • Rapidly progressive or terminal illness • Insufficient knowledge of German • Chronic respiratory disease • Condition or therapy that weakens the immune system • Cancer • Serious obesity (Body Mass Index (BMI) > 40 kg/m²)

Each game was comprised of ten difficulty levels, and all participants started at level one. They advanced to the next difficulty level if they achieved scores above 90%, or if the average of the last three game scores was above 75%. Average scores below 50% would imply a decrease in the difficulty level of the game [325]. Before starting the intervention, all participants were assessed for demographics and medical history, cognitive tests, physical tests, and gait. Tests already administered for the purpose of selecting study participants were not administered again.

The study was approved by the Institutional Ethics Committee of ETH Zurich (registration: 2020-00578) and followed the ethical code for research with human beings as stated by the Declaration of Helsinki. Written informed consent was obtained from all participants. No compensation for participation was given.

6.2.2 Assessments

A health questionnaire was administered to obtain self-reported information about the participants, including, sociodemographic variables (age, gender, weight, height, and education), general health questions (hearing and vision problems), medical history (diagnosed diseases to confirm exclusion criteria), and physical abilities and activity (single question about fear of falling, number of falls in the last six months, and physical activity level). Participants were also asked about video game use and previous experience with exergames.

Global cognitive performance was assessed using the Montreal Cognitive Assessment (MoCA), administered by qualified physiotherapists from Physio SPArtoS. MoCA is a widely used and preferred tool for cognitive screening in primary care settings [22], making it suitable for use in this study. This test has a maximum score of 30 points and consists of several tasks organized

in seven cognitive domains: visuospatial/executive, naming, attention, language, abstraction, delayed recall, and orientation [166]. The total score was corrected for low education (≤ 12 years) by adding an extra point [166]. Both the total score and the individual domain scores were used in this study.

Functional physical performance was assessed by trained physiotherapists using the Short Physical Performance Battery (SPPB). SPPB consists of three subtests focused on balance, walking, and lower limb strength. SPPB has a maximum score of 12 points [329, 330], with scores lower than 10 indicative of mobility impairment [326, 327].

Gait was measured at participants' preferred speed, walking along a straight 20 m distance path. To objectively assess gait parameters, inertial sensors were mounted on the shoes, and the method described in Chapter 3 was employed. Gait speed was extracted as a comprehensive measure of individuals' functional capacity and mobility [84, 331].

6.2.3 Stepping Exergames

The exergame included 8 stepping-based minigames, four of which were designed to stimulate cognition, and the other four to stimulate balance (Figure 6.1). Cognitive games included tasks that challenged specific cognitive domains related to attention and executive functions. The Healthy Snacks game (Figure 6.1-a) challenged reaction time and response inhibition control, which were needed to select healthy food as fast as possible while avoiding the selection of unhealthy food. The Shopping List game (Figure 6.1-b) required short-term memory to memorize and later be able to recognize the list of products previously shown. The Pizza game (Figure 6.1-c) challenged selective attention by requiring players to ignore irrelevant distractors (pizza slices pointing in a common direction) and focus on the relevant information (the pizza slice pointing in an opposite direction). Finally, by requiring participants to switch between two cognitive processes—one to interpret the shape, and the other to interpret the color of the objects—the Shopping game (Figure 6.1-d) challenged cognitive flexibility. Balance games, i.e., Labyrinth (Figure 6.1-e), Mommy Chicken (Figure 6.1-f), Falling Books (Figure 6.1-g), and Music (Figure 6.1-h), focused more on the execution of multidirectional steps. Balance games required more frequent steps and transitions between right and left feet to properly command the main character of the game. A full description of the minigames is provided in [325].

For each step performed within the games, we recorded its length (maximum displacement in the horizontal plane), height (maximum displacement in the vertical—gravity—axis), and duration (time elapsed since the initiation of the step to foot-contact), as determined by the inertial sensors. These metrics were measured using the methods described in Chapter 5, using the same wearable inertial measurement units. Averages were then calculated for each game and level played. Step speed was calculated by dividing step length by step duration. For the cognitive games, we have also recorded step reaction times (the time from the presentation of the stimulus until the initiation of the step) and whether it has or has not led to a correct answer within the game. The average reaction time (incl. the average reaction time for the correct and for the incorrect answers) was calculated for the cognitive games and registered for each game and level played. Games scores

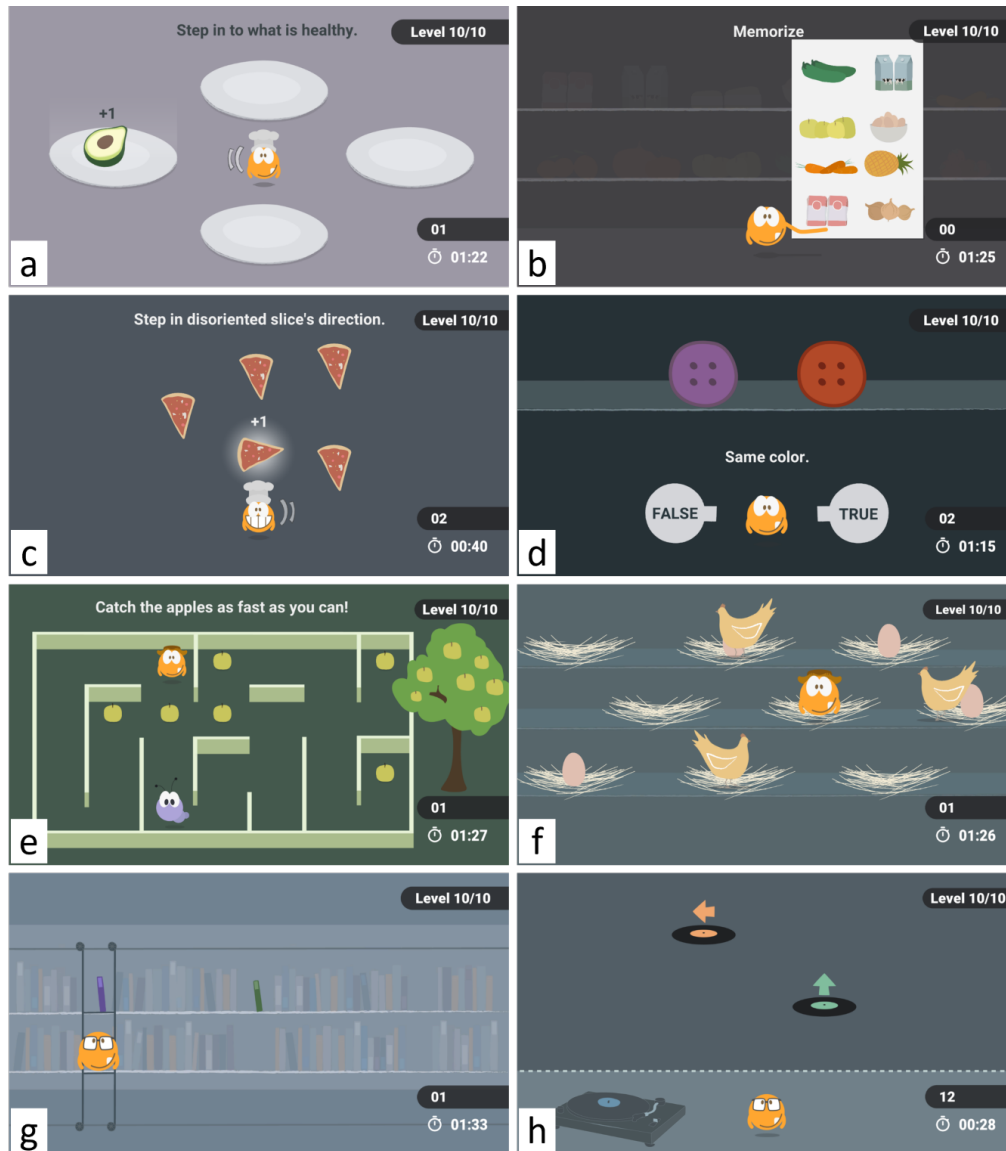


Figure 6.1: Stepping-based minigames. *Cognitive games:* (a) Healthy Snacks (cognitive inhibition), (b) Shopping List (short-term memory), (c) Pizza (selective attention), (d) Shopping (cognitive flexibility). *Balance games:* (e) Labyrinth, (f) Mommy Chicken, (g) Falling Books, and (h) Music.

(a value between 0 and 100% calculated based on the number of correct answers or the number of elements picked) [325] were also registered for the cognitive games as possible indicators of cognitive performance.

6.2.4 Statistical Analysis

The packages pingouin (v0.4.0), StatsModels (v0.13.0), SciPy (v1.7.1), and NumPy (v1.21.5) for Python 3 were used for the statistical analysis. A p-value of less than 0.05 was considered to indicate statistical significance.

Demographic and medical characteristics were summarized using either means and standard deviations, or frequencies and percentages, as appropriate. As height was missing in one participant, its value was replaced by the average.

Since demographics and medical characteristics were obtained at the beginning of the longitudinal study and considering that not all participants played the same number of repetitions of a game level (due to dropout, or due to their performance), we have only analyzed the stepping and game metrics obtained on the first time the participant played the first level of difficulty (the entry-level performance). A correlation analysis was conducted to examine the association between gameplay metrics and global physical and cognitive performances. We calculated the partial correlation between step characteristics (step height, step length, step duration, and step speed) and measures of cognitive and physical performance (MoCA total score, SPPB score, and gait speed) whilst controlling for participants' height. We have also calculated the partial correlations between cognitive games' performance metrics (scores and reaction times) and cognitive test scores (MoCA scores without correction for low education and individual domain scores). These correlations were controlled for education as both individual cognitive domains and the total MoCA score may be affected by the educational level [332]. We used Spearman's rank correlations because most metrics were not normally distributed according to the Shapiro-Wilk test. Significant correlations were interpreted based on their strength as follows: 0.0 to 0.2 (0.0 to -0.2) negligible, 0.2 to 0.4 (-0.2 to -0.4) weak, 0.4 to 0.7 (-0.4 to -0.7) moderate, 0.7 to 0.9 (-0.7 to -0.9) strong, and 0.9 to 1.0 (-0.9 to -1.0) very strong [333].

Two groups were formed based on MoCA scores, using the cut-offs proposed by Thomann et al. [168] and O'Caoimh et al. [334], as follows: individuals were considered cognitively normal if they had MoCA score ≥ 24 (healthy control group); they were considered cognitively impaired if they had MoCA score < 24 (cognitively impaired group). The differences between groups concerning their progression within the games were analyzed by visually inspecting difficulty levels' progression over the intervention period. Time was represented by the number of days since the start of the intervention, rounded to multiples of 5 days for improved data aggregation. The differences in reaction times (for both correct and incorrect answers) and step speed evolution across difficulty levels were also visually represented. Reaction times and step speeds were calculated as average values for each difficulty level, participant, and game. For the comparison between groups, we reported averages and 95% confidence intervals.

6.3 Results

A total of 14 participants were allocated to the intervention group, but only the data from 13 participants were analyzed in this study since one patient dropped out before playing any game. Three participants dropped out from the study after having completed a single training session. Two participants completed 5 and 8 training sessions before dropping out. The eight participants who finished the intervention study performed between 22 and 32 sessions (the average was 27.4 ± 3.2 sessions).

6.3.1 Descriptive Statistics

The demographic and clinical characteristics of the participants are presented in Table 6.2. The study included participants with an average age of 81.6 ± 7.3 years old, an average MoCA score of 23.5 ± 3.2 , and an average gait speed of 0.9 ± 0.2 m/s. Nearly half of the participants fell more than once in the last 6 months. The average gait speed was below 1.0 m/s, possibly denoting an increased risk of falls [203]. All participants had an SPPB score below 10 (average of 7.5 ± 1.4) indicating impaired mobility [326, 327]. Participants had very low experience with the use of video games, and, in particular, with exergames, but they were physically active. Their education spanned from 8 to 24 years (average of 12.0 ± 4.0 years).

6.3.2 Steps Characteristics

The partial correlations between steps characteristics and global physical performance and cognitive tests scores are shown in Table 6.3. Of the four step characteristics measured, only step length and step speed presented significant correlations with MoCA scores: step length had moderate positive correlations with MoCA scores in two cognitive games (Pizza and Shopping List) and strong positive correlations in two balance games (Mommy Chicken and Music); step speed presented moderate positive correlations with MoCA scores in two cognitive games (Pizza and Shopping List) and a strong positive correlation in a balance game (Mommy Chicken). Although not always significant, the correlations of step length and speed with MoCA scores were all ≥ 0.40 . Step height and step length were both significantly correlated with SPPB scores in Pizza, Shopping, Books, Labyrinth, and Music; these correlations were all positive and moderate or strong. Speed was significantly correlated with SPPB scores in the same games except Shopping. Additionally, speed had a significant correlation with SPPB scores in Healthy Snacks. The correlations between SPPB scores and step speed were all positive and strong. Of these 15 significant correlations with SPPB scores, only six were also significantly correlated with gait speed, presenting moderate or strong positive associations. Most of them were not significant, even though their correlation strengths were still at a high level (>0.40). Step duration was not significantly correlated with any of the global physical performance or cognitive tests scores.

Table 6.2: Demographic and clinical characteristics of the intervention group.

Participant Characteristics	Intervention group (n=13)
Age (years)	81.6 $\pm$ 7.3 (64-91)
Weight (Kg)	71.1 $\pm$ 14.3 (52-97)
Height (cm)	165.2 $\pm$ 10.1 (148-178)
Education (years)	12.0 $\pm$ 4.0 (8-24)
Female, n (%)	8 (61.5)
Hearing problems, n (%)	5 (38.5)
Vision problems, n (%)	4 (30.8)
Play video games, n (%)	1 (7.7)
Tried exergames before, n (%)	2 (15.4)
Fear of falling	
Never, n (%)	5 (38.5)
Sometimes, n (%)	5 (38.5)
Often, n (%)	2 (15.4)
Always, n (%)	1 (7.7)
Number of falls in the last 6 months	
Zero, n (%)	6 (46.2)
One, n (%)	1 (7.7)
More than one, n (%)	6 (46.2)
Physical activity	
>3 times/week, n (%)	9 (69.2)
1-3 times/week, n (%)	4 (30.8)
1 time/week, n (%)	0 (0.0)
Never, n (%)	0 (0.0)
MoCA	
Total score (0–30)	23.5 $\pm$ 3.2 (20-30)
Visuospatial/executive (0–5)	3.5 $\pm$ 1.5 (1-5)
Naming (0–3)	2.9 $\pm$ 0.3 (2-3)
Attention (0–6)	5.3 $\pm$ 0.8 (4-6)
Language (0–3)	2.1 $\pm$ 0.9 (1-3)
Abstraction (0–2)	0.6 $\pm$ 0.7 (0-2)
Delayed recall (0–5)	2.6 $\pm$ 1.7 (0-5)
Orientation (0–6)	5.8 $\pm$ 0.4 (5-6)
Physical tests	
SPPB Score	7.5 $\pm$ 1.4 (4-9)
Gait speed (m/s)	0.9 $\pm$ 0.2 (0.6-1.3)

Data are mean values $\pm$ standard deviation (range), except otherwise indicated.

MoCA: Montreal Cognitive Assessment; SPPB: Short Physical Performance Battery.

6.3.3 Cognitive Games' Metrics

The partial correlations between cognitive games' metrics and MoCA total and individual domain scores are shown in Table 6.4. MoCA total score only presented a significant correlation with metrics of the Pizza game, namely a moderate negative correlation with the reaction time of correct answers and a moderate positive correlation with the game score. The visuospatial/executive component of MoCA was significantly correlated with reaction times in the Healthy Snacks game (strong correlation), with reaction time in the Shopping game (moderate correlation), and with reaction time and score in the Shopping List game (moderate correlation); the correlations were negative for reaction times and positive for games scores. The abstraction component of MoCA

Table 6.3: Correlations between step characteristics and global physical and cognitive performance tests.

Game	Step characteristic	MoCA score	SPPB score	Gait speed
Healthy Snacks (n=11)	Height	0.57	0.39	0.03
	Length	0.53	0.37	0.13
	Duration	-0.02	-0.23	-0.55
	Speed	0.41	0.70*	0.56
Pizza (n=12)	Height	0.48	0.63*	0.45
	Length	0.60*	0.73*	0.61*
	Duration	-0.27	-0.15	-0.26
	Speed	0.67*	0.74**	0.57
Shopping (n=11)	Height	0.22	0.68*	0.39
	Length	0.40	0.69*	0.45
	Duration	-0.24	0.03	-0.42
	Speed	0.45	0.63	0.48
Shopping List (n=12)	Height	0.49	0.58	0.44
	Length	0.64*	0.36	0.34
	Duration	0.14	0.04	-0.07
	Speed	0.64*	0.52	0.50
Books (n=13)	Height	0.30	0.64*	0.65*
	Length	0.53	0.70*	0.45
	Duration	-0.38	-0.30	-0.46
	Speed	0.48	0.76**	0.61*
Labyrinth (n=10)	Height	0.62	0.83**	0.67*
	Length	0.65	0.73*	0.74*
	Duration	0.30	0.11	0.10
	Speed	0.66	0.81**	0.63
Mommy Chicken (n=12)	Height	0.58	0.53	0.25
	Length	0.79**	0.38	0.33
	Duration	-0.08	-0.06	-0.43
	Speed	0.82**	0.34	0.34
Music (n=10)	Height	0.40	0.77*	0.57
	Length	0.78*	0.67*	0.56
	Duration	-0.30	-0.42	-0.54
	Speed	0.58	0.83**	0.72*

Shown are the Spearman partial correlations, adjusted for height. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, bold values indicate significance. The number of participants playing each game varied due to patient dropout. MoCA: Montreal Cognitive Assessment; SPPB: Short Physical Performance Battery.

presented a strong negative correlation with reaction times and a strong positive correlation with the scores of the Pizza game. The remaining components of MoCA (i.e., naming, attention, language, delayed recall, and orientation) did not present any significant correlation with cognitive games' metrics.

6.3.4 Game Progression

Six participants were considered healthy controls (control group) and seven had cognitive impairment (CI group) according to MoCA. As some participants dropped out of the study or could not reach the most advanced difficulty levels, the number of participants decreased as the difficulty

Table 6.4: Correlations between cognitive games' metrics and MoCA scores.

Game	Game metric	MoCA score [†]	MoCA visuospatial /executive	MoCA naming	MoCA attention	MoCA language	MoCA abstraction	MoCA delayed recall	MoCA orientation
Healthy Snacks (n=11)	RT	-0.07	-0.72*	-0.50	-0.10	0.02	0.04	0.34	-0.37
	RT (correct)	0.09	-0.57	-0.50	-0.07	0.06	0.31	0.39	-0.18
	RT (incorrect)	0.04	-0.78*	-0.41	0.23	0.44	0.27	0.23	-0.36
	Score	0.12	0.55	0.53	0.13	-0.37	0.19	-0.19	0.19
Pizza (n=12)	RT	-0.48	-0.17	- [‡]	-0.44	-0.47	-0.86***	-0.12	-0.03
	RT (correct)	-0.62*	-0.20	- [‡]	-0.51	-0.47	-0.84**	-0.43	0.18
	RT (incorrect)	-0.23	-0.02	- [‡]	-0.29	-0.42	-0.39	0.11	-0.19
	Score	0.69*	0.45	- [‡]	0.57	0.59	0.79**	0.31	0.26
Shopping (n=11)	RT	-0.19	-0.38	-0.10	-0.33	-0.21	-0.41	0.18	0.05
	RT (correct)	-0.55	-0.66*	-0.51	-0.61	-0.11	-0.47	-0.05	0.05
	RT (incorrect)	0.19	-0.31	-0.32	0.20	-0.25	0.66	0.32	-0.04
	Score	0.33	0.32	0.30	0.53	0.21	0.39	0.02	-0.02
Shopping List (n=12)	RT	-0.31	-0.65*	- [‡]	-0.34	0.19	-0.47	0.09	0.23
	RT (correct)	-0.05	-0.52	- [‡]	0.01	0.33	-0.27	0.28	0.13
	RT (incorrect)	-0.14	-0.48	- [‡]	-0.13	-0.23	-0.28	0.12	-0.35
	Score	0.27	0.67*	- [‡]	0.36	-0.04	0.44	-0.23	-0.15

Shown are Spearman partial correlations, adjusted for education. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, bold values indicate significance. The number of participants playing each game varied due to patient dropout. [†]MoCA scores without correction for low education. [‡]Could not be computed because one of the variables was constant. RT: reaction time; MoCA: Montreal Cognitive Assessment.

levels progressed. The number of participants playing each game level, stratified by their cognitive status, is shown in Table 6.5.

The progression of each group within cognitive games is shown in Figure 6.2, where we can observe some differences between the two groups. The CI group required, in general, more time than the control group to advance to the next difficulty level (Figure 6.2–left). Reaction times were generally higher (i.e., slower) in the CI group (Figure 6.2–right).

As can be observed in Figure 6.3, step speed was slower in the CI group, a tendency that spanned across all difficulty levels and all balance and cognitive games.

6.4 Discussion

We investigated the potential of stepping and game performance metrics to assess the motor-cognitive status of older adults by evaluating their association with standard cognitive and physical performance metrics. This study was the first to evaluate steps characteristics like length, height, or speed using inertial sensors while interacting with exergames and investigating their role as non-invasive indicators of cognitive and physical performance in older adults. In line with our initial hypothesis, several stepping and gameplay metrics presented a significant association with cognition and mobility performance indicators, denoting the potential value of stepping-based exergames as a clinical tool for the screening of motor-cognitive status in older adults.

Table 6.5: Number of participants playing each game level, stratified by their cognitive status.

Game	Group	Difficulty levels									
		1	2	3	4	5	6	7	8	9	10
Healthy Snacks	Control	5	5	5	5	5	5	4	4	4	4
	CI	6	5	5	5	5	4	4	4	4	3
Pizza	Control	6	5	5	5	5	5	5	5	5	4
	CI	6	3	3	3	3	3	3	3	3	3
Shopping	Control	5	5	5	5	5	5	5	5	5	4
	CI	6	4	4	4	3	3	3	3	3	3
Shopping List	Control	6	5	5	5	5	5	5	4	4	4
	CI	6	4	3	3	3	3	2	0	0	0
Books	Control	6	5	5	5	4	4	4	4	4	4
	CI	7	3	3	2	2	2	1	1	0	0
Labyrinth	Control	5	5	5	5	5	4	4	4	4	4
	CI	5	4	4	3	3	3	3	3	3	2
Mommy Chicken	Control	6	4	4	2	2	1	1	1	1	1
	CI	6	3	1	1	1	1	1	1	0	0
Music	Control	5	5	5	5	5	4	4	4	4	4
	CI	5	4	4	4	4	3	3	3	3	3

CI: Cognitive impairment.

6.4.1 Association with Physical Performance

We examined the correlations between stepping metrics and standard tests of physical performance. We anticipated that each exergame and each difficulty level would elicit distinct steps characteristics, due to the different challenges they pose. For this reason, we performed the correlation analysis for each game individually, considering only the metrics obtained the first time the participant played the first level of difficulty. In the results presented by Skjæret-Maroni et al. [335], steps characteristics changed depending on the exergame played, difficulty level, and the number of times the exergame was played.

In line with our initial hypothesis, several steps characteristics presented significant correlations with physical tests scores (Table 6.3). Step height, length, and speed presented moderate or strong positive associations with SPPB scores and gait speed in several games, suggesting that the better the mobility of the person, the faster, the larger, and the higher the steps were performed while interacting with balance and cognitive exergames. Step duration was not significantly associated with physical tests scores. Faster speeds—resulting from the simultaneous contribution of increased step lengths and/or decreased steps duration—provided a better indication of physical performance than step duration. Significant correlations were more frequent on balance than cognitive exergames, denoting their potential superior value as indicators of mobility in older adults.

Previous works studied the role of volitional (or voluntary) stepping on mobility performance assessment. Both the maximal step length test and the rapid step test were shown to correlate significantly with measures of mobility performance [336], balance, and fall risk in older adults [324], which is in line with our results. Another study showed that the ability to step rapidly declines with age, being stepping performance also affected by step direction and the need to decide on the most appropriate direction [337], which is also required to play exergames.

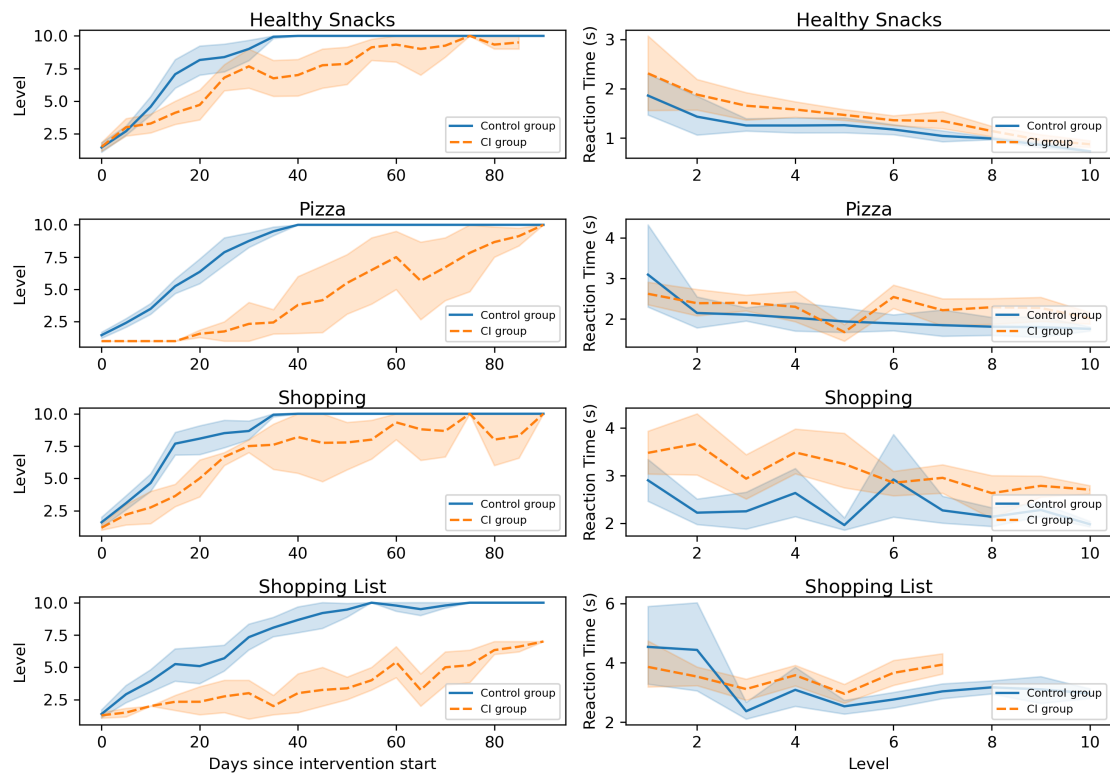


Figure 6.2: Progression within cognitive games: (left) difficulty levels over time, and (right) reaction times across difficulty levels. Error bands represent 95% confidence intervals.

Stepping parameters could be extracted directly from the inertial sensors, using the methods described in Chapter 5, avoiding the need to use independent motion capture solutions as was the case in the work presented by Skjæret-Maroni et al. [335]. In fact, most of the existing stepping exergames rely on the use of step-sensitive platforms that are not able to measure these stepping parameters [297, 338]. As shown in our study, stepping parameters, as measured by inertial sensors, can be used as indicators of mobility in older adults.

6.4.2 Association with Cognitive Performance

Although cognitive games were not purposefully designed to assess cognition, the tasks within these games were designed to provide some cognitive demand related to attention and executive functions. In line with our initial hypothesis, these games presented significant associations with cognitive performance.

One of the cognitive games (Pizza game) was moderately associated with overall MoCA scores (Table 6.4), indicating that the higher the game scores and the faster the reaction times, the better the global cognitive performance. Additionally, the Pizza game was strongly associated with the abstraction component of MoCA, being faster reaction times and better scores associated with better performance on abstract reasoning. Within MoCA, the abstraction component refers to a two-item verbal abstraction task [166]. Learning this abstract rule requires an efficient allocation

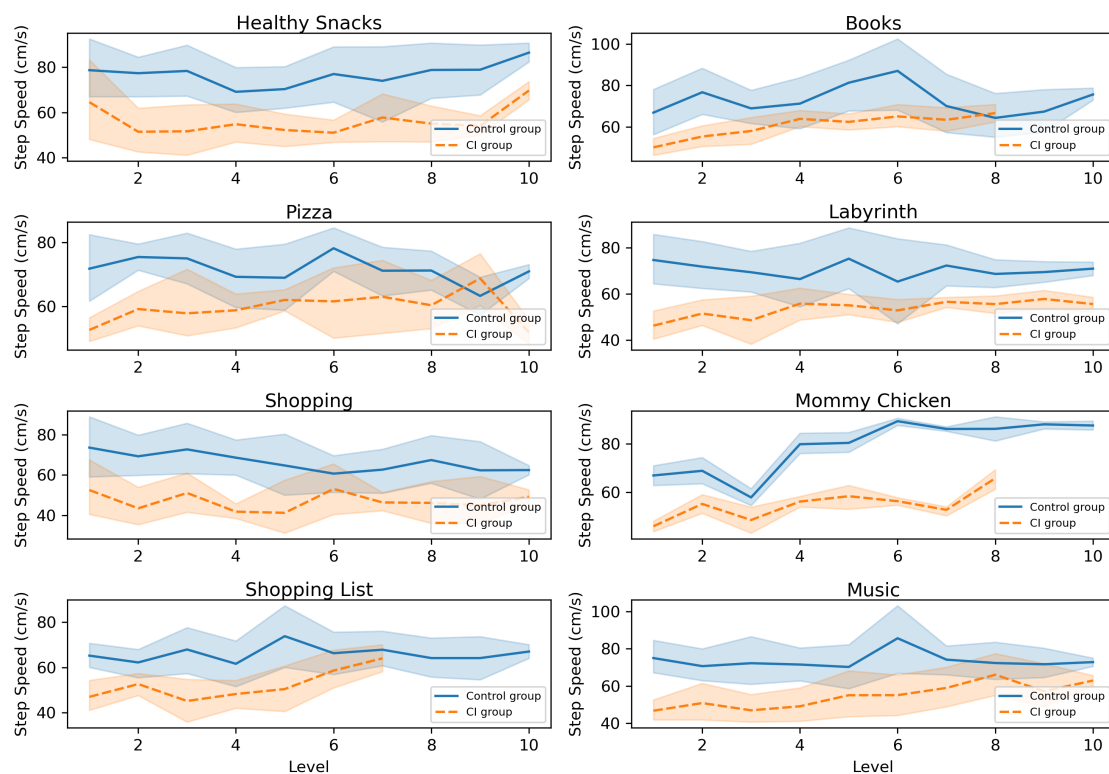


Figure 6.3: Step speed across difficulty levels presented for all balance and cognitive games. Error bands represent 95% confidence intervals.

of attention, i.e., focusing on some aspects of information while ignoring others [339]. This ability is usually described as selective attention and may be required to identify the common attributes within the verbal abstraction task [339, 340]. The Pizza game, designed to challenge selective attention, may also require these abilities to identify the pizza slice that is pointing in the wrong direction. Reflecting on abstraction, philosophers often conceive abstraction as being a form of selective attention [341].

As expected, the ability to correctly interpret visual information within the games and the ability to act accordingly depends on visuospatial and executive skills as reflected by the associations with the visuospatial/executive component of MoCA. The games Healthy Snacks, Shopping, and Shopping List were significantly associated with the visuospatial/executive component of MoCA, being better visuospatial/executive skills associated with higher scores and faster reaction times. The visuospatial/executive component of MoCA includes a modified trail-making test, the copy of the cube, and the clock drawing test [166]. In the trail-making test, the respondent is asked to draw a line connecting an alternating sequence of numbers and letters [166]. Besides visuomotor and visuperceptual skills, the trail-making test requires mental flexibility to alternate between numbers and letters [342–344]. This flexibility of task switching is also demanded by the Shopping game in the ability to switch between shape or color matching [340]. To copy the cube, spatial planning and visuomotor coordination are needed [342, 345], which makes a parallelism with the

abilities required to plan and step in an appropriate direction while interacting with the games. Coordinated visuomotor stepping control is also required to walk safely and avoid falls [346,347]. Finally, the performance of the clock drawing task demands planning skills, conceptualization, symbolic representation, and inhibitory control [342, 348, 349]. In our results, Healthy Snacks had stronger correlations with the visuospatial/executive component of MoCA than Shopping and Shopping List games, since it simultaneously required visuoperceptual skills (to rapidly identify the objects), visuomotor coordination (to step in an appropriate direction) and inhibitory response (to avoid selection of unhealthy food).

Although the Shopping List game was designed to challenge short-term memory, it did not present significant associations with the delayed recall component of MoCA. In the Shopping List game, the person had to identify whether the objects shown matched or not the list presented previously. To answer this task, recent memory, in particular, recognition was required [11]. In contrast, MoCA asked participants to recall as many items as possible from a list of words, requiring recent memory but, in particular, free recall [11], which differs from the task in the game. Moreover, while the Shopping List game challenges delayed recall through visual memory, MoCA challenges delayed recall through verbal memory, utilizing different and potentially competing cognitive resources [350,351]. According to our results, the Shopping List game mainly challenges the ability to recognize and identify the pictures in the shopping list, which is more related to the visuospatial/executive component of MoCA.

The remaining components of MoCA, i.e., naming, attention, language, and orientation were not associated with any of the cognitive games' metrics. Although individual MoCA tests are organized under the subheadings identified in Table 6.4, tests scores can be grouped using distinct structures to represent the cognitive domains [167, 352]. The correspondence between each individual test and its assumed cognitive domain is not always robust because each cognitive task requires abilities from multiple cognitive domains [167, 352, 353]. It is then not surprising that cognitive games challenge multiple cognitive skills, and, for that reason, their scores and metrics may reflect multiple cognitive constructs [320, 323]. The tasks used to test naming, attention, language, and orientation components within MoCA—i.e., a naming task, the digit span forward and backward, letter A tapping, serial-7 subtraction, sentence repetition, letter fluency, and orientation in time and place [166]—are very distinct from the tasks presented in the games. These components of MoCA were thus not related to the cognitive mechanisms underlying cognitive games.

As hypothesized, stepping characteristics also presented significant associations with cognitive performance (Table 6.3). Longer step lengths and faster speeds were associated with better global cognitive performance (MoCA scores) in several balance and cognitive games. Previous studies have also identified a relationship between mobility and cognitive performance in older adults, suggesting that gait speed could be used to predict cognitive impairment [39–41, 84]. Gait metrics determined under dual-task (i.e., walking while performing a cognitive task) could identify gait changes related to cognitive impairment that could otherwise go unnoticed [52, 53, 203]. Considering that exergames provide motor-cognitive training (i.e., they require the simultaneous

execution of stepping to solve a cognitive challenge), it is not surprising that stepping characteristics may also serve as an indicator of cognitive performance. Like gait parameters measured under single and dual-task analysis, stepping characteristics extracted while playing exergames may be used to assess performance on both cognitive and mobility functions.

Due to the low number of participants, we were not able to objectively assess differences between healthy and cognitively impaired groups as determined by MoCA scores. Nevertheless, we observed that step speed was generally lower in the group with cognitive impairment (Figure 6.3), which is consistent with previous studies on gait analysis [39–41, 52, 53, 84, 203]. This tendency spanned across all difficulty levels and all games. The group with cognitive impairment also required more time to advance to the next difficulty level, being slower to answer the tasks of the cognitive games (Figure 6.2). A previous study has examined visuospatial executive function and learning in older adults with and without cognitive impairment. The study showed that relative to healthy controls, individuals with cognitive impairment made more exploratory/learning errors, displaying slower learning curves [354], which is in line with our results. Another study concluded that game progression was slower in older participants, being strongly influenced by age and, thus, sensitive to changes in cognitive ability [355]. According to Jakobsen et al. [356], complex reaction time (with inhibition control) can be used to reflect cognitive function in healthy subjects and patients. Reaction time slowing has been considered an early sign of cognitive impairment [357]. Taken together, the results of these studies strengthen our provisional results indicating that gameplay metrics and the path of progression within the games may be able to differentiate between both groups. Yet, additional tests are needed.

6.4.3 Limitations

The present study was mostly limited by the small number of participants. The correlation analysis performed in this study should be considered of exploratory nature, as it focused on the analysis of the results of the first gameplay and did not include an objective analysis of all difficulty levels. On one side, this entry-level performance can be seen as a reflection of the first impact on gameplay and ensures a fair comparison between participants. In this regard, we need to highlight that all participants had very low previous experience with playing exergames, and, as such, they could all be considered inexperienced users. On the other side, the first impact on gameplay may not entirely reflect the real performance of the player as a simple misunderstanding of the rules or of the interaction with the game can negatively impact players' performance. Moreover, the lowest difficulty level may not challenge participants as needed to detect cognitive impairment. More advanced difficulty levels could offer a more adequate challenge for a specific patient's performance (avoiding ceiling effects) [323], and could potentially strengthen the relationship between variables. Considering that the difficulty levels automatically adapted to each person (to ensure an optimal load in line with performance) [358], the number of times the participants played a single difficulty level differed, making it difficult to compare the different levels. A more structured analysis of difficulty levels and repetitions would be needed to more thoroughly assess construct validity, test-retest reliability, and sensitivity to change, as proposed by Litz et al. [323].

Given the small sample size and the existence of study dropouts, the most complex levels had even lower sample sizes (Table 6.5) and, as such, lack of statistical power. Therefore, it was not possible for us to objectively evaluate exergame progression and its relation to the motor-cognitive status of the participants. The preliminary visual analysis revealed the potential ability of exergame progression to discriminate between groups with distinct cognitive skills (according to their MoCA scores), however, it requires further investigation.

Although MoCA is the most common and preferable tool for cognitive screening in primary care settings [22], it cannot be considered an alternative for more in-depth neuropsychological assessment, let alone its domain-specific scores [352, 353, 359]. In this study, we used domain-specific scores as presented by MoCA, however, several alternative test grouping schemes are proposed in the literature, reflecting an inconsistent correspondence between individual MoCA tests and their assumed cognitive domains [352, 353, 359]. Our results should, thus, be interpreted with caution.

Finally, our study has only included older adults with mobility limitations (as determined by their SPPB scores) and its results cannot be generalized to other populations. Despite their mobility limitations, they were still quite active and most of them have never or only sometimes experienced fear of falling. A larger and more diverse sample—including participants with more distinct characteristics, higher SPPB scores, and/or cognitive impairment as determined by neuropsychologists—would be needed to better evaluate the generalization of our results and confirm the ability of stepping and exergame metrics to assess the motor-cognitive status of older adults.

6.4.4 Future Work

Further research is required to comprehensively evaluate the new assessment approach based on stepping and gameplay metrics measured while playing stepping exergames. We should include a higher number of participants and formally evaluate their cognitive status by resorting to expert diagnosis, which should also include potential neurodegenerative conditions and a domain-specific evaluation. Moreover, older adults without mobility limitations should be included to formally assess whether gameplay and stepping metrics are able to differentiate people with and without mobility impairment. More specific measures of mobility (e.g., static and dynamic balance) should be considered as well.

The distinct difficulty levels and the individual progression of the participants should be more thoroughly analyzed to allow more frequent cognitive screening. Future research should address how to differentiate between the effects of learning on a patient's performance and his/her actual cognitive status, particularly relevant when the exergame is available to be played repeatedly at home. Besides its construct validity, we should assess its test-retest reliability and the sensitivity to change in a longitudinal study design purposefully thought to answer all the questions above.

6.5 Conclusion

Although stepping exergames were originally designed for the training of physical and cognitive functions, stepping and game performance metrics extracted while playing may also serve as indicators of the motor-cognitive status in older adults.

In this study, we used inertial sensors to extract relevant stepping characteristics like height, length, duration, and speed while interacting with exergames. Moderate-to-strong correlations were obtained between game-derived measures and physical and cognitive performance tests. Faster, longer, and higher steps were associated with better mobility scores. Longer and faster steps were associated with better cognitive scores. Cognitive game metrics (e.g., reaction times and scores) presented significant associations with cognitive performance. People with cognitive impairment seemed to exhibit slower progression within the game, and slower reaction times and step speed when compared to people without cognitive impairment.

Although the results in this study should be considered exploratory, they provide some important indications on the potential value of stepping exergames as a clinical tool for the screening of motor-cognitive status in older adults. Exergames like those tested in this work may be able to assess and monitor motor-cognitive function in older adults, potentially allowing assessments to be more frequent, more affordable, and more enjoyable.

Chapter 7

Using Gait to Assess Cognition: a Cross-Sectional Study

This chapter is based on the following publication:

V. Guimarães, I. Sousa, E. D. de Bruin, J. Pais, and M. V. Correia, “Minding your steps: a cross-sectional pilot study using foot-worn inertial sensors and dual-task gait analysis to assess the cognitive status of older adults with mobility limitations,” *BMC Geriatrics*, vol. 23, no. 1, p. 329, May 2023, doi: 10.1186/s12877-023-04042-6.

This pilot study aimed to investigate the potential of the instrumented gait analysis solution proposed in Chapters 3 and 4 to capture and differentiate gait performance in the presence of cognitive impairment based on single- and dual-task gait assessments.

We analyzed demographic and medical data, cognitive tests scores, physical tests scores, and gait metrics acquired from 29 older adults with mobility limitations. Gait metrics were extracted using the newly developed gait analysis approach and recorded in single- and dual-task conditions. Participants were stratified into two groups based on their Montreal Cognitive Assessment (MoCA) global cognitive scores. Statistical analysis was performed to assess differences between groups, discrimination ability, and association of gait metrics with cognitive performance.

The addition of the cognitive task influenced the gait performance of both groups, but the effect was higher in the group with cognitive impairment. Multiple dual-task costs, dual-task variability, and dual-task asymmetry metrics presented significant differences between groups. Also, several of these metrics provided acceptable discrimination ability and had a significant association with MoCA scores. The dual-task effect on gait speed explained the highest percentage of the variance in MoCA scores. None of the single-task gait metrics presented significant differences between groups.

Our preliminary results show that the newly developed gait analysis solution based on foot-worn inertial sensors is a pertinent tool to evaluate gait metrics affected by the cognitive status of

older adults relying on single- and dual-task gait assessments. Further evaluation with a larger and more diverse group is required to establish system feasibility and reliability in clinical practice.

7.1 Introduction

Previous studies have shown that the quality of gait under dual-task testing is affected by cognitive impairment and can distinguish MCI from preserved cognition in older adults, exposing symptoms that would otherwise be invisible [36, 51, 52]. However, some studies show contradictory results, as is the case of [202] and [208]. Most studies differ in the protocols used, including the cognitive tasks and the gait evaluation methods (incl. the instruments used to objectively measure gait). However, not all cognitive tasks may be adequate to stress cognition in older adults [53, 203] and not all gait evaluation methods may be adequate for a robust evaluation of gait metrics [58].

As discussed previously, the technology used for an objective and multidimensional assessment of gait—e.g., instrumented walkways, and optical motion capture systems—is expensive and requires dedicated spaces not being ideal for routine screening within clinical settings [58]. Additionally, they have a limited coverage area [58], which may potentially compromise the assessment of metrics like gait variability due to insufficient walking distance [53]. Wearable technology based on inertial sensors offers a low-cost alternative for the assessment of gait, allowing gait parameters to be captured both at the clinic or in free-living conditions [58]. Due to advances in data analytics, wearables are now widely proposed as practical tools to aid the diagnosis and treatment of a range of neurological disorders [61]. However, the ability of wearables and their associated algorithms to detect cognitive impairment has yet to be established [58, 64].

In Chapters 3 and 4, we proposed a new approach for the gait analysis resorting to inertial sensors placed on the shoes. In these chapters, we proposed a new data analysis approach that is independent of the orientation of the sensor on the shoes, and that allows the evaluation of an extensive set of gait metrics and foot—heel and toe—trajectories.

This pilot study aims to investigate the potential of the newly developed foot-worn inertial sensor-based gait analysis solution to capture and differentiate gait performance in older adults with and without cognitive impairment. We hypothesized that (i) people with cognitive impairment would demonstrate more significant dual-task effects on gait metrics than older adults without cognitive impairment, (ii) the effects would be visible in multiple gait parameters, (iii) dual-task gait metrics and dual-task costs would better differentiate between groups than single-task gait metrics, and (iv) multiple gait parameters would be associated with the global cognitive performance. This work constitutes the first step toward developing a screening tool for cognitive impairment that does not heavily rely on cognitive testing.

7.2 Materials and Methods

This study provides a secondary analysis of anonymized baseline data (pre-measurements) collected under the scope of a Randomized Controlled Pilot Trial. The trial was originally designed

to assess the feasibility of a newly developed exergame in patients with mobility impairment. The full details of the trial can be found on ClinicalTrials.gov (identifier: NCT04587895). In this study, we used the baseline data from the trial to investigate the potential of foot-worn inertial sensors and gait analysis to assess gait metrics that are affected by the cognitive status of the participants. We adhere to guidelines to report the results of these cross-sectional data [360].

7.2.1 Participants and Procedure

Participants were recruited through local contact persons at the physiotherapy clinic Physio SPAr-tos (Interlaken, Bern, Switzerland) and public advertisements in local newspapers in the surroundings of Interlaken (Bern, Switzerland). The study took place on Physio SPAr-tos' premises.

Participants were included if they (i) were aged 60+ years old, (ii) lived independently, in a residency dwelling, or with care, (iii) could stand straight for at least 10 minutes without aids, (iv) had visual acuity with correction sufficient to work on a TV screen, and (v) had Short Physical Performance Battery (SPPB) test score below 10 as an indication of impaired mobility [326,327].

Subjects manifesting one or more of the following criteria were excluded from the study: (i) Mobility impairments that did not allow them to play the exergame [325]; (ii) Severe cognitive impairment (MoCA scores below the 1st percentile according to [328]); (iii) Acute or unstable chronic diseases (e.g., recent cardiac infarction, uncontrolled high blood pressure or cardiovascular disease, uncontrolled diabetes); (iv) Orthopedic or neurological diseases that inhibited exergame training; (v) Rapidly progressive or terminal illness; (vi) Insufficient knowledge of German; (vii) Chronic respiratory disease; (viii) Condition or therapy that weakens the immune system; (ix) Cancer; and (x) Serious obesity (Body Mass Index (BMI) $> 40 \text{ kg/m}^2$).

Baseline data were acquired in a single visit to the study site, lasting around 1 to 1.5 hours. All eligible participants were assessed for demographics and medical history, cognitive tests, physical tests, and gait tests under single- and dual-task conditions.

The study was approved by the Institutional Ethics Committee of ETH Zurich (registration: 2020-00578) and followed the ethical code for research with human beings as stated by the Declaration of Helsinki. Written informed consent was obtained from all participants. No compensation for participation was given.

7.2.2 Demographic Data and Medical History

A health questionnaire was applied to describe general information about the participants, including age, sex, weight, height, and years of education. The questionnaire also covered general health, medical history, and physical activity, including self-reported variables such as pain, medical conditions, level of physical activity, fear of falling, past falls, and walking aids. Body mass index (BMI, in Kg/m^2) was calculated from weight and height.

7.2.3 Cognitive Tests

General cognition was assessed using the Montreal Cognitive Assessment (MoCA) [166]. MoCA is the neuropsychological test recommended for the screening of MCI [22]. The score is corrected for low education (≤ 12 years) by adding an extra point [166]. Although the initially proposed cut-off of < 26 has shown good sensitivity and specificity for detecting MCI [166], the cut-off score of < 24 resulted in better classification rates in German-speaking, Swiss, older adults [168].

Specific cognitive domains related to executive functions were assessed using (i) the Trail Making Test (TMT) Part A—for general information processing speed—and Part B—for mental flexibility [361]; (ii) the Color Word Interference Test (CWIT), to evaluate response inhibition and interference [362], and (iii) the Digit Backward Task of the Wechsler Memory Scale-Revised (WMSR), to evaluate working memory [363].

7.2.4 Physical Tests

The Short Physical Performance Battery (SPPB) was applied to evaluate functional physical performance. SPPB consists of three subtests: (i) maintaining balance with the feet in tandem, semi-tandem, and side-by-side, (ii) walking 4 meters, and (iii) rising and sitting five times on a chair without the help of the upper limbs. The test score is obtained by adding the scores of the subtests, resulting in values between 0 (worst performance) and 12 (best performance) [329, 330]. The SPPB has been used to characterize functional capacity in older adults with and without cognitive impairment [364], showing little change in relation to aspects such as culture, language, or educational level [88, 365]. An SPPB score lower than 10 is predictive of all-cause mortality [366] and is usually accepted as a threshold for functional impairment [367]. All participants in this study had SPPB scores < 10 .

Additionally, the 1-Minute Sit to Stand Test (MSTS) was employed to measure exercise capacity. In this test, the participants cross their arms over the chest and rise from a chair as often as possible in one minute. The score is the number of complete sit-to-stand cycles performed in 1 min [368].

7.2.5 Gait Assessment

Participants walked at their preferred speed (single task) and while performing a secondary cognitive task (dual task), along a straight 20m distance path. In the dual-task condition, participants had to count backward in steps of seven from a random given number between 200 and 250 while they were walking. This cognitive task has been previously recommended to stress cognition in dual-task tests [52, 53, 53, 203]. Participants had to count out loud for the test to be considered valid. No prioritization instruction was given. The order of the tasks was maintained for all the participants, starting with the single task. The wearable devices used to capture walking data were equipped with a tri-axial gyroscope and a tri-axial accelerometer (Bosch BMI160) and were developed in our lab [268]. They were connected via Bluetooth to the computer and were used to capture raw acceleration and angular rate, which were recorded in a text file. Inertial sensors were

placed on the shoes near the instep region (as described in Chapters 3 and 4). Sensor data were sampled at a frequency of 100 Hz. Participants wore their own, low-heeled, regular shoes.

The extraction of gait metrics was performed using the method described in Chapter 3. Very briefly, the method resorts to an orientation-invariant approach that comprises the steps of detecting zero velocity intervals (using the angular rate energy detector), estimating sensor orientation (using the Euston Complementary Filter), obtaining displacements (using direct and reverse integration), detecting gait events (from the low-pass filtered acceleration magnitude and vertical acceleration), and, finally, estimating spatiotemporal gait metrics [268]. Further details about the method can be found in Chapter 3. Data processing was performed offline, taking into account all the data available in recorded text files. After excluding the first and last two strides to focus the analysis on steady-state walking [44], the following metrics were extracted per stride: Stride time (s), Swing time (s), Stance time (s), Swing ratio (%), Stance ratio (%), Foot flat ratio (%), Pushing ratio (%), Loading ratio (%), Stride length (m), Gait speed (m/s), Strike angle (degrees), Liftoff angle (degrees), and Minimum toe clearance (cm). The stance ratio was calculated as a proportion of stride time. The ratios of foot flat (the time during which the foot is in full contact with the ground), pushing (the time from foot flat to toe-off), and loading (the time from heel strike to foot flat) were calculated as proportions of stance time. The liftoff angle was the angle made by the foot at toe-off, relative to the ground; the strike angle was determined at heel strike. The angles were signed and were negative for the strike. Minimum toe clearance was defined as the minimum height of the toes during the swing phase of walking, as presented in Chapter 3.

The method described in Chapter 4, resorting to a deep learning approach, was used to estimate heel and toe trajectories, from which the corresponding 3D path lengths were calculated. The method used a stacked bidirectional long short-term memory (LSTM) recurrent neural network implemented in Python 3 using conda TensorFlow 2.3 to estimate trajectories from acceleration and angular rate data. To describe the variability of the trajectories, we first resampled the strides by expressing their time in percentual terms (between 0 and 100%). Then, we calculated the mean absolute deviation (MAD) for each stride, trajectory (heel and toe), and component (forward, sideward, and vertical) according to Equation 7.1:

$$MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}_i| \quad (7.1)$$

where n is the number of samples in a stride, x_i is the value of the trajectory at sample i , and $\bar{x}_i$ is the average value of all strides at sample i .

For each gait test, we reported the average of the metrics along all the strides (including the right and left sides). We have also calculated the variability of the metrics, quantified using the coefficient of variation ($CoV[\%] = 100 \times (\text{standard deviation}/\text{mean})$). The symmetry index was used to assess asymmetry between left and right gait parameters using the formula $SI[\%] = 100 \times |\text{left side} - \text{right side}| / (0.5 \times (\text{left side} + \text{right side}))$. Moreover, for each subject and each gait parameter, we calculated the dual-task effect (DTE) of walking as a percentage of loss (or gain) relative to the single-task walking performance, according to the formula

$DTE[\%] = 100 \times (\text{dual-task score} - \text{single-task score}) / \text{single-task score}$. A positive DTE corresponded to an increase in the value of the gait parameter under dual-tasking, whereas a negative DTE corresponded to a decrease.

7.2.6 Statistical Analysis

Participants were stratified into two groups, according to their MoCA scores, based on the cut-offs proposed by [168] and [334], as follows: they were considered controls with normal cognition if they had MoCA score ≥ 24 (control group); they were considered to have cognitive impairment (CI) when MoCA score < 24 (CI group). In the CI group, two persons could not finish the TMT B test, as it took too long. For this reason, their values were replaced by the maximum.

The statistical analysis was performed using SPSS 27.0 for Mac OS (SPSS Inc, Chicago, Illinois, USA) and Python 3 (using the packages pingouin v0.4.0, StatsModels v0.13.0, SciPy v1.7.1, and NumPy v1.21.5). A p-value of less than 0.05 indicated statistical significance.

Demographic and medical characteristics were summarized using either means and standard deviations, or frequencies and percentages, as appropriate. Comparisons between groups were made using the Independent Samples T-test or the Mann-Whitney U test for continuous variables, and Pearson's chi-square test (χ^2) or Fisher's exact test for categorical variables. The Mann-Whitney U test was used when data were not normally distributed or when outliers were present. Fisher's exact test was used when at least one cell of the contingency table had an expected frequency of less than 5. Asymptotic significance was reported for the Mann-Whitney U. The assumption of homogeneity of variance was checked using Levene's test; when $p < 0.05$, equal variances could not be assumed, and the Welch's t-test was computed instead of the standard Independent Samples T-test.

To assess differences between single-task and dual-task conditions within each group (CI and control), we used Paired Samples T-tests or Wilcoxon signed-rank tests. Wilcoxon signed-rank tests were used when paired data differences were not normally distributed or when outliers were present. Asymptotic significance was reported for Wilcoxon.

Group differences in gait parameters were examined using the Independent Samples T-test or the Mann-Whitney U test. Interaction between gait condition (single-task vs dual-task) and group (control group vs CI group) was assessed by comparing the DTEs between groups.

Data were tested for normality using the Shapiro-Wilk test. Outliers were detected using box plots.

To examine how well gait performance could distinguish between CI and control groups, we fitted logistic regression models. Separate models were built for each gait parameter that differed significantly between groups in previous steps. Subsequently, receiver operating characteristic (ROC) curves were constructed, and the sensitivity, specificity, and area under the curve (AUC) were estimated. The optimal cut-point was considered the point of the ROC curve closest to the upper left corner (0,1). Mathematically, the optimal sensitivity and specificity were defined as those yielding the minimum value of $(1 - \text{sensitivity})^2 + (1 - \text{specificity})^2$ [369]. AUC values were interpreted as follows: $AUC = 0.5$, no discrimination; $0.5 < AUC < 0.7$, poor discrimination;

$0.7 \leq AUC < 0.8$, acceptable discrimination; $0.8 \leq AUC < 0.9$, excellent discrimination; and $AUC \geq 0.9$, outstanding discrimination [370].

Gait metrics that differed significantly between groups were included as predictors of MoCA scores (global cognition) in bivariate linear regression models to test their association with a progressive score. To obtain the best set of variables associated with MoCA scores, we used a stepwise forward approach, successively entering variables with $p < 0.05$ and removing those with $p > 0.1$. The stepwise regression analysis was repeated controlling for age, sex, and education.

7.3 Results

A total of 29 participants (average age 81.24 ± 8.35 years old, 48.3% women) were enrolled in the study, from which 13 were considered healthy controls and 16 had cognitive impairment according to MoCA.

7.3.1 Descriptive Statistics

Table 7.1 presents the descriptive statistics for the demographic and medical data, stratified by cognitive status (healthy controls *vs* cognitively impaired). Participants differed in global cognition (MoCA score), TMT A and B scores, and CWIT inhibition times. The differences in MoCA scores between groups were expected as they defined membership in each group. The times of TMT A, TMT B, and CWIT inhibition were significantly higher in the group with MoCA score < 24 , indicating worse performance on these tests. The remaining variables, including demographic characteristics (e.g., age, BMI, education, etc.) and physical test metrics (SPPB and MSTs scores) presented no significant differences between the two groups. For this reason, these variables were not accounted as covariates in the subsequent steps of the statistical analysis.

7.3.2 Differences in Gait Between Groups

The differences in average temporal gait metrics between groups—control *vs* CI group—and within groups—single- *vs* dual-task conditions—are summarized in Table 7.2. Compared to the single-task condition, walking under dual-task led to significant changes in all temporal gait metrics in the CI group. The largest effect was observed on the foot-flat ratio in the CI group, where the addition of the dual task increased the foot-flat ratio by 21.7%. Accordingly, the addition of the dual task decreased the pushing ratio by 9.7% and the loading ratio by 9.9% in the CI group. The average stride time increased with the addition of the cognitive task, as did the swing time and stance time. However, the increase in stance time (of 17.0%) was higher than the increase in swing time (8.5%), which was reflected as an increase in stance ratio and a decrease in swing ratio. Similar effects were observed in the control group, but these effects were less pronounced. For instance, the addition of the dual task only increased the foot flat ratio by 8.0% in the control group (versus 21.7% in the CI group). Not all temporal gait metrics in the control group were significantly influenced by the dual-task condition.

None of the single-task or dual-task average temporal gait metrics alone presented significant differences between groups, except dual-task swing time whose average was higher in the CI group. The DTE on stride time, swing time, stance time, foot flat ratio, and pushing ratio differed significantly between groups.

Differences in average spatial gait metrics are shown in Table 7.3. As observed in temporal metrics, all spatial gait metrics were significantly affected by the addition of the cognitive task in the CI group. The addition of the dual task led to a significant decrease in all average spatial metrics, being the largest effects observed on strike angle (25.2%) and gait speed (24.4%) in the CI group. In the control group, only minimum toe clearance was not significantly affected by the dual-task condition. None of the single-task or dual-task average spatial gait metrics presented significant differences between groups. The DTE on gait speed, liftoff angle, minimum toe clearance, and heel 3D path length differed significantly between groups.

Figure 7.1 shows interaction plots for some of the gait metrics whose DTEs presented a statistically significant difference between groups in Tables 7.2 and 7.3. Compared to single-task, the dual-task condition had a significant effect on all gait parameters and groups presented in Figure 7.1, except the pushing ratio in the control group (Table 7.2). The effects were more pronounced in the CI group leading to significant differences in the DTEs. In Figure 7.1, the blue solid lines (CI group) had higher slopes than the orange dashed lines (control group) which denoted these effects.

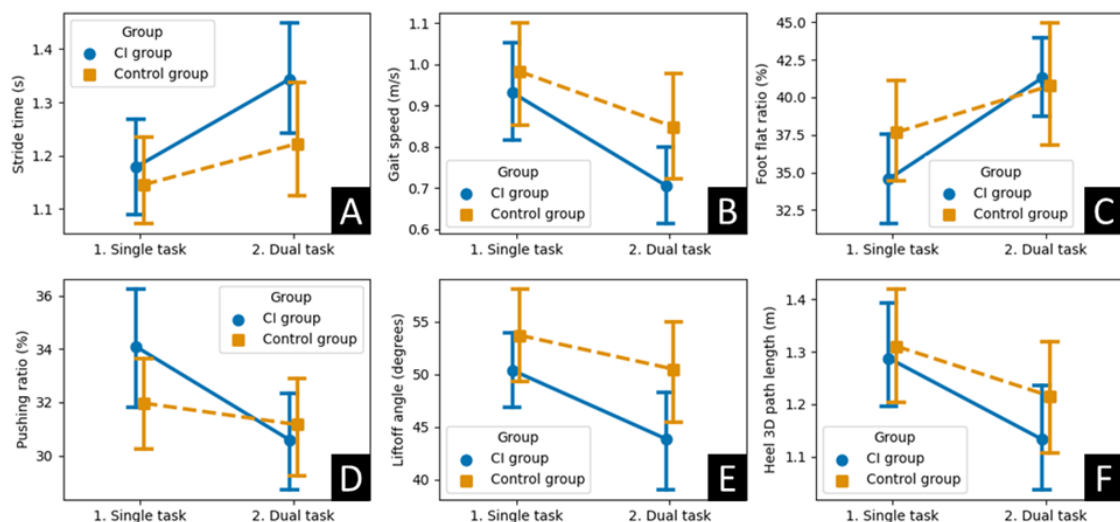


Figure 7.1: Group $\times$ task interaction for (A) stride time, in seconds, (B) gait speed, in m/s, (C) foot flat ratio, in %, (D) pushing ratio, in %, (E) liftoff angle, in degrees, and (F) heel 3D path length, in meters. Errors bars represent 95% confidence intervals.

Differences between groups on dual-task variability metrics are shown in Table 7.4. Some of the metrics presented significant differences between groups, being their averages higher in the CI group. Among temporal metrics, stride, stance, and swing times variability presented significant

differences between groups, whereas none of the gait phase ratios did. Speed and liftoff angle variabilities also presented significant differences between groups, although the differences in liftoff angle variability were marginally significant. The differences in stride length variability were significant and consistent with the differences in heel and toe trajectory forward MAD, and heel and toe 3D path length variability—all of them related to increased variability of foot trajectories in the CI group under dual-task gait testing.

Table 7.5 shows the differences between groups on dual-task asymmetry metrics. Only two asymmetry measures (on loading ratio and liftoff angles) presented significant differences between groups, although these differences were marginally significant. The average values of loading ratio asymmetry and liftoff angles asymmetry were higher in the CI group.

7.3.3 Prediction of Cognitive Impairment

Gait metrics that differed significantly between groups in the previous section were included in the ROC analysis to test their capability to accurately identify clinical groups. Table 7.6 shows the classification performance of each gait parameter based on ROC analysis.

All metrics, except speed DTE and heel 3D path length DTE, had statistically significant ROC curves (AUC significantly higher than chance) and provided acceptable discrimination according to [370] ($0.7 \leq AUC < 0.8$). The best combination of sensitivity and specificity was achieved with dual-task heel and toe trajectories forward MAD, both with $AUC = 0.745$ ($p < 0.05$) and able to detect CI group membership with a sensitivity of 81.3% and specificity of 76.9%. The pushing ratio DTE had the highest AUC, followed by foot flat ratio DTE, dual-task stride length variability, and dual-task stride time variability, all with $AUC > 0.75$ ($p < 0.05$).

7.3.4 Association with Global Cognitive Score

Gait metrics that differed significantly between groups were included as predictors of MoCA scores in multiple bivariate linear regression models to test the association of each gait metric with the global cognitive score (progressive cognitive impairment). Results are shown in Table 7.7.

All tested DTEs presented a significant association with MoCA scores. Additionally, dual-task heel and toe path length variabilities, and dual-task loading ratio asymmetry presented a significant negative association with MoCA. The DTE on gait speed explained the highest percentage (over 25%) of the variance in MoCA scores ($R^2 = 0.258$).

The model resulting from the stepwise forward regression analysis only included the DTE on speed ($B = 0.114$, $p = 0.005$) as an independent variable. For each 10% decrease in speed DTE (%), MoCA scores decreased roughly by 1 point (Figure 7.2). None of the potential covariates (i.e., age, sex, or education) were selected for the final model.

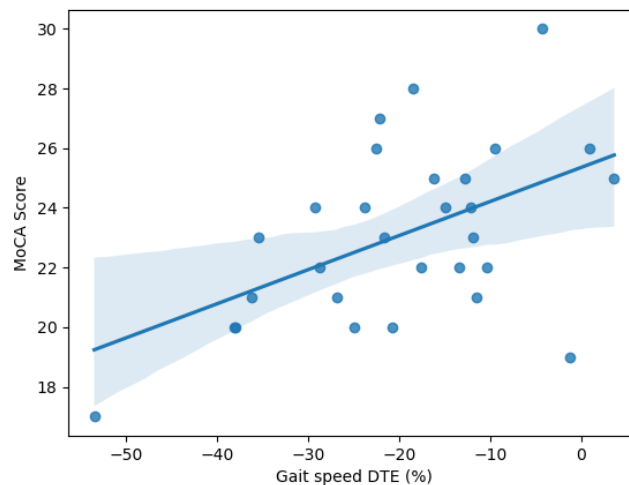


Figure 7.2: Regression model for the prediction of MoCA scores.

7.4 Discussion

This study investigated the potential of the newly developed foot-worn inertial sensor-based gait analysis solution to assess gait metrics that are affected by the cognitive status of older adults using single- and dual-task gait assessments. Our contributions were the following: (i) we inspected whether the novel foot-worn inertial sensor-based gait analysis solution was able to distinguish between older adults with and without cognitive impairment; (ii) we explored differences between these two groups considering a comprehensive set of gait parameters, including dual-task metrics and dual-task costs—some of them not yet explored in the literature; (iii) we investigated which gait parameters measured by the foot-worn inertial sensors were the most relevant to assess cognitive status; and (iv) for the first time, we investigated these topics in a group of older adults with mobility limitations: a very common condition among older adults, with an estimated prevalence of 23% to 47% [67].

In line with our initial hypothesis, the results of this study showed that the CI group tended to be more influenced by the dual-task than the control group, which translated into significant differences between groups on several DTEs (Tables 7.2 and 7.3) and some dual-task variability (Table 7.4) and asymmetry (Table 7.5) metrics. Also, several of these metrics provided acceptable discrimination ability (according to the ROC analysis, Table 7.6) and had a significant association with MoCA scores (Table 7.7). The signs of the regression coefficients revealed positive or negative associations as expected due to the detrimental effects of the cognitive task on gait performance. The dual-task effect on gait speed explained the highest percentage of the variance in MoCA scores ($> 25\%$). As hypothesized, dual-task gait metrics and dual-task costs had better discrimination ability than single-task gait metrics. None of the single-task gait metrics presented significant differences between groups.

Overall, our preliminary results confirmed that the newly developed gait analysis solution is a pertinent tool to evaluate gait metrics that are affected by the cognitive status of older adults relying on single- and dual-task gait assessments.

7.4.1 The Effect of the Dual Task on Walking

This study examined the effects of dual tasking on several gait metrics. The observed increase in stance and swing times, the decrease in swing ratio, the increase in foot flat ratio, and the decrease in all average spatial metrics—including gait speed—are consistent with a poorer gait performance due to the adoption of a more cautious gait under dual-tasking, i.e., when less attentional resources were available for walking [371]. Slower gait speed due to dual tasking was also reported previously in [372] and in [56].

Gait speeds below 1.0 m/s are commonly used to identify individuals with increased risk of falls [203], which is relevant under dual-task testing as most of the activities of daily living involve the simultaneous performance of cognitive and motor tasks [188,203]. In our study, both mobility-impaired groups had an average gait speed < 1.0 m/s, which was further lowered in dual-task conditions and in the CI group, potentially denoting an increased fall risk under real-life situations. The fact that our participants had mobility impairment may explain the lower gait speeds even under single-task conditions.

The fact that cognitively normal older adults are also affected by the dual task is not surprising, as with aging structural changes occur in the brain and gait becomes less automatic [42,188]. Moreover, our cognitively normal group had mobility limitations (SPPB score < 10). Considering that motor and cognitive functions are interrelated, and that intact cognition is required for motor control, it is expected that in the CI group the quality of walking while performing an attention-demanding task is further decreased, which may serve as an early indicator of cognitive impairment [43,44,52].

According to several past studies, the effect of the dual task depends on the complexity or load of the cognitive task used. The cognitive task of counting backward by sevens has been consistently reported as one of the most complex tasks, leading to greater dual-task effects than less complex tasks such as counting backward by ones or naming animals [52,53,203–205,372]. The task of counting backward used in our study seems to be appropriate and explains the high dual-task effects observed in our participants. According to [53], a high cognitive load is required in a dual-task protocol for making MCI-specific gait changes emerge.

7.4.2 Differences Between Groups

None of the single-task gait metrics presented significant differences between groups in our study. While most of the studies agreed that dual-task gait testing could create higher differences in gait between cognitively impaired and cognitively healthy groups, some studies could still find some differences under single-task conditions. For instance, gait speed was significantly lower in patients with MCI under single-task in [205], [373], [59] and [57]. In [203] gait speed under

single-task gait testing did not present significant differences between groups, but, under dual-task, the MCI group differed from the control group. Similar results were obtained for gait speed in [57], [56], [206], and [374], and for stride time variability in [56]. A recent meta-analysis showed that the differences between groups on gait speed under dual-task were more obvious than in single-task, with effects sizes of -0.89 and -0.74 , respectively [52].

Muir et al. [203] showed that the DTE on gait speed was higher in MCI than in the cognitively normal group. Other studies reported significant interactions between gait condition (single-task vs dual-tasks) and group (CI group vs control group) on gait speed [56, 374], stride time [55, 204], and stride time variability [56, 204], suggesting that the effect of the dual task was greater in the CI group than in the control group. In our study, the DTEs on gait speed and stride time were significantly different, which is in line with past studies' results. In [204], however, the interaction was not significant for gait speed, which contradicts the results obtained in our study and other studies in this field. Contradictory results were also obtained in [202], in which group $\times$ condition interaction was not significant for all metrics reported, i.e., gait speed, stride time, and double support ratio. Besides having included a very low number of participants (low power), in [202] the cognitive tasks (counting backward by ones, short story recall, and a phonemic fluency task) may not have been challenging enough to cause a significant effect on gait [53] (as discussed above).

In addition to the DTE on gait speed and stride time, the DTE on foot flat ratio, liftoff angle, minimum toe clearance, and heel 3D path length presented significant differences between groups in our study. None of these metrics have been studied in the past, although they are coherent with a poorer gait performance and the adoption of a more cautious gait, well evidenced by the slower speeds and shorter displacements of the feet during swing caused by higher motor-cognitive interferences in the CI group. Metrics derived from heel and toe trajectories that provided significant differences between groups were obtained by resorting to a deep learning-based approach (as documented in Chapter 4), which denotes the potential of these novel techniques (and metrics) and their role in the identification of cognitive impairment.

The differences between groups were also significant for several dual-task variability metrics (according to Table 7.4) and marginally significant for two asymmetry metrics (according to Table 7.5). In [203] and in [56], stride time variability under dual-task gait testing was significantly higher in the MCI group, which is consistent with our results. In [206], symmetry was significantly affected by the dual task in the MCI group, although the differences between groups were not significant. Gait variability is known to rely on higher cortical brain control, such that an increased gait variability reflects cognitive and cortical deterioration [375]. Consistently, gait variability under dual-tasking was further exacerbated in the CI group according to our results.

7.4.3 Cognitive Status Evaluation

Most of the studies agreed that dual-task gait testing could better differentiate between groups than single-task gait tests [53]. In our study, most dual-task gait metrics and DTEs could differentiate

between groups, providing acceptable discrimination ($0.7 \leq AUC < 0.8$). In [374], the discrimination ability of the dual-task gait speed was classified as fair-to-good ($AUC = 0.738$) and was higher than the single-task gait speed, which was classified as poor-to-fair ($AUC = 0.674$). In [55], the dual-task gait performance on gait time could distinguish subjects with subjective complaints from patients with MCI, producing an AUC value of 0.63. In [205], the DTE on gait speed differed between MCI and control groups, with the MCI group experiencing higher costs than the controls. According to [52], the DTE on gait speed provides high sensitivity for the detection of MCI. In our study, the DTE on gait speed could not discriminate between CI and control groups, although other DTE metrics could (namely, the DTE on stride time, foot flat ratio, pushing ratio, liftoff angle, and minimum toe clearance, all with $AUC > 0.70$ —Table 7.6). In contrast, the DTE on gait speed presented a significant (positive) association with MoCA scores, explaining the highest percentage (over 25%) of the variance in MoCA scores (Table 7.7).

All tested dual-task variability metrics and dual-task asymmetry metrics provided acceptable discrimination, but only three of them (i.e., dual-task heel and toe 3D path length variability, and dual-task loading ratio asymmetry) presented a significant (negative) association with MoCA scores. All tested DTEs presented a significant association with global cognitive score (Table 7.7). Past studies have also reported significant associations of gait metrics with global cognitive scores. For instance, in [41] all gait parameters (i.e., speed, stride length, cadence, and duration of stance, swing, and double support phases) had a significant correlation with cognitive status as expressed by the Addenbrooke's Cognitive Examination Revised (ACE-R) score. In [373] gait speed and stride length were positively correlated with ACE-R. In [376], MoCA scores were significantly correlated with gait speed and stride length.

7.4.4 Study Strengths and Limitations

Most of the previous studies relied on conventional gait analysis instruments like instrumented walkways (e.g., in [377], [203], [51], [59], and [205]), or accelerometers placed on the lower trunk (e.g., in [57], [206], [378], [177], [373], and [41]) to evaluate the gait of older adults with and without cognitive impairment. Although foot-worn inertial sensors and a newly developed method for gait analysis have been employed in this study, our results concerning the assessment of cognitive status showed no inferior discrimination ability, which denotes the potential of the newly developed method. The proposed solution is convenient for use in clinical settings due to its simplicity and flexibility.

In the past, studies have barely explored the utility of DTEs to discriminate between older adults with and without cognitive impairment. In our study, the DTEs on several gait metrics provided significant differences between groups (Table 7.2 and Table 7.3), could discriminate between groups (Table 7.6), and presented a significant association with MoCA scores (Table 7.7). DTEs offer great potential to distinguish between groups as they compare dual-task gait performance with baseline (single-task) gait performance, thereby eliminating individual baseline characteristics and representing only the effects due to the addition of the cognitive task. Other metrics, like foot flat, pushing, and loading ratios, strike and liftoff angles, minimum toe clearance, and heel

and toe trajectories, have never been explored in related literature which constitutes a novel contribution of our work. Parameters that are dependent on the interrelationship between both feet (e.g., double support time) could not be evaluated in our study since inertial sensors were not temporally synced. According to past works, double support time could also be an important metric to discriminate older adults with cognitive impairment [41,377].

Metrics known to influence gait parameters, like age, sex, or height, were not considered in the analysis of group differences, because they did not differ between groups in the descriptive analysis. Yet, other potential covariates (e.g., the size of the leg according to [59], medication effects, or others) were not considered in this study, which may affect the results. For instance, some studies have considered gait speed as a covariate to control for the influence on variability parameters [59,375] although, according to [379], the greater variability observed in the gait of older adults may result more from loss of strength and flexibility than from their slower speeds. Nevertheless, the use of the DTEs may mitigate some of the effects of the unknown or unconsidered covariates as their influences on baseline gait data are not expected to affect the dual-task decrements, i.e., study participants are considered their own baselines [55].

The utility of discrete gait characteristics to discriminate between groups was determined by the AUC of ROC curves. AUC values were between 0.7 and 0.8, indicating that none of the separate gait parameters had a strong individual predictive power. It should be noted that the reported AUC values may not accurately represent the generalization performance of the logistic regression models, as they were not evaluated in an independent test set. Although previous studies have also reported AUC values based on logistic regression models trained on the entire dataset (e.g., [375], [64], [59], [374], or [55]), it is important to interpret the results cautiously due to the potential presence of overfitting, even though only one feature (one gait metric) was used to fit the models (less prone to overfitting). Previous works suggested that the combination of metrics in logistic regression models could provide better discrimination ability [59,60]. Considering that we had rather few participants in this study and to further avoid the issue of overfitting, we did not combine gait metrics. On the other side, several gait metrics might be highly related and should not provide additional discrimination ability. In several studies, gait metrics were summarized by their independent factors using factor analysis, yielding 3 to 5 independent domains, like pace, rhythm, variability, asymmetry, and postural control [60,187,375,377], that differed significantly between groups [375,377]. Due to the limited size of the sample in our study, we were not able to assess the independence of the gait characteristics and reduce the dimensionality of the problem. Our results thus warrant further investigations with larger and more heterogenic groups of older adults, specifically focused on assessing more complex relationships between variables and evaluating the models on independent test sets.

Also, we could not compare our results with a strict control group, i.e., a sample without mobility impairment, although mobility impairment is known to affect gait performance [380]. We should investigate whether the dual-task assessments are or are not specific to the detection of cognitive impairment when people without mobility impairment (as determined by the SPPB scores) are included. The newly developed gait analysis solution was able to find differences between

those with and without cognitive impairment in a sample with mobility limitations. Considering that mobility limitations are very common in older adults—and frequently co-exist with cognitive impairment [36, 37, 48]—it is important to ensure that the developed solution can reliably assess gait in participants with these characteristics. Further research is, however, required.

In addition to the relatively small number of participants involved (low statistical power), other study limitations should be pointed out. First, CI and control groups were stratified according to their MoCA scores—the diagnosis was not provided by the neuropsychologists. Although MoCA is considered the preferred screening tool for primary care settings, it has a sensitivity of 81-97% [22], which may not be reliable. Additionally, the order of the gait tests (single-task and dual-task) was fixed, which may have served as a limitation if individuals with cognitive impairment would fatigue more easily than those without cognitive impairment. Although gait was affected by the dual task and could discriminate between groups, the effects of walking on the cognitive task were not evaluated in this study. According to [200], the correct interpretation of the dual-task interference requires an evaluation of both tasks under single- and dual-task conditions as an effect might also be observed on the performance of the cognitive task that may support the evaluation of the cognitive status (as in [177], [374], and [55]).

7.4.5 Future Work

Further research is required to demonstrate the validity and reliability of gait metrics for clinical decision-making. We should acquire data from a higher number of participants, include older adults without mobility limitations, and ask neuropsychologists to provide a formal clinical diagnosis of the participants. We should explore more advanced data analysis techniques, including data-driven modeling approaches, e.g., based on machine learning, to unveil more complex relationships between variables and more complex patterns within the data. The models developed to detect cognitive impairment should be evaluated in an independent test set (never utilized in the modeling phase) to assess the generalization performance of the model. Feature selection and dimensionality reduction techniques should be applied to reduce the number of input variables and ensure their independence. We should not only evaluate gait but also evaluate the performance of the cognitive task under single- and dual-task testing to ensure a complete evaluation of the motor-cognitive interference. Additional cognitive tasks might be explored to maximize the interference of the dual-task condition and its effects on the performance of gait and cognition. Finally, we should investigate the performance and the usefulness of the newly developed methods in a real application scenario using a longitudinal study design to estimate the predictive value of cognitive decline based on gait measures.

7.5 Conclusion

The current pilot study strengthens the current evidence that dual-task gait assessments can be an important tool for the assessment of cognitive impairment. By employing foot-worn inertial sensors and the newly developed gait analysis approach, the preliminary results of this study could

capture differences in multiple gait metrics related to differences in cognitive performance. Several dual-task effect metrics were explored for the first time, revealing an increased susceptibility of the cognitive impairment group to dual-task effects when compared to the control group. Further evaluation with a larger and more diverse sample is required to establish the feasibility and reliability of the system in practical scenarios. The validation of the solution in the future could make standardized dual-task assessments and foot-worn inertial sensor-based gait analysis applicable as a fast-screening tool in primary care settings to improve the early detection of cognitive impairment.

Table 7.1: Demographic and clinical characteristics of study participants stratified by cognitive status.

Participant Characteristics	Control Group (n=13)	CI group (n=16)	p-value
Age [years]	78.5 ± 9.2 (64-93)	83.5 ± 7.1 (68-96)	0.107
BMI [Kg/m ²]	28.0 ± 4.5 (21.3-33.6) [†]	25.3 ± 2.5 (20.9-31.0)	0.114
Weight [Kg]	80.5 ± 15.1 (58-110)	71.4 ± 13.2 (52-95)	0.095
Height [cm]	169.7 ± 6.4 (162-182) [†]	167.6 ± 11.6 (148-184)	0.529
Education [years]	13.3 ± 4.2 (8-24)	11.5 ± 1.9 (9-15)	0.232
Female, n (%)	8 (61.5)	6 (37.5)	0.198
Hearing problems, n (%)	7 (53.8)	6 (37.5)	0.379
Vision problems, n (%)	7 (53.8)	7 (43.8)	0.588
Dizziness, n (%)	3 (23.1)	8 (50.0)	0.249
Gait problems, n (%)	8 (61.5)	10 (62.5)	1.000
Fear of falling			
Never, n (%)	3 (23.1)	3 (18.8)	0.755
Sometimes, n (%)	7 (53.8)	9 (56.3)	
Often, n (%)	1 (7.7)	3 (18.8)	
Always, n (%)	2 (15.4)	1 (6.3)	
Number of falls in the last 6 months			
Zero, n (%)	8 (61.5)	9 (56.3)	1.000
One, n (%)	2 (15.4)	3 (18.8)	
More than one, n (%)	3 (23.1)	4 (25.0)	
Pain			
No pain, n (%)	1 (7.7)	2 (12.5)	0.869
Less often than daily, n (%)	6 (46.2)	9 (56.3)	
Daily, n (%)	6 (46.2)	5 (31.3)	
Pain severity			
Low, n (%)	6 (50.0)	6 (42.9)	0.640
Medium, n (%)	4 (33.3)	7 (50.0)	
Sometimes unbearable, n (%)	2 (16.7)	1 (7.1)	
Walking aids			
None, n (%)	8 (61.5)	5 (31.3)	0.275
Cane, n (%)	2 (15.4)	4 (25.0)	
Walker, n (%)	3 (23.1)	7 (43.8)	
Physical activity			
>3 times/week, n (%)	6 (46.2)	9 (56.3)	0.387
1-3 times/week, n (%)	7 (53.8)	5 (31.3)	
1 time/week, n (%)	0 (0.0)	2 (12.5)	
Never, n (%)	0 (0.0)	0 (0.0)	
Cognitive tests			
MoCA Score	25.7 ± 1.8 (24-30)	21.0 ± 1.6 (17-23)	<0.001*
TMT A Time [s]	51.1 ± 12.6 (30.5-68.8)	77.8 ± 32.5 (42.3-155.0)	0.016*
TMT B Time [s]	147.5 ± 74.0 (55.1-340.8)	264.4 ± 99.9 (91.4-386.6)	0.002*
CWIT inhibition time [s]	76.0 ± 24.2 (56.4-136.7)	110.3 ± 40.6 (59.5-187.5)	0.010*
WMSR Score	5.7 ± 1.2 (4-8)	4.9 ± 1.2 (3-7)	0.078
Physical tests			
SPPB Score	7.9 ± 1.3 (5-9)	7.0 ± 1.6 (4-9)	0.076
MSTS Nr of repetitions	14.3 ± 8.5 (0-26)	14.9 ± 6.2 (5-32)	0.742

Data are mean values ± standard deviation (range) or the number of participants per category (absolute and relative frequency) when indicated. Group differences were evaluated using the Independent Samples T-test, Mann-Whitney U test, Pearson's chi-square, or Fisher's exact test. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance. [†]Missing data for this variable in one individual in this group. BMI: Body Mass Index; MoCA: Montreal Cognitive Assessment; TMT: Trail Making Test; CWIT: Color Word Interference Test; WMSR: the Digit Backward Task of the Wechsler Memory Scale-Revised; SPPB: Short Physical Performance Battery; MSTs: 1-Minute Sit to Stand Test.

Table 7.2: Comparison of average temporal gait parameters within groups (single- vs dual-task) and between groups (control vs CI group), including dual-task effects (DTEs).

Gait metric	Control group (n=13)	CI group (n=16)	Statistic	p-value
Stride time DTE (%)	6.49 ± 6.63	14.28 ± 10.69	T = -2.291	0.03*
ST stride time (s)	1.15 ± 0.16	1.18 ± 0.19	U = 98.5	0.809
DT stride time (s)	1.22 ± 0.21	1.34 ± 0.23	U = 64.0	0.079
	T = -3.424; p = 0.005*	T = -4.804; p <0.001*		
Swing time DTE (%)	2.47 ± 6.45	8.52 ± 7.55	U = 56.0	0.035*
ST swing time (s)	0.35 ± 0.04	0.38 ± 0.05	U = 72.0	0.160
DT swing time (s)	0.36 ± 0.05	0.41 ± 0.06	U = 55.0	0.032*
	W = 25.5; p = 0.162	T = -4.175; p <0.001*		
Stance time DTE (%)	8.31 ± 7.4	16.96 ± 12.36	T = -2.216	0.035*
ST stance time (s)	0.79 ± 0.13	0.8 ± 0.14	U = 106.5	0.913
DT stance time (s)	0.86 ± 0.17	0.93 ± 0.18	U = 74.5	0.196
	W = 3.0; p = 0.003*	W = 5.0; p = 0.001*		
Swing ratio DTE (%)	-3.49 ± 3.12	-4.64 ± 3.74	T = 0.881	0.386
ST swing ratio (%)	31.06 ± 2.01	32.36 ± 1.91	T = -1.778	0.087
DT swing ratio (%)	29.98 ± 2.18	30.88 ± 2.44	U = 71.0	0.148
	T = 4.114; p = 0.001*	T = 5.354; p <0.001*		
Stance ratio DTE (%)	1.57 ± 1.37	2.18 ± 1.58	T = -1.079	0.290
ST stance ratio (%)	68.94 ± 2.01	67.64 ± 1.91	T = 1.778	0.087
DT stance ratio (%)	70.02 ± 2.18	69.12 ± 2.44	U = 137.0	0.148
	T = -4.114; p = 0.001*	T = -5.354; p <0.001*		
Foot flat ratio DTE (%)	7.99 ± 7.92	21.74 ± 18.37	T = -2.7	0.013*
ST foot flat ratio (%)	37.67 ± 6.25	34.56 ± 6.73	U = 137.0	0.148
DT foot flat ratio (%)	40.76 ± 7.67	41.29 ± 5.77	U = 108.0	0.861
	T = -3.637; p = 0.003*	T = -5.208; p <0.001*		
Pushing ratio DTE (%)	-2.50 ± 6.45	-9.70 ± 8.44	U = 159.0	0.016*
ST pushing ratio (%)	31.97 ± 3.25	34.09 ± 4.88	T = -1.342	0.191
DT pushing ratio (%)	31.16 ± 3.67	30.59 ± 3.84	U = 117.0	0.569
	W = 23.0; p = 0.116	W = 3.0; p <0.001*		
Loading ratio DTE (%)	-8.18 ± 6.3	-9.89 ± 10.47	U = 130.0	0.254
ST loading ratio (%)	30.36 ± 5.0	31.35 ± 5.5	U = 98.0	0.792
DT loading ratio (%)	28.07 ± 6.02	28.12 ± 5.08	U = 106.5	0.913
	W = 2.0; p = 0.002*	W = 13.0; p = 0.004*		

Data are mean values ± standard deviation. Between-group differences were evaluated using the Independent Samples T-test (T) or the Mann-Whitney U test (U), as appropriate. Differences between single- and dual-task measures were evaluated using Paired Samples T-tests (T) or Wilcoxon signed-rank tests (W), as appropriate. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance. DTE: dual-task effect; ST: single-task; DT: dual-task.

Table 7.3: Comparison of average spatial gait parameters within groups (single- vs dual-task) and between groups (control vs CI group), including dual-task effects (DTEs).

Gait metric	Control group (n=13)	CI group (n=16)	Statistic	p-value
Stride length DTE (%)	-9.07 ± 7.48	-15.33 ± 9.14	U = 144.0	0.079
ST stride length (m)	1.10 ± 0.19	1.06 ± 0.18	U = 118.5	0.525
DT stride length (m)	1.00 ± 0.21	0.91 ± 0.2	U = 131.0	0.236
	T = 4.803; p <0.001*	T = 8.160; p <0.001*		
Speed DTE (%)	-13.94 ± 9.79	-24.39 ± 13.56	T = 2.325	0.028*
ST speed (m/s)	0.98 ± 0.24	0.93 ± 0.25	T = 0.560	0.580
DT speed (m/s)	0.85 ± 0.23	0.71 ± 0.21	T = 1.743	0.093
	W = 3.0; p = 0.003*	T = 6.853; p <0.001*		
Strike angle DTE (%)	-17.04 ± 20.04	-25.24 ± 30.65	U = 129.0	0.273
ST strike angle (°)	-20.20 ± 7.94	-21.13 ± 7.09	T = 0.335	0.74
DT strike angle (°)	-17.48 ± 8.21	-17.14 ± 7.72	U = 105.0	0.965
	T = -4.210; p = 0.001*	T = -7.123; p <0.001*		
Liftoff angle DTE (%)	-6.05 ± 5.28	-13.56 ± 9.58	U = 153.0	0.032*
ST liftoff angle (°)	53.69 ± 8.77	50.38 ± 7.29	T = 1.110	0.277
DT liftoff angle (°)	50.48 ± 8.93	43.87 ± 9.31	T = 1.936	0.063
	W = 5.0; p = 0.005*	T = 6.598; p <0.001*		
Min toe clearance DTE (%)	-6.35 ± 18.26	-18.36 ± 22.54	U = 149.0	0.048*
ST min toe clearance (cm)	1.14 ± 0.63	1.60 ± 1.1	U = 65.0	0.087
DT min toe clearance (cm)	1.07 ± 0.62	1.25 ± 0.8	U = 90.0	0.539
	T = 1.238; p = 0.239	W = 12.0; p = 0.004*		
Heel 3D path length DTE (%)	-7.28 ± 5.85	-12.23 ± 6.21	T = 2.190	0.037*
ST heel 3D path length (m)	1.31 ± 0.2	1.29 ± 0.21	T = 0.297	0.769
DT heel 3D path length (m)	1.22 ± 0.2	1.13 ± 0.21	T = 1.067	0.296
	T = 4.531; p <0.001*	T = 8.063; p <0.001*		
Toe 3D path length DTE (%)	-9.06 ± 7.23	-14.6 ± 7.63	T = 1.991	0.057
ST toe 3D path length (m)	1.21 ± 0.23	1.19 ± 0.24	T = 0.277	0.784
DT toe 3D path length (m)	1.11 ± 0.24	1.02 ± 0.24	U = 125.0	0.357
	T = 4.689; p <0.001*	W = 0.0; p <0.001*		

Data are mean values ± standard deviation. Between-group differences were evaluated using the Independent Samples T-test (T) or the Mann-Whitney U test (U), as appropriate. Differences between single- and dual-task measures were evaluated using Paired Samples T-tests (T) or Wilcoxon signed-rank tests (W), as appropriate. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance. DTE: dual-task effect; ST: single-task; DT: dual-task.

Table 7.4: Comparison of dual-task variability measures between groups.

Gait metric	Control group (n=13)	CI group (n=16)	Statistic	p-value
Stride time variability (%)	4.81 ± 4.77	7.23 ± 5.78	U = 51.0	0.020*
Swing ratio variability (%)	6.74 ± 3.86	8.43 ± 4.73	U = 67.5	0.109
Stance ratio variability (%)	2.81 ± 1.42	3.63 ± 1.63	U = 65.0	0.087
Swing time variability (%)	6.87 ± 3.45	10.81 ± 8.02	U = 49.0	0.016*
Stance time variability (%)	6.62 ± 7.03	8.91 ± 6.37	U = 56.5	0.037*
Foot flat ratio variability (%)	11.75 ± 4.67	14.71 ± 5.32	T = -1.570	0.128
Pushing ratio variability (%)	9.80 ± 4.29	14.45 ± 7.38	U = 67.0	0.105
Loading ratio variability (%)	12.90 ± 5.45	16.81 ± 7.71	U = 74.0	0.188
Speed variability (%)	7.91 ± 4.81	10.76 ± 4.33	U = 56.0	0.035*
Stride length variability (%)	5.62 ± 3.57	7.63 ± 2.97	U = 50.0	0.018*
Strike angle variability (%)	-29.18 ± 38.95	8.45 ± 104.41	U = 106.0	0.930
Liftoff angle variability (%)	8.57 ± 4.27	13.00 ± 7.09	U = 58.0	0.044*
Min toe clearance variability (%)	42.57 ± 13.87	44.56 ± 20.68	T = -0.297	0.769
Heel trajectory forward MAD (mm)	25.22 ± 8.94	32.23 ± 8.54	U = 53.0	0.025*
Heel trajectory sideward MAD (mm)	2.65 ± 0.57	2.47 ± 0.40	T = 1.007	0.323
Heel trajectory vertical MAD (mm)	5.61 ± 2.02	6.47 ± 1.93	U = 80.0	0.293
Toe trajectory forward MAD (mm)	25.28 ± 9.29	32.03 ± 7.85	U = 53.0	0.025*
Toe trajectory sideward MAD (mm)	2.80 ± 0.89	2.57 ± 0.82	U = 117.0	0.569
Toe trajectory vertical MAD (mm)	2.41 ± 0.82	2.69 ± 0.61	U = 77.0	0.236
Heel 3D path length variability (%)	4.70 ± 3.13	6.47 ± 3.86	U = 59.0	0.048*
Toe 3D path length variability (%)	5.55 ± 3.28	7.41 ± 2.97	U = 53.0	0.025*

Data are mean values ± standard deviation. Group differences were evaluated using the Independent Samples T-test (T) or the Mann-Whitney U test (U), as appropriate. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance. MAD: Mean Absolute Deviation.

Table 7.5: Comparison of dual-task asymmetry measures between groups.

Gait metric	Control group (n=13)	CI group (n=16)	Statistic	p-value
Stride time asymmetry (%)	0.35 ± 0.32	0.52 ± 0.65	U = 92.5	0.614
Stance ratio asymmetry (%)	1.74 ± 1.16	2.72 ± 2.55	U = 85.0	0.405
Swing time asymmetry (%)	3.99 ± 2.91	6.27 ± 6.09	U = 89.5	0.525
Foot flat ratio asymmetry (%)	10.01 ± 7.86	13.30 ± 10.06	T = -0.960	0.345
Pushing ratio asymmetry (%)	7.95 ± 6.43	14.06 ± 14.70	U = 77.0	0.236
Loading ratio asymmetry (%)	7.47 ± 6.96	13.19 ± 7.83	U = 59.0	0.048*
Stride length asymmetry (%)	1.57 ± 0.99	2.79 ± 2.28	U = 66.5	0.100
Speed asymmetry (%)	1.20 ± 1.02	2.59 ± 2.36	U = 61.0	0.059
Strike angle asymmetry (%)	-29.78 ± 52.78	-2.01 ± 42.12	U = 79.0	0.273
Liftoff angle asymmetry (%)	5.55 ± 5.74	11.09 ± 8.18	U = 57.0	0.039*
Min toe clearance asymmetry (%)	48.14 ± 24.12	40.22 ± 32.78	T = 0.725	0.475
Heel 3D path length asymmetry (%)	2.00 ± 1.99	3.16 ± 3.62	U = 86.0	0.430
Toe 3D path length asymmetry (%)	2.79 ± 1.98	3.53 ± 2.56	T = -0.856	0.400

Data are mean values ± standard deviation. Group differences were evaluated using the Independent Samples T-test (T) or the Mann-Whitney U test (U), as appropriate. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance.

Table 7.6: Classification performance based on logistic regression and ROC analysis (n=29).

Gait metric	AUC	SE	p-value	Sens./Spec.
Stride time DTE (%)	0.740	0.097	0.028*	0.813 / 0.692
Foot flat ratio DTE (%)	0.760 [†]	0.090	0.018*	0.625 / 0.846
Pushing ratio DTE (%)	0.764 [†]	0.092	0.016*	0.750 / 0.769
Speed DTE (%)	0.707	0.096	0.059	0.625 / 0.692
Liftoff angle DTE (%)	0.736	0.094	0.032*	0.688 / 0.769
Min toe clearance DTE (%)	0.716	0.097	0.048*	0.750 / 0.615
Heel 3D path length DTE (%)	0.712	0.097	0.054	0.625 / 0.769
DT Stride time variability (%)	0.755 [†]	0.095	0.020*	0.750 / 0.769
DT Stride length variability (%)	0.760 [†]	0.101	0.018*	0.750 / 0.769
DT Speed variability (%)	0.731	0.101	0.035*	0.750 / 0.692
DT Liftoff angle variability (%)	0.721	0.098	0.044*	0.688 / 0.846
DT Heel trajectory forward MAD (mm)	0.745	0.097	0.025*	0.813 / 0.769
DT Toe trajectory forward MAD (mm)	0.745	0.100	0.025*	0.813 / 0.769
DT Heel 3D path length variability (%)	0.716	0.100	0.048*	0.750 / 0.692
DT Toe 3D path length variability (%)	0.745	0.098	0.025*	0.750 / 0.692
DT Loading ratio asymmetry (%)	0.716	0.102	0.048*	0.938 / 0.538
DT Liftoff angle asymmetry (%)	0.726	0.095	0.039*	0.750 / 0.692

Sensitivity (Sens.) and Specificity (Spec.) were obtained for the optimal cut-point based on ROC curve analysis. p-values under the null hypothesis that AUC equals 0.5. * $p < 0.05$, bold values indicating significance. [†] $AUC > 0.75$. AUC: Area Under the Curve; ROC: Receiver-operating characteristic; SE: Standard Error (under nonparametric assumption); DTE: Dual-task effect; DT: Dual-task; MAD: Mean Absolute Deviation.

Table 7.7: Associations between gait metrics and MoCA scores in multiple bivariate linear regression analysis (n=29).

Gait metric	Model Summary			Coefficients		
	R ²	Adj. R ²	SE	B	β	p
Stride time DTE (%)	0.193	0.163	2.660	-0.131	-0.439	0.017*
Speed DTE (%)	0.258	0.230	2.550	0.114	0.508	0.005*
Foot flat ratio DTE (%)	0.173	0.142	2.693	-0.076	-0.416	0.025*
Pushing ratio DTE (%)	0.162	0.130	2.711	0.140	0.402	0.031*
Liftoff angle DTE (%)	0.247	0.219	2.570	0.166	0.497	0.006*
Min toe clearance DTE (%)	0.166	0.135	2.705	0.056	0.407	0.028*
Heel 3D path length DTE (%)	0.244	0.216	2.575	0.222	0.494	0.006*
DT Stride time variability (%)	0.031	-0.004	2.914	-0.096	-0.177	0.357
DT Stride length variability (%)	0.097	0.064	2.813	-0.271	-0.312	0.099
DT Speed variability (%)	0.086	0.052	2.831	-0.181	-0.293	0.123
DT Liftoff angle variability (%)	0.124	0.092	2.771	-0.163	-0.353	0.061
DT Heel trajectory forward MAD (mm)	0.093	0.059	2.820	-0.096	-0.305	0.108
DT Toe trajectory forward MAD (mm)	0.064	0.029	2.865	-0.081	-0.252	0.187
DT Heel 3D path length variability (%)	0.163	0.132	2.708	-0.326	-0.404	0.030*
DT Toe 3D path length variability (%)	0.153	0.121	2.725	-0.356	-0.391	0.036*
DT Loading ratio asymmetry (%)	0.180	0.150	2.681	-0.157	-0.425	0.022*
DT Liftoff angle asymmetry (%)	0.067	0.032	2.861	-0.099	-0.258	0.177

* $p < 0.05$, bold values indicate significance. DTE: Dual-task effect; DT: Dual-task; MAD: Mean Absolute Deviation; R²: Coefficient of determination; Adj. R²: Adjusted R²; SE: Standard error of the estimate; B: unstandardized coefficients; β : standardized coefficients.

Chapter 8

A Machine Learning Approach for the Detection of Mild Cognitive Impairment

This chapter is based on the following work:

V. Guimarães, I. Sousa, R. Cunha, R. Magalhães, Á. Machado, V. Fernandes, S. Reis, and M. V. Correia, “One-Class Classification with Confound Control for Cognitive Screening in Older Adults Using Gait, Fingertapping, Cognitive, and Dual Tasks,” submitted for publication.

Early detection of cognitive impairment is crucial for timely clinical interventions aimed at delaying the progression to dementia. In this study, we investigate the use of gait, fingertapping, cognitive, and dual tasks to detect mild cognitive impairment in an imbalanced dataset by employing one-class classification. Models are constructed based on the behavior of the majority group (healthy controls) to establish a baseline for normal behavior; instances that deviate from this baseline are considered abnormal. Confounder effects are addressed through confound regression embedded in the classification pipeline. Our results show that one-class classification has a higher predictive accuracy for mild cognitive impairment, whereas two-class classification performs better in identifying healthy controls. The combination of individual models leads to better performance than the combination of the features. Notably, the one-class majority voting approach exhibits a sensitivity of 87.5% and a specificity of 75.7%, which outperforms the performance of a standard cognitive screening test. In contrast, two-class majority voting fails to recover the low sensitivities achieved by the individual models due to the low representation of the impaired group. Our results thus support the use of one-class classification with confound control to detect abnormal patterns of gait, fingertapping, cognitive, and dual tasks, to improve the early detection of cognitive impairment.

8.1 Introduction

The detection of MCI using machine learning methods is usually framed as a supervised classification task, in which a model is built to differentiate between normal and abnormal cases using labeled training data derived from other normal and abnormal cases. Several recent works have explored the use of machine learning, most commonly using Support Vector Machines (SVM), to detect CI by analyzing gait characteristics [381]. Examples of such studies can be found in [225], [229], and [232], which achieved sensitivities of 82.0%, 87.0%, and 83.3%, and specificities of 67.7%, 85.0%, and 56.7%, respectively. In another study, a recurrent neural network was used to evaluate sequences of gait features, achieving a sensitivity of 97.5% and an F1-score of 97.4% in detecting a higher risk of CI [230]. Even though most of these studies present imbalanced training data (e.g., [225], [229], and [230]), only [230] accounts for dataset imbalance using under- and over-sampling techniques.

When faced with highly imbalanced datasets, traditional binary classification methods may struggle to accurately model and detect instances of the minority class due to their bias towards the majority class. One approach to addressing this issue is to use one-class classification (OCC), which constructs models using only the data from one class [382, 383]. Instances that deviate from the learned representation are regarded as anomalies, or novelties, and are classified as such. Experiments using synthetic and real-world data demonstrate that binary classifiers' performance decreases as data imbalance increases, while OCCs' performance remains relatively stable [382]. OCC has been effectively used in several real-world problems where data are only abundant for one class [383], such as fraud detection [384], fault detection [385], and medical diagnosis [386–391]. It should be noted that the use of normative data as a reference against which an individual's performance can be compared, typically employed in statistical analysis, is an established approach for detecting cognitive disorders in clinical and cognitive neuroscience [388]. Yet, the potential of OCC-based approaches for detecting MCI through gait, fingertapping (FT), and dual-tasking has not yet been explored.

Another frequent problem with existing machine learning models is the lack of confound control. Baseline characteristics of the subjects, such as age, education, or height may influence the characteristics of the features derived from gait and/or cognitive tests [162, 378]. Thus, when confounding effects are not appropriately addressed, biases related to demographic groups are likely reflected in machine learning models [392]. Although most studies on the detection of CI report differences between groups concerning age and education characteristics (e.g., as in [229], [232], and [230]), none of these studies take measures to account for the effects of these variables on the features, which constitutes a limitation. Despite being commonly used in statistics, the incorporation of confound adjustment in machine learning experiments is still uncommon and understudied [392].

This study aimed to explore OCC for MCI screening using data from gait, FT, cognitive, and dual tasks. For that purpose, we collected data from older adults from two groups: i) persons without MCI diagnosis and with or without cognitive complaints (control group), and ii) persons

with confirmed MCI diagnosis. The data collected comprised of inertial sensor data from gait, smartphone touch events from FT, the answers from a cognitive task, and the data from the dual tasks. To account for the minority class represented by MCI patients, we proposed modeling the majority group (the control group) and assessing whether the data from MCI patients were correctly classified as anomalies. We compared the performance of the different tasks, as well as the combination of features (early fusion) and models (late fusion). Moreover, we introduced feature adjustment based on confound control during pipeline training in cross-validation, allowing for the adjustment based solely on training data from the control group. We hypothesized that: i) gait, FT, cognitive, and dual-task performance metrics in healthy subjects are homogeneous enough to allow for the definition of a normal representation; ii) this representation is altered in MCI patients; iii) the amount of departure from the decision boundary is related to the severity of the CI as measured by a standard cognitive screening test; iv) the performance of the OCC surpasses that of the standard two-class classification (TCC); and v) dual-task performance metrics lead to better classification performances than single-task metrics. The performance of the machine learning models was compared with that of a standard cognitive screening test.

8.2 Methods

8.2.1 Participants

Data were collected from older adults with MCI (amnesic and non-amnesic), and healthy controls. The control group included older adults with subjective complaints (SC) without an MCI diagnosis (confirmed by neuropsychological tests) and older adults without any known cognitive changes. Participants with MCI and SC were recruited from a pool of individuals being followed at the neurology service of Braga Hospital. Healthy controls were recruited from the community through local contact persons and public advertisements promoted by the Clinical Academic Center (2CA-Braga), Braga, Portugal, and the School of Medicine from the University of Minho.

Participants were included if they were (i) aged 65+ years old, (ii) able to understand simple instructions, (iii) able to walk without support for at least 20 m, and iv) able to understand the objectives of the study and provide their informed consent. Persons were excluded from the study if they had any of the following exclusion criteria: (i) mobility impairment preventing autonomous walking, (ii) diagnosis of moderate or major CI (dementia), (iii) history of diseases that affect walking (e.g., Parkinson's diseases, orthopedic diseases, severe depression, etc.), iv) use of psychotropic medication, except antidepressants, short-acting benzodiazepines, and benzodiazepine derivatives, or v) inability to understand the study and provide informed consent.

The study was conducted in compliance with the national and international guidelines regarding clinical research, particularly, the Portuguese Law 21/2014 (changed by Law 73/2015 and Law 49/2018), EU Regulation 536/2014, and the ICH E6 (R2) Guidelines for Good Clinical Practice. The study was also developed in accordance with the national and international guidelines regarding data protection, namely, the Portuguese Law 58/2019 and the EU Regulation 2016/679.

The study was approved by the Data Protection Officer and Ethics Committee for Health of Braga Hospital (approval number 122-2021). Data collection was performed after obtaining informed consent.

8.2.2 Protocol

Data were acquired in a single visit to the study site, lasting around 40 minutes, that comprised the following procedures: i) demographics and health questionnaire, ii) Montreal Cognitive Assessment (MoCA), iii) Short Physical Performance Battery (SPPB), and iv) smartphone-guided tests.

8.2.2.1 Demographics and Health Data

Demographic data (age, sex, school years) and health (weight, height, physical activity, medication, falls in the last 12 months, fear of falling, and gait aids) were assessed using a questionnaire.

8.2.2.2 MoCA

The MoCA is a brief cognitive screening instrument, comprised of six subscores: i) executive function; ii) memory; iii) language; iv) visuospatial ability; v) orientation; and vi) attention, concentration, and working memory [166]. The Clinical Study Psychologists of 2CA-Braga administered the MoCA to all participants. Obtained MoCA scores were compared with age- and education-specific norms for the Portuguese population, as presented in [393]. Scores below 1.5 SD were considered indicative of CI according to test guidelines [393].

8.2.2.3 SPPB

The SPPB evaluates functional physical performance (lower limbs) by combining the results of gait speed, chair stand, and balance tests [329, 330]. Scores lower than 10 out of 12 indicate the presence of mobility limitations [326].

8.2.2.4 Smartphone-Guided Tasks

Smartphone-guided tasks consisted of a cognitive task, a walking task, FT, and dual tasks, which were performed with the support of two smartphones (Samsung Galaxy A52), a microphone, and two wearable inertial sensors [268]. One of the smartphones was used by the clinical study researcher to control the tests' execution. The other smartphone was connected to the microphone (by cable) and two inertial sensors (by Bluetooth) that were used by the participant to collect tests' performance data (Figure 8.1). The order of the three tests (when as single tasks) was randomized.

Cognitive Task: This task consisted of listing aloud for 20 s the result of serial subtractions, 7 by 7, starting from a number between 200 and 250, randomly generated by the smartphone app. During the test, audio was recorded using an external microphone connected to the smartphone.

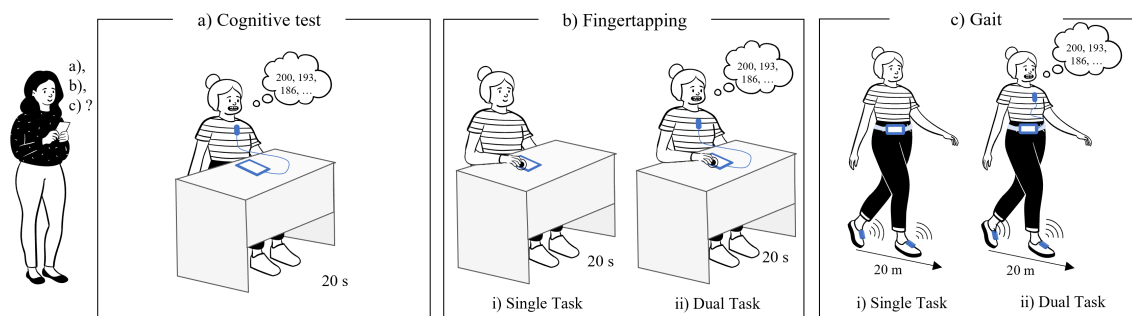


Figure 8.1: Data collection procedure: a clinical study researcher explains and initializes the tests (a, b, or c) according to the random order presented by the smartphone; the other smartphone is connected to an external microphone and two inertial sensors (placed on foot dorsum) to collect tests' performance data.

Gait: This task consisted of walking, at a comfortable speed, along 20 m in a straight line. During this task, data were collected from the inertial sensors—accelerometer and gyroscope—placed on both feet and from the smartphone placed at the waist (Figure 8.1), using a sampling rate of 100 Hz. The test's start and end were controlled by the clinical researcher. After completing this task (as a single task), participants were asked to repeat it while simultaneously performing the cognitive task (dual task). No prioritization instruction was given regarding the performance of the two tasks.

Fingertapping: The FT consists of the execution of successive touches on the smartphone screen, at a rhythm of the participant's choice, for 20 s. To perform the test, the participants should sit down and rest their forearms on a table. Screen touches should be executed using the index finger of the dominant hand. After completing the task, participants were asked to repeat it while simultaneously performing the cognitive task (dual task). No prioritization instruction was given regarding the performance of the two tasks.

8.2.3 Feature Extraction

Features were extracted from each smartphone-guided task as presented in Table 8.1. The same features were extracted from single- and dual-task conditions. Moreover, we have calculated the dual-task effect (DTE) by comparing a single-task metric (ST) with its equivalent in dual-task (DT), as follows:

$$DTE = DT - ST \quad (8.1)$$

Details on how features were extracted from each task are provided next.

8.2.3.1 Cognitive Task

Audio recordings were manually transcribed by annotating the numbers spoken by the participants. From the annotated text files, we extracted the number of correct and incorrect answers. Additionally, we separated subtractions into two groups: basic subtractions (e.g., 107-7, 108-7,

Table 8.1: Features extracted from smartphone-guided tasks.

Task	Features extracted
Cog	number of correct answers per second (in regrouping, basic, and all answers); number of errors per number of answers (in regrouping, basic, and all answers).
Gait	number of strides; test time; MAD of heel and toe trajectories (x, y, z); AVG, ASYM, MIN, MAX, IQR, CoV, SD and MAD of stride time, swing time, stance time, stride length, gait speed, swing width, strike angle, liftoff angle, cadence, stance ratio, swing ratio, foot-flat ratio, loading ratio, pushing ratio, double support ratio, min toe clearance, 3D length of heel and toe trajectories.
FT	number of taps; touch dispersion; AVG and SD of time between contacts and contact time.

AVG: average, CoV: coefficient of variation, ASYM: asymmetry, MAD: mean absolute deviation, MIN: minimum, MAX: maximum, IQR: interquartile range, SD: standard deviation, FT: fingertapping.

109-7) and regrouping subtractions (e.g., 100-7, 101-7, 102-7, etc.) [394, 395]. We have reported the number of correct answers (overall, basic, and regrouping) divided by the total test time and the number of errors (overall, basic, and regrouping) divided by the total number of answers. In case no answers were provided, the error rate was set at 100%. Metrics were extracted from text files using a Python routine.

8.2.3.2 Gait

Raw data from foot-worn inertial sensors (accelerometer and gyroscope) were automatically processed to extract spatiotemporal gait parameters, using the method described in Chapter 3. Moreover, foot trajectory-related features and minimum toe clearance were determined relying on the method presented in Chapter 4. The metrics were extracted from the periods of steady-state walking after excluding the first and last two strides [44]. In addition to the calculation of the average (AVG), we have also extracted the minimum (MIN), the maximum (MAX), the interquartile range (IQR), and the standard deviation (SD). Moreover, we have calculated the symmetry index (SI) [82], the coefficient of variation (CoV) [80], and the mean absolute deviation (MAD) [396].

8.2.3.3 Fingertapping

The smartphone application registered touch and release events, including timings and screen positions. From these data, we have calculated the number of touch events, the average and standard deviation of the time between touch events, and the average and standard deviation of the time in contact with the screen. Additionally, we have calculated touch dispersion as an average of the distances between all touch position pairs (Table 8.1).

8.2.4 Classification Pipeline

Features extracted in the previous step were used to develop machine learning models for the classification of MCI. Since the extracted features contained no missing values, no measures were taken to account for missing data. The classification pipeline included the following sequential

steps: confound control, data standardization, dimensionality reduction, sampling, and classification. The pipeline was implemented in Python 3 using the scikit-learn v1.1.3 and the imbalanced-learn v0.9.1 packages. Details are presented next for all the steps in the pipeline.

8.2.4.1 Confound Control

The literature offers several strategies for addressing confounders, including confound regression [392, 397], sample matching [398], sample weighting [399], and post-hoc corrections [400]. Among these, confound regression is one of the most commonly used techniques [392, 401], and, thus, it was adopted in this study. To implement this method, a linear regression model was constructed using the confounders as independent variables and the features as the dependent variable. After estimating the regression coefficients, the resulting residuals were utilized as inputs to the prediction models instead of the original features. According to [392], feeding confounding-adjusted variables to machine learning experiments relying on the entire dataset may result in leakage, as models are created using information beyond the training set. Thus, in a cross-validated setting, the coefficients need to be estimated using only the training data of each split [401]. To avoid removing disease-related confounders-features interactions, only control subjects in the training set were used to determine regression coefficients, as indicated in [397].

8.2.4.2 Standardization

Scaling the input features is crucial for certain steps in the pipeline, such as dimensionality reduction and classification. Therefore, prior to applying these steps, the data were standardized to have zero mean and unit variance. The means and standard deviations required to standardize the data were estimated from the training set.

8.2.4.3 Dimensionality Reduction

High data dimensionality can hinder learning in classification problems [385]. Therefore, dimensionality reduction techniques are usually required. In this work, we used the Principal Component Analysis (PCA), which transforms a set of variables into a new set of uncorrelated variables based on linear projections [385]. Since PCA is an unsupervised method, it is well-suited to the context of learning from only one class. Besides PCA, we also tested the Intra-Class Distance (ICD) as a feature selection method for the OCC. ICD measures the mean Euclidean distance from all training samples in one class to the centroid of the class. Lower ICDs thus indicate that samples in the class are closer to each other. The reduction of ICD when one feature is removed is used as a measure of its importance, being lower reductions qualified as higher feature rankings [402, 403]. When both classes were used for training, we used as an alternative to the ICD the mutual information (MI) feature selection [404].

8.2.4.4 Sampling

Over- and under-sampling are common strategies used to deal with the class imbalance in the setting of two-class classification (TCC). In this work, we tested the use of the Synthetic Minority Oversampling Technique (SMOTE) algorithm to generate synthetic examples of the minority class [405]. SMOTE was applied after feature selection, as recommended in [405], and was only applied during training. We also evaluated the performance of the TCC without resampling; in this case, the adjustments for class imbalance were introduced in the classification algorithm via class weighting.

8.2.4.5 Classification

OCC can be achieved using density-based, reconstruction-based, or boundary-based methods [383, 386]. Since large training sets are usually required in density- and reconstruction-based methods [386], we opted for using a boundary-based approach, the One-Class SVM (OC-SVM). OC-SVM can be viewed as a standard two-class SVM (TC-SVM) where all the training data lie in the first class, and only the origin is taken as a member of the second class [385, 406, 407]. The TC-SVM learns the decision boundary (linear hyperplane) that best separates the two classes in a non-linearly transformed feature space [387]. Details about both methods can be found in [384], [406], and [390].

In this work, the OC-SVM was learned only from the control group, and MCI samples were regarded as anomalies/outliers. Thus, examples falling inside the boundary were classified as normal, whereas examples outside the boundary were classified as MCI. The OC-SVM uses the hyperparameter ν to control for the trade-off between the margin size and the number of training points that are allowed to be inside the margin [387, 389, 406]. In addition, we tested the standard TC-SVM that used the hyperparameter C to control for the strength of the regularization. To map the training samples to a higher-dimensional space we used the Gaussian kernel, also called the radial basis function (RBF). RBF is a non-linear kernel whose width is controlled by hyperparameter γ [386, 387].

8.2.5 Evaluation

We used nested cross-validation (CV) to optimize pipeline hyperparameters and evaluate the performance of the resulting models, as illustrated in Figure 8.2. The selection of the hyperparameters was conducted in the inner loop resorting to a 5-fold grid search CV. The best-performing hyperparameters were selected based on the average balanced accuracy obtained in the validation folds. The whole outer training set was then used to fit the pipeline. Average balanced accuracy, sensitivity, specificity, F1-scores, and area under the receiver operating characteristic curve (AUC) were determined by the performances achieved in the test folds of the outer loop. Table 8.2 shows the hyperparameters tested in the grid search.

The CVs were stratified to maintain consistent group proportions across all splits. However, for OCC, only the control group was included in both the inner and outer training sets. Although

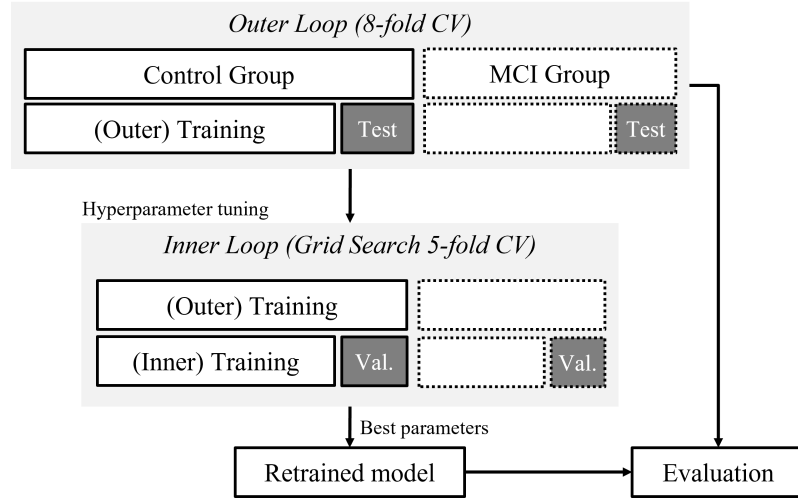


Figure 8.2: Nested cross-validation for OCC. In TCC, the blank boxes from the MCI group are also used for (outer and inner) training.

the MCI group was also divided, the partition allocated for training was not used as input for the OCC; these partitions were only used for the TCC. The selection of the best-performing models, feature selectors, dimensionality reduction methods, and sampling strategies (in the TCC) was based on the average balanced accuracy obtained on test partitions, which included samples from both groups.

Besides comparing the performance of OC-SVM and standard TC-SVM, we have also compared the performance of the different smartphone-guided tasks. For each individual task, we have assessed the performance of i) using ST measures only, ii) using DT measures only, iii) using DTEs only, or iv) combining ST, DT, and DTE measures. Only the performance of the best-performing feature set was reported for each included task. In addition, we evaluated the effectiveness of combining features from distinct tasks in an early fusion approach that combined gait and cognitive tasks, FT and cognitive tasks, and all tasks.

Given that OC-SVM enables the computation of a distance to the decision boundary that measures the degree of abnormality from the training samples, we hypothesized that these distances would be correlated with MoCA scores. To test this hypothesis, we trained one of the best-performing pipelines on a subset of the control subjects, using the average hyperparameters derived from nested CV. We then tested the trained model on the remaining subjects. Retraining was required to ensure that predictions could be compared to each other as they were all computed in relation to the same classifier and decision boundary [388]. We computed Spearman correlations (and the corresponding p-values) between the distances to the decision boundary and MoCA scores (adjusted for age and education based on Portuguese normative data). Since the choice of the training subset can impact the results, we tested our hypothesis on 500 different training subsets that were randomly chosen.

Finally, we have compared the effectiveness of feature combination (early fusion) with that of model combination (late fusion). Model combination entails fusing the output of multiple

Table 8.2: Hyperparameters search space.

Methods	Hyperparameters	Search Values
PCA	Explained variance	[0.75, 0.85, 0.95]
ICD and MI	k	[3, 8, 16]
OC-SVM	γ	[0.001, 0.01, 0.1, 1, 10, 100, 1000]
	ν	[0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5]
TC-SVM	γ	[0.001, 0.01, 0.1, 1, 10, 100, 1000]
	C	[0.001, 0.01, 0.1, 1, 10, 100, 1000]

classifiers to generate a final prediction. To perform this comparison, we employed a majority voting approach, which involves combining the outputs of individual models based on a majority vote. The performances of the majority voting were obtained after pooling the predictions of each CV split into a single set [408]. The resulting performances were also compared with the performances obtained with MoCA considering the normative data of the Portuguese population as a reference.

8.2.6 Statistical Analysis

Demographics and health data were summarized using either means and standard deviations, or frequencies and percentages, as appropriate. The comparisons between control and MCI groups were performed using the Independent Samples T-test or the Mann-Whitney U test for continuous variables, and Pearson's chi-square test (χ^2) for categorical variables. The Mann-Whitney U test was used when data were not normally distributed or when outliers were present outside ± 1.5 of the IQR range. The data were tested for normality using the Shapiro-Wilk test.

To assess differences between different machine learning experiments by comparing the performances achieved in nested CV, we used Paired Samples T-tests or Wilcoxon signed-rank tests. Wilcoxon signed-rank tests were used when paired data differences were not normally distributed or when outliers were present. The McNemar's Test [409] was used to compare pooled machine learning predictions and/or classifications according to MoCA.

Statistical analysis was performed resorting to the packages statsmodels v0.13.2, mlxtend v0.21.0, SciPy v1.10.1, and NumPy v1.24.2 for Python 3. A p-value of less than 0.05 indicated statistical significance.

8.3 Results

8.3.1 Descriptive Statistics

From March 2022 to March 2023, data were collected from 38 healthy controls, and 16 persons with MCI. One participant in the control group was excluded from the analysis due to unreliable sensor data, resulting in 37 healthy controls and 16 persons with MCI. The dataset was highly imbalanced, with the MCI group representing less than half of the control group. Descriptive statistics of the included participants are shown in Table 8.3.

Table 8.3: Descriptive statistics (n=53).

Characteristic	Control (n=37 [†])	MCI (n=16 [‡])	p-value
Age [years]	72.59 ± 4.95	75.88 ± 5.49	0.041*
Height [cm]	1.59 ± 0.08	1.59 ± 0.08	0.816
Weight [Kg]	77.54 ± 17.39	68.62 ± 10.8	0.068
Female, n (%)	20 (54.1)	10 (62.5)	0.569
Education [years]	6.97 ± 3.22	5.5 ± 3.26	0.070
Nr. of medications	4.73 ± 2.65	4.5 ± 2.09	0.899
Has fear of falling, n (%)	21 (56.8)	8 (50.0)	0.650
Nr. of falls (last 12 months)	0.89 ± 2.51	0.81 ± 1.59	0.972
Does physical activity, n (%)	28 (75.7)	9 (56.2)	0.157
Physical activity [h/wk.]	5.09 ± 7.11	4.03 ± 5.48	0.622
SPPB	9.35 ± 1.36	8.38 ± 1.96	0.066
MoCA	18.89 ± 4.57	13.44 ± 4.85	< 0.001*

Data are mean values ± standard deviation or the number of participants per category (absolute and relative frequency) when indicated. Group differences were evaluated using the Independent Samples T-test, Mann-Whitney U test, or Pearson's chi-square. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance. [†]8 with subjective complaints. [‡]10 amnesic and 6 non-amnesic.

Age presented significant differences between the groups. As expected, MoCA scores were also significantly different. The average SPPB score was below 10 in both groups denoting possible mobility limitations [326]. Half of the participants reported having fear of falling.

8.3.2 Confound Control

Age, height, weight, sex, and education are potential confounders that may affect gait metrics and/or cognitive performance [162, 378]. Since age, education, and weight presented significant or nearly significant differences between the groups (Table 8.3), only these three variables were considered confounders in the classification pipeline.

Figure 8.3 shows the number of features in each of the smartphone-guided tasks that remained significant after confound control. As can be observed, the confound control strategy reduced the number of features presenting significant differences between control and MCI groups in all tasks. After confound adjustment, significant differences could be found in gait metrics such as stride length (IQR, ST), loading ratio (ASYM, ST), strike angle (IQR and ASYM, DT and DTE), foot-flat ratio (AVG, DT), pushing ratio (AVG, DT and MAX, DT and DTE), stride time (MIN, DTE), stance time (MIN, DTE), gait speed (IQR, DTE), and minimum toe clearance (ASYM, DTE). Most of these metrics were determined under DT conditions, or using DTEs. Concerning cognitive metrics, only the DTE of FT on the number of correct answers (basic) per second displayed significant differences between groups. As for the FT, only contact times (SD, ST) and time between contacts (SD, DT) presented significant differences between groups after confound control.

Despite the reduction in the number of significant features, the OC-SVM models were still able to effectively learn from the data after accounting for confounds. Figure 8.4 compares the best performances achieved by the models with and without confound control. Although the

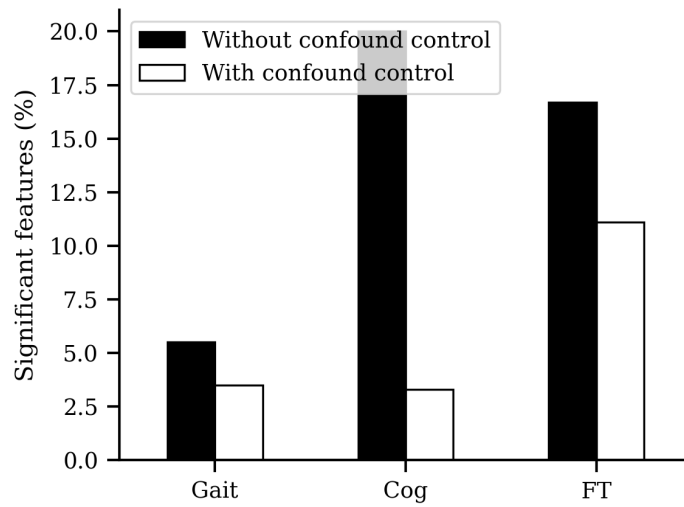


Figure 8.3: Proportion of features with significant differences between groups before and after confound regression for the gait, cognitive (Cog), and fingertapping (FT) tasks.

differences were not statistically significant in any of the presented conditions ($p > 0.05$), we did notice a slight decline in the models' performance after confound control, except for two of the conditions. To ensure control for confounders' effects, all subsequent machine learning experiments included confound regression.

8.3.3 One-Class Classification Performance

The results of the OCC, including the best-performing dimensionality reduction methods, are shown in Table 8.4. The best results were achieved with gait features (ST, DT, and DTE), resulting in a balanced accuracy of $72.2 \pm 12.8\%$, a sensitivity of $87.5 \pm 21.7\%$, and a specificity of $56.9 \pm 19.8\%$. The second best performing model resulted from the combination of FT and cognitive DTEs, with a balanced accuracy of $70.9 \pm 17.5\%$, a sensitivity of $81.2 \pm 24.2\%$, and a specificity of $60.6 \pm 27.2\%$. All conditions presented sensitivity values above specificity. The average number of features/components used by the models did not exceed 12.

8.3.4 Comparison with Two-Class Classification

The best results achieved with TCC, including the best-performing dimensionality reduction and sampling methods, are shown in Table 8.5. The best results were achieved with FT and cognitive features (DTEs), resulting in a balanced accuracy of $70.6 \pm 23.3\%$, a sensitivity of $62.5 \pm 41.5\%$ and a specificity of $78.8 \pm 17.5\%$. Contrarily to the OCC, all tested conditions in the TCC had specificity values above sensitivity.

Figure 8.5 shows an objective comparison of the balanced accuracy, sensitivity, and specificity achieved through OCC and TCC for each data source considered. Additionally, the figure compares the performance of TCC with and without data sampling using SMOTE. The statistical

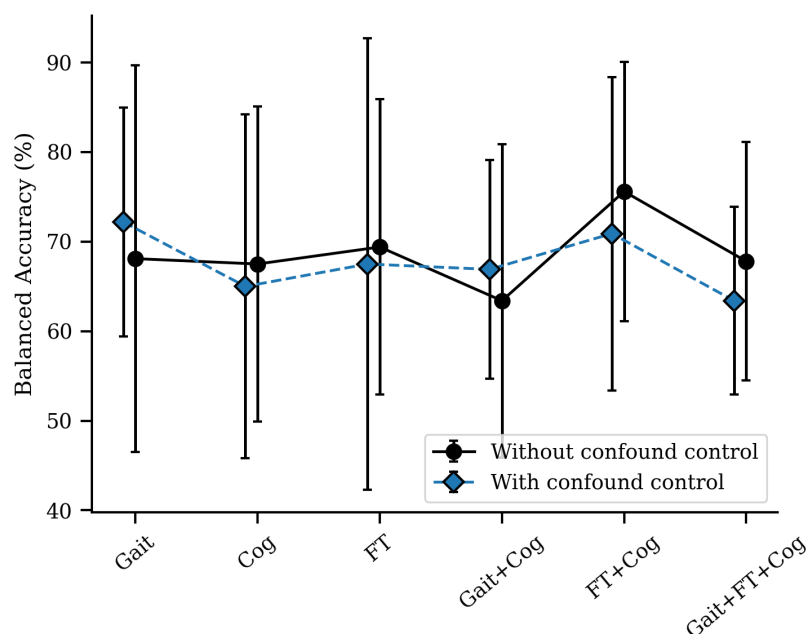


Figure 8.4: Comparison of the best OC-SVM results with and without confound control. Vertical lines represent standard deviation.

analysis focused on the differences between OCC and TCC (without sampling), and between using or not SMOTE in TCC. The results indicate that the differences in sampling strategies were not significant in any of the considered tests. Significant differences in sensitivity and specificity between OCC and TCC (without sampling) were observed for some of the test conditions that included gait features. However, for all experiments, sensitivity was higher in OCC than in TCC. Conversely, the specificity of the TCC was higher than that of the OCC.

8.3.5 Decision Boundary

As mentioned previously, the best results in nested CV using OC-SVM were obtained using gait features (ST, DT, and DTE, Table 8.4). Thus, the average hyperparameters obtained from this OC-SVM pipeline were used to retrain the model using a subset of control subjects. To evaluate the performance, 16 samples from the control group were set aside for testing, while 21 were used for training. The test group was thus comprised of 16 samples from the control group plus 16 samples from the MCI group. Besides classification performance, we evaluated the correlations between adjusted MoCA scores and the distances to the decision boundary. Figure 8.6 shows the distributions of performance, correlations, and p-values obtained from 500 random control group splits.

As can be observed, performance metrics exhibited some variability due to the different data splits. The MoCA sensitivity remained constant at 81.2% as the MCI group remained unchanged. The average performance achieved on the 500 runs (average balanced accuracy of $61.0 \pm 7.1\%$, sensitivity of $69.3 \pm 11.2\%$, and specificity of $52.8 \pm 15.3\%$) was lower than that obtained in the nested CV. The correlations with adjusted MoCA scores were non-significant with $p > 0.05$.

Table 8.4: Results of the One-Class SVM

	Gait	Cog	FT	Gait+Cog	FT+Cog	Gait+FT+Cog
Feature set	ST+DT+DTE	DT (with FT)	ST	ST+DT+DTE	DTE	ST+DT+DTE
Dim. Red.	PCA	ICD	ICD	PCA	None	ICD
Bal. Acc. (%)	72.2 ± 12.8	65.0 ± 19.2	67.5 ± 25.2	66.9 ± 12.2	70.9 ± 17.5	63.4 ± 10.5
Sens. (%)	87.5 ± 21.7	68.8 ± 34.8	68.8 ± 34.8	93.8 ± 16.5	81.2 ± 24.2	81.2 ± 24.2
Spec. (%)	56.9 ± 19.8	61.3 ± 22.2	66.2 ± 26.4	40.0 ± 17.9	60.6 ± 27.2	45.6 ± 22.8
F1 score (%)	61.3 ± 13.6	51.7 ± 27.3	56.9 ± 33.5	57.3 ± 12.4	61.7 ± 20.3	53.5 ± 11.1
AUC (%)	72.8 ± 18.3	69.1 ± 22.0	65.6 ± 33.1	73.1 ± 19.5	72.5 ± 19.0	60.3 ± 15.9
# Features	11.1 ± 4.6	3.0 ± 0.0	3.0 ± 0.0	12.0 ± 4.6	12.0 ± 0.0	8.1 ± 5.0

Table 8.5: Results of the Two-Class SVM

	Gait	Cog	FT	Gait+Cog	FT+Cog	Gait+FT+Cog
Feature set	DTE	DT (with FT)	DT	DTE	DTE	ST+DT+DTE
Dim. Red.	PCA	MI	PCA	PCA	PCA	None
Sampling	SMOTE	SMOTE	None	SMOTE	None	None
Bal. Acc. (%)	56.6 ± 18.3	67.2 ± 20.3	65.6 ± 20.4	58.4 ± 12.7	70.6 ± 23.3	61.9 ± 25.6
Sens. (%)	43.8 ± 30.0	62.5 ± 41.5	43.8 ± 39.0	37.5 ± 21.7	62.5 ± 41.5	43.8 ± 46.4
Spec. (%)	69.4 ± 24.0	71.9 ± 20.5	87.5 ± 17.1	79.4 ± 26.3	78.8 ± 17.5	80.0 ± 21.8
F1 score (%)	38.8 ± 26.6	50.5 ± 31.9	41.2 ± 38.9	38.3 ± 24.6	54.6 ± 33.0	39.2 ± 40.3
AUC (%)	61.3 ± 16.3	63.1 ± 32.8	61.6 ± 23.7	61.6 ± 27.9	75.0 ± 25.2	73.1 ± 31.0
# Features	13.1 ± 4.4	3.0 ± 0.0	4.5 ± 0.5	13.4 ± 4.6	5.1 ± 1.5	450.0 ± 0.0

Figure 8.7 shows the results of an optimist test partition achieving performances comparable to MoCA (in the same partition), with a balanced accuracy of 78.1%, a sensitivity of 81.2%, and a specificity of 75.0%. In this figure, positive values are classified as controls, whereas negative values are classified as MCI. Despite achieving equal performance, we observe that misclassifications occurred in different test subjects. The correlation between MoCA and the distances to the decision boundary was also not significant in this test partition.

8.3.6 Majority Voting

Table 8.6 compares the performance of MoCA with that of the majority voting approach, which combines the predictions of models trained on individual tasks, such as gait, cognition, and FT. Both one-class and two-class majority voting outperformed the combination of features (Gait + FT + Cog) as shown in Tables 8.4 and 8.5. The one-class majority voting had a higher balanced accuracy than the two-class majority voting. Additionally, the one-class majority voting was able to recover the lower specificities observed in Table 8.4, while the two-class majority voting was unable to recover the lower sensitivities observed in Table 8.5. Furthermore, all performance metrics of the one-class majority voting were superior to those obtained with MoCA. Despite these differences, McNemar tests revealed no significant differences in misclassifications (all with $p > 0.05$) when comparing pairwise the classifications of MoCA, the predictions from one-class majority voting, and the predictions from two-class majority voting.

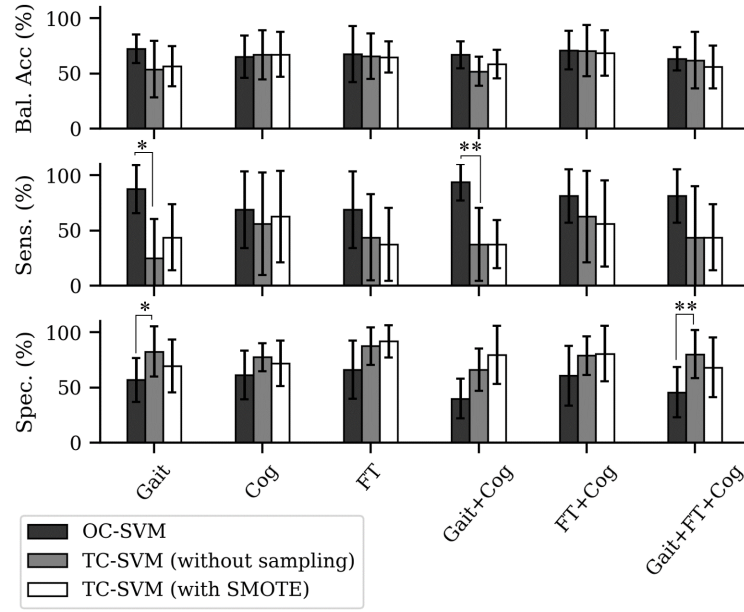


Figure 8.5: Comparison between OC-SVM and TC-SVM, with the latter being evaluated both with and without sampling with SMOTE (* $p < 0.05$, ** $p < 0.01$).

8.4 Discussion

In this work, we investigated the use of OCC for the detection of MCI relying on the analysis of gait, FT, and a cognitive task, in which confound control was implemented through confound regression. We compared OCC with the standard TCC approach and investigated whether the use of oversampling techniques, like SMOTE, could tackle the problem of class imbalance in our dataset. Moreover, we investigated whether an early fusion approach based on the combination of features would provide better results than a late fusion approach based on the combination of models' predictions. We compared our results with the classifications provided by MoCA.

In line with our initial hypothesis, gait, FT, cognitive, and dual-task performance metrics allowed the definition of a normality boundary and the classification of MCI with high sensitivity. The OC-SVM achieved better sensitivity than the TC-SVM, but specificity was higher in the TC-SVM. The use of oversampling techniques, like SMOTE, to increase the representation of the MCI group during training, did not allow to compensate for the low sensitivity in TC-SVM. According to previous studies, the OCC is more appropriate to use when there is a large class imbalance in the dataset [382, 383]. According to [382], TCC performance starts deteriorating when imbalances ratios over 1:2.5 are observed; according to [387], the benefit threshold is at about half the number of cases in the minority class. Our results are thus consistent with prior research, as an imbalance ratio of 1:2.3 was present in our dataset. Furthermore, the estimated prevalence of MCI in the population is 27.5% [410], indicating that the class imbalance is naturally present. Our results support the use of OC-SVM for cognitive screening, particularly when the primary objective is to detect all potential cases for further evaluation [237]. Across the reported experiments, DT metrics and DTEs were generally favored over ST metrics, which aligns with previous studies [52, 53].

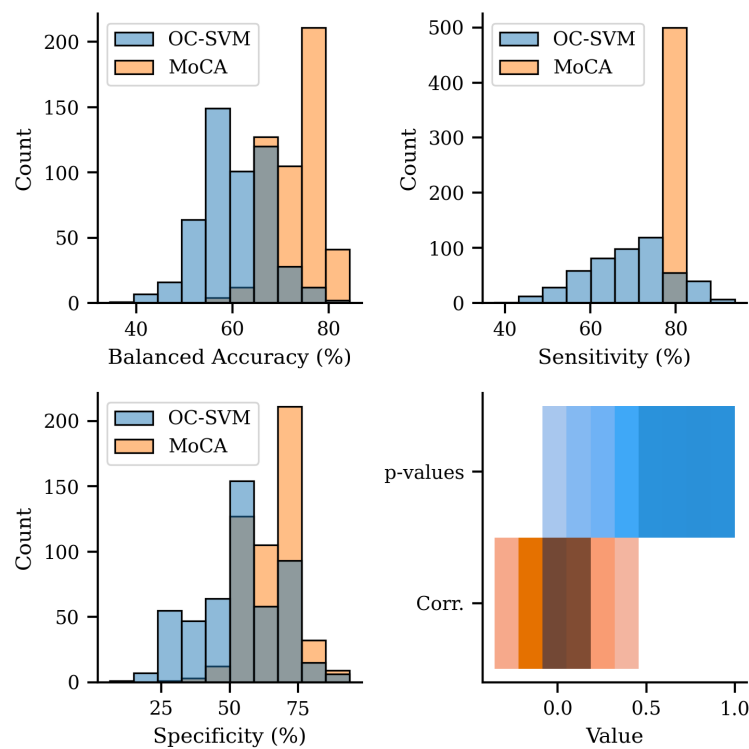


Figure 8.6: Distribution of performance metrics and correlations between distances to decision boundaries and MoCA scores.

Contrary to our expectations, the distances to the decision boundary in OC-SVM using gait features, which was one of the best-performing experiments, was unrelated to the severity of CI as assessed by MoCA. Although MoCA is recognized as one of the most reliable and rapid methods of cognitive screening [22, 173], it has limited discrimination power and its scores may not entirely reflect the extent of CI, especially in cases of milder impairment [22]. To gain a better understanding of the significance of decision boundaries compared to standard cognitive tests, further assessments involving a more diverse range of subjects with significant CI or dementia should be conducted.

Although the OC-SVM produced highly sensitive results for the detection of MCI, the experiments had low specificity in general. However, combining models using the majority voting approach compensated for the low specificity and resulted in better outcomes than the combination of features. It is worth noting that combining TC-SVM classifiers did not recover the low sensitivities of the models in the TCC, despite overall performance improvement over the combination of features in TC-SVM. Although misclassifications were not significantly different according to McNemar's Test, the performance of the OC-SVM majority voting approach surpassed that of MoCA in all performance metrics. Moreover, the reported performances were comparable or even superior to other studies employing SVMs [225], [229], and [232]. OC-SVM was thus regarded as a viable option for tackling the problem of cognitive screening in older adults using gait, FT, cognitive, and dual tasks. Compared to MoCA, these smartphone-guided tasks are simpler, faster, and more objective; their relevance in screening is highlighted by the fact that they don't require

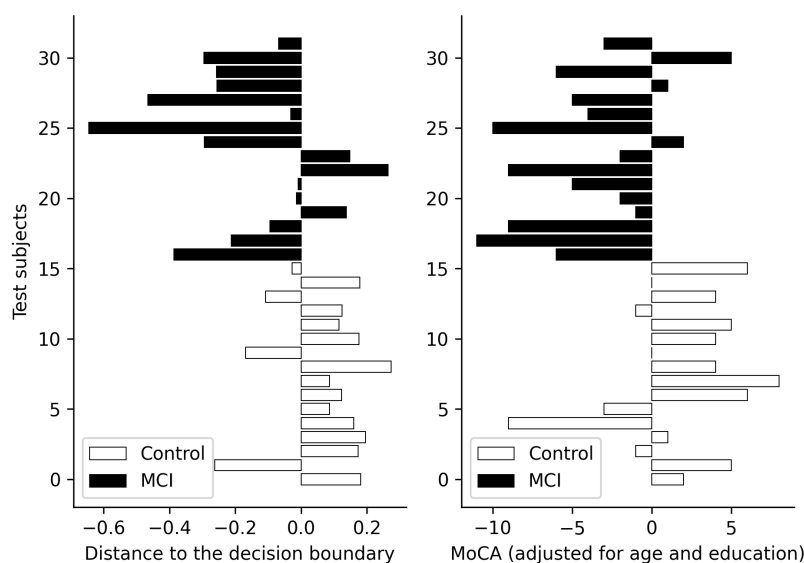


Figure 8.7: Test predictions in an optimistic dataset partition according to (*left*) the distance to the decision boundary obtained with OC-SVM and (*right*) adjusted MoCA scores.

scoring by experts [212].

8.4.1 Study Limitations

The primary limitation of our study was the low amount of available data, which was partially addressed by using a nested CV. However, the small sample size may affect the robustness of the models and their ability to generalize. Due to the limited data, we were unable to systematically evaluate the performance of OCC and TCC methods at consecutively higher imbalance ratios to determine the optimal point at which OCC would be most beneficial. Moreover, we were unable to investigate whether the subjective complaints group exhibited a different pattern than the normal group, or if differences in amnesic and non-amnesic groups could be identified using gait, FT, cognitive, and/or dual-tasks. Additionally, the effects of the confound adjustment could not be extensively tested. Although confound adjustment has been implemented as recommended by past studies (e.g. [392] and [397]), since a linear regressor was used to describe the relationship between confounders and features, it is possible that non-linear relationships still exist and were captured by the machine learning models [400]. To address this, non-linear regressors or adversarial learning approaches could be utilized to mitigate potential biases resulting from confounding [411, 412]. Further evaluation of the confound adjustment should be performed with additional data.

The cognitive task used in this study presents another limitation. While the literature supports the use of subtractions by sevens as a sufficiently challenging task to assess dual-task performance [52], it induced anxiety in many of our participants and was heavily dependent on individual skill [188]. Therefore, future studies should explore alternative cognitive tasks. Additionally, the participants of the control group that were recruited from the community were not formally assessed by the medical team from the Neurology Department of Braga Hospital using the clinical practice methods to evaluate CI. As cognitive assessments performed by medical doctors are

Table 8.6: Performance using MoCA and a majority voting approach.

Metric	MoCA (normative data)	One-Class Majority Voting (Gait+FT+Cog)	Two-Class Majority Voting (Gait+FT+Cog)
Bal. Acc. (%)	73.1	81.6	70.0
Sens. (%)	81.2	87.5	56.2
Spec. (%)	64.9	75.7	83.8
F1 score (%)	61.9	71.8	58.1

usually triggered by a subjective complaint, the participants in the control group were assumed to have normal cognition. Thus, from all the participants in the control group, only the participants with subjective complaints were confirmed not to have MCI by the clinical practice cognitive assessments.

8.4.2 Future Work

Besides collecting and analyzing results on additional data, further experiments could be performed. For example, as suggested by Jung et al. [230], incorporating sequential gait parameters and recurrent neural networks may improve classification results. Deep neural networks, such as generative adversarial networks, are also compatible with OCC and could provide a promising solution to modeling these types of problems [391,413]. Therefore, these methodologies should be explored in future research. Sequential patterns could also be analyzed with respect to FT performance and dual tasks. Additionally, we should compare the results achieved using gait parameters extracted from foot-worn inertial sensors with those from smartphone inertial sensors to determine which technology is most suitable for clinical use. Furthermore, alternative cognitive tasks and the differential patterns of amnesic and non-amnesic MCI sub-groups should be investigated. The differences between cognitively healthy and subjective complaints participants should also be investigated after an objective assessment of the cognitively healthy group. Finally, the effects of the confound regression should be further evaluated by comparing the performance of the models in subcohorts formed according to participants' baseline characteristics.

8.5 Conclusion

In this study, we evaluated the use of OCC for the detection of MCI using gait, FT, cognitive, and dual tasks, while accounting for confounder effects using confound regression. Our findings demonstrate that OCC is a promising approach for cognitive screening in imbalanced datasets, as it provides higher sensitivities compared to TCC. Moreover, we observed that a majority voting approach combining OCC models outperforms feature combination, resulting in performance metrics superior to those achieved with the MoCA. In contrast, the TCC yields higher specificities than sensitivities, which cannot be improved by oversampling. Although the use of one-class majority voting enables specificity recovery, two-class majority voting fails to recover sensitivity. These results suggest that OCC can improve the detection of MCI in highly imbalanced data. The

proposed tasks are simple to administer, objective, and do not require expert scoring, making the method more feasible in a clinical setting (including its use as a screening method at the primary healthcare level for a more efficient selection of the persons that should be evaluated and followed by neurologists). However, further research is needed to confirm the feasibility and effectiveness of the OCC approach in clinical practice.

Chapter 9

Detection of Cognitive Impairment in Post-Stroke Patients

This chapter is based on the following work:

V. Guimarães, I. Sousa, and M. V. Correia, “Detecting Cognitive Impairment in Cerebrovascular Disease Using Gait, Dual Tasks, and Machine Learning,” submitted for publication.

Cognitive impairment is common after a stroke, but it can often go undetected. In this study, we investigated whether using gait and dual tasks could help detect cognitive impairment after stroke.

We analyzed gait and neuropsychological data from 47 participants who were part of the Ontario Neurodegenerative Disease Research Initiative. Based on neuropsychological criteria, participants were categorized as impaired ($n=29$) or cognitively normal ($n=18$). Nested cross-validation was used for model training, hyperparameter tuning, and evaluation. Grid search with cross-validation was used to optimize the hyperparameters of a set of feature selectors and classifiers. Different gait tests were assessed separately.

The best classification performance was achieved using a comprehensive set of gait metrics, measured by the electronic walkway, that included dual-task costs while performing subtractions by ones. Using a Support Vector Machine (SVM), we could achieve a sensitivity of 96.6% and a specificity of 61.1%. An optimized threshold of 27 in the Montreal Cognitive Assessment (MoCA) revealed lower classification performance than the gait metrics, although differences in classification results were not significant. Combining the classifications provided by MoCA with those provided by gait metrics in a majority voting approach resulted in a higher specificity of 72.2% and a high sensitivity of 93.1%.

Overall, our results suggest that gait analysis can be a useful tool for detecting cognitive impairment in patients with cerebrovascular disease, serving as a suitable alternative or complement to MoCA in the screening for cognitive impairment.

9.1 Introduction

Cognitive impairment (CI) due to cerebrovascular disease (CVD), also known as vascular cognitive impairment (VCI) is common after stroke. It affects mostly attention and executive functions [158,414], ranging from mild to major CI or dementia [158,415,416]. Vascular dementia, which is the second most common cause of dementia after Alzheimer's disease [157–159], can significantly impact patients' quality of life and independence [157,417], being associated with increased mortality, disability, and institutionalization [237].

Despite significant progress in stroke care in recent decades, CI after stroke remains a common and underdiagnosed problem [417–419]. Studies estimate that 30 to 50% of stroke survivors experience CI [417], but recent systematic reviews suggest that the prevalence may be even higher, with more than half of stroke survivors experiencing some degree of CI [420]. In the first year after a stroke, the prevalence of mild CI ranges from 17.5% to 54.9%, with a pooled prevalence of 38% [421]. As the population ages and the number of stroke survivors increases, effective management of post-stroke CI will become increasingly important for both patients and caregivers [418].

The first step in the management of CI is diagnosis. The gold standard for diagnosis is a comprehensive neuropsychological examination [418] that is not feasible at a population level. In practice, a two-step approach is advised, starting with a primary screening test such as the Montreal Cognitive Assessment (MoCA) [237,416]. While MoCA is commonly used for this purpose, its low specificity, and moderate to good sensitivity for the diagnosis of post-stroke CI coupled with a lack of standardization still limit its use [159,418].

Researchers have been exploring alternative methods to improve the screening of CI. Gait analysis, for instance, was shown to distinguish normal aging from mild CI in non-stroke patients [52]. Further research showed that gait performance is affected by the cognitive load in people with CI due to cognitive-motor, or, dual-task, interference [51,422]. A recent meta-analysis showed that dual-task gait characteristics allow better differentiation between groups, with an increased sensitivity for the detection of mild CI [52].

Because attention is often impaired, patients with VCI may struggle with multitasking [418]. Studies suggest that performance on dual tasks can be compromised in patients with VCI [423], and suggest that gait assessment under dual-task conditions could be a useful tool for early detection of VCI [424]. Dual-task assessment is attractive because it is less influenced by educational level, is practical, fast, and easy to administer in clinical practice [212], and may be a good alternative to screening tests like MoCA.

Digital biomarker technologies, often combined with machine learning-based predictive models, are increasingly explored for early detection of CI in community settings [381]. Support Vector Machines (SVMs) are often used for this purpose by analyzing gait characteristics [381]. For example, Boettcher et al. [225] used dual-task gait data collected from an electronic walkway to differentiate subjects with mild CI from healthy individuals, achieving an accuracy of 77.2%. Aoki et al. [212] used the Microsoft Kinect sensor to capture dual-task gait, achieving an AUC of 74.7% in differentiating subjects with different levels of cognitive performance, as measured by

the mini-mental state examination (MMSE). Ghoraani et al. [229] employed an SVM in a one-vs-one manner with a majority vote to differentiate healthy subjects, older adults with mild CI, and older adults with Alzheimer’s disease. An accuracy of 86.0% was achieved in the differentiation of cognitively impaired and healthy subjects using gait features extracted from an electronic walkway. Shahzad et al. [232] used an inertial sensor-based gait analysis and a machine learning framework to distinguish mild CI due to Alzheimer’s disease from healthy subjects, finding that dual-task walking provided a better distinction than single-task, with a classification accuracy of 70.0%. Finally, Jung et al. [230] used sequential gait characteristics extracted from inertial sensors coupled with a long short-term memory network to categorize community-dwelling older adults into three groups based on their scores of the MMSE; they achieved improved F1 scores of 97.4% using temporal gait features extracted from usual and fast pace walking trials. Taken together, these studies suggest that machine learning and gait assessments have the potential to be objective tools for cognitive screening that do not heavily rely on cognitive testing.

Although some research has investigated the use of gait characteristics to detect CI, it is unclear whether machine learning and gait assessments can effectively distinguish CI from cognitively normal in post-stroke patients. In this study, we aim to objectively evaluate the clinical utility of gait in classifying CI in post-stroke patients using a machine learning-based approach. We used as input data from different gait tests, including dual tasks and fast walking conditions. We also tested multiple machine learning classifiers and two different gait assessment technologies that provided varying levels of detail about gait metrics. Finally, we objectively compared the classification results with those obtained using MoCA and proposed a majority voting approach to combine MoCA with gait. We hypothesized that: i) machine learning models based on gait metrics differentiate VCI from healthy individuals with a performance comparable to MoCA; ii) the dual-tasks provide better classification results than the single-tasks; iii) a more comprehensive analysis of gait allows better differentiation of the groups than simpler gait metrics; and iv) the combination of MoCA with gait metrics leads to better performances than using each modality alone.

The structure of this chapter is as follows: in Section 9.2 we describe our methodological approach; in Section 9.3 we present the results of our experiments; in Section 9.4 we discuss the results, limitations, and future prospects for this work; and in Section 9.5 we provide the main conclusions.

9.2 Materials and Methods

9.2.1 Dataset

The data used in this study were obtained from the Ontario Neurodegenerative Disease Research Initiative (ONDRI) [425, 426]. ONDRI is a multi-site prospective cohort study investigating Alzheimer’s dementia and mild CI (AD/MCI), amyotrophic lateral sclerosis (ALS), frontotemporal disorders (FTD), Parkinson’s disease (PD), and cerebrovascular disease (CVD). The ONDRI

protocol includes assessments for genomics, neuroimaging, ocular function, gait and balance, and neuropsychological testing. In the present study, we have only used the baseline data from the cerebrovascular disease (CVD) cohort, including demographic, clinical, neuropsychological, and gait data.

Participants in the CVD group experienced a mild to moderate ischemic (or silent) stroke event verified on neuroimaging 3 or more months before enrollment. They also met the following inclusion criteria: a) age between 55 and 85 years old, b) proficient in speaking and understanding English, c) 8 or more years of formal education, d) mild-moderate stroke severity defined by scores 0-3 in the modified Rankin scale, e) a MoCA score of at least 18. Exclusion criteria included vascular cause of symptoms, large cortical strokes, severe CI, aphasia, history of dementia prior to the stroke, inability to write, and/or severe functional disability limiting the ability to perform assessments [425,427,428].

The current study received approval from the Ethics Committee of the Faculty of Medicine of the University of Porto (51/CEFMUP/2022). Access to the ONDRI Foundational Study Data was granted after approval by Brain-CODE's Data Access Committee [429].

9.2.2 Assessments

9.2.2.1 Demographic and Clinical Data

The dataset included demographic data, such as age, education (reported in years), height, weight, and leg length. Additionally, it included a rate of depression, as assessed by the Quick Inventory of Depressive Symptomatology (QIDS) [430].

Study partners rated the participant's ability to function independently across eight instrumental activities of daily living (iADLs; i.e., telephone use, shopping, food preparation, housekeeping, laundering, use of transportation, medication management, and financial management) and six basic ADLs (i.e., feeding, dressing, grooming, ambulating, bathing, and toileting). The percent of independence on relevant items was provided for iADLs and ADLs.

9.2.2.2 MoCA

MoCA is a paper-and-pencil screening tool for CI [166]. It evaluates multiple cognitive domains and its score is corrected for low education (≤ 12 years) by adding an extra point, having a maximum score of 30 points [166]. With a high sensitivity of 83 to 97%, MoCA is currently the preferred tool for the screening of CI at the primary care level [22]. The cutoff of < 26 is commonly used to detect CI [166,237,428].

9.2.2.3 Neuropsychological Assessment

The neuropsychological assessment consisted of a standardized battery administered to all participants in the ONDRI study. The ONDRI study followed standard quality assurance and quality control procedures, as described in [431], ensuring the rigor and accuracy of the data.

To characterize areas of CI in individual participants, the tests were categorized into five cognitive domains, as proposed by [432] after a consensus agreement among the ONDRI Clinical Neuropsychologists. In addition to raw scores from the neuropsychological tests, the ONDRI dataset provided standard scores (z-scores, t-scores, percents, or scaled scores) based on published normative data (education- and/or age-adjusted). Normative values were not available in two of the considered tests (BDAE: Semantic probe and BVMT-R: Copy trial) and they were, thus, excluded from the analysis. The resulting cognitive domains and associated test measures are shown in Table 9.1.

We categorized participants as impaired or as cognitively normal using the comprehensive criteria proposed by [433]. According to this criteria, individuals were classified as cognitively normal if, at most, performance on one measure within one or two cognitive domains fell more than 1 SD below age/education-appropriate norms [433].

Before applying the comprehensive criteria, we filled in missing values using k-Nearest Neighbors (KNN). Missing neuropsychology data were imputed using the mean value from the 5 nearest neighbors weighted by the inverse of their Euclidean distance. Since KNN is based on distances, before imputation, we standardized the data, i.e. removed their mean and scaled to unit variance, to generate unbiased estimations of missing values [434]. After imputation, the values were transformed to the original scale.

9.2.2.4 Gait

Quantitative gait parameters were assessed using wearable inertial sensors (Gulf Coast Inc.; Shimmer Inc.) worn bilaterally on the ankles and at the hip. In two study sites, gait performance was additionally assessed using electronic walkway systems (GAITRite® or PKMas®) [425,435]. To facilitate a comparison between inertial sensors and electronic walkway systems, we have only considered the participants who were assessed by the two systems.

Gait was assessed under three conditions, using the following order: 1) preferred walking speed, or single-task walking (SS), 2) dual-task walking (DT), and 3) fast walking (F). The secondary tasks were administered in the following order: 1) subtracting serial ones from 100 (DS1), 2) naming animals (DAn), and 3) subtracting serial sevens from 100 (DS7). The cognitive tasks were performed out loud, without any prioritization instruction. All participants walked along a 6-meter path, starting and finishing one meter away to ensure that only steady-state walking was captured [425,435].

The preferred walking speed condition was repeated three times, and average values were calculated. Gait characteristics such as speed, cadence, and total number of steps were reported for the participants assessed with the accelerometer. The participants assessed using the electronic walkways had additional gait characteristics, including, gait speed, stride velocity, step and stride time, step and stride length, double support time, swing time, step and stride width, cadence, and corresponding variabilities, calculated using the coefficient of variation, CoV, obtained from $[(\text{standard deviation}/\text{mean}) \times 100]$ [435].

Table 9.1: Neuropsychological assessment.

Cognitive Domains	Tests Included
Attention and working memory	Symbol digit modality test (coding)
	Trail making test–Part A (time)
	WAIS-III: Digit span forward (longest span)
	WAIS-III: Digit span backward (longest span)
	WAIS-III: Digit span total
	DKEFS: Color naming (time)
Executive function	DKEFS: Word reading (time)
	Trail making test–Part B (time)
	DKEFS: Interference (time)
	DKEFS: Inhibition/switching (time)
	DKEFS: Letter fluency
	DKEFS: Category fluency
Language	WASI-II: Matrix
	Boston naming – 15 item (pro-rated)
	TAWF: Verb naming
	BDAE: <i>Semantic probe (excluded)</i>
Verbal memory	WASI-II: Vocabulary
	RAVLT: Immediate
	RAVLT: Long-delay
Visuospatial awareness	RAVLT: Recognition discrimination
	Judgement of line orientation
	VOSP: Incomplete letters
	BVMT-R: <i>Copy trial (raw) (excluded)</i>

BDAE: Boston Diagnostic Aphasia Examination; BVMT-R: Brief Visuospatial Memory Test-Revised; DKEFS: Delis-Kaplan Executive Function System; RAVLT: Rey Auditory Verbal Learning Test; TAWF: Test of Adolescent/Adult Word Finding; VOSP: Visual Object and Space Perception Battery; WAIS-III: Wechsler Adult Intelligence Scale–Third Edition; WASI-II: Wechsler Abbreviated Scale Intelligence–Second Edition.

Since some of the metrics extracted from the electronic walkway provide similar information (e.g., stride time and cadence, step and stride metrics), we excluded step metrics and cadence from the analysis. The metrics included were, thus, stride length, double support time, stride time, stride velocity, stride width, swing time, and their respective variabilities.

For all these metrics we calculated dual-task costs, or effects (DTE, in %) using the formula $[(\text{single-task metric} - \text{dual-task metric}) / \text{single-task metric}] \times 100$, which quantifies the magnitude of the effect of cognitive load on gait performance [435]. We also calculated the capacity index comparing preferred speed with fast walking trials using the formula $[(\text{fast metric} - \text{single-task metric}) / \text{single-task metric}] \times 100$.

9.2.3 Machine Learning Pipeline

We developed a machine learning pipeline for classifying participants into normal or cognitively impaired using the metrics from gait tests. We assessed each gait test separately (i.e., SS, DAn, DS1, DS7, and F) to compare each test's ability to differentiate both groups. Additionally, we tested the combinations of dual-task tests (and fast-walking test) with SS, i.e., DAn+SS, DS1+SS,

DS7+SS, and F+SS. To this purpose, we analyzed and reported the best-performing experiment from the following three possibilities: a) using dual-task costs (or capacity indexes) alone, b) combining dual-task costs (or capacity indexes) and SS metrics, and c) combining dual-task costs (or capacity indexes) and dual-task (or fast-walking) test metrics. The five gait tests and the four combinations resulted in a total of nine reported experiments.

The experiments were performed using metrics provided by the accelerometer and repeated using the metrics provided by the electronic walkway. Since our goal was to compare the limited set of features provided in this study by the accelerometer with the more comprehensive set provided by the electronic walkway (described in Section 9.2.2.4), we have only included participants assessed by the two systems.

One participant in the cognitively normal group had no accelerometer data on DAn, DS7, and F tests and was, as such, excluded from experiments including these data. The same happened with one participant in the cognitively impaired group that had no accelerometer data on the DAn test. All the remaining variables were present, yielding 29 participants in the impaired group (or 28 when data were missing) and 18 in the cognitively normal group (or 17 when data were missing).

To avoid overly optimistic estimations and account for the limited number of participants in this study, we used a nested cross-validation (CV) approach [436], which consists of an outer and inner CV loop, as shown in Figure 9.1. The dataset was first divided into 20 equally-sized outer folds, yielding 42 to 45 participants' data for hyperparameter tuning and model training (training sets) and 2 to 3 participants' data for evaluation (test sets). In the inner loop, the outer training data were further divided into 5 folds, forming the inner training and validation folds. The inner loops were used for hyperparameter tuning in a grid search. When the best set of hyperparameters was found, the whole (outer) training set was used to retrain the model. The model was then used to classify the test set in the outer loop.

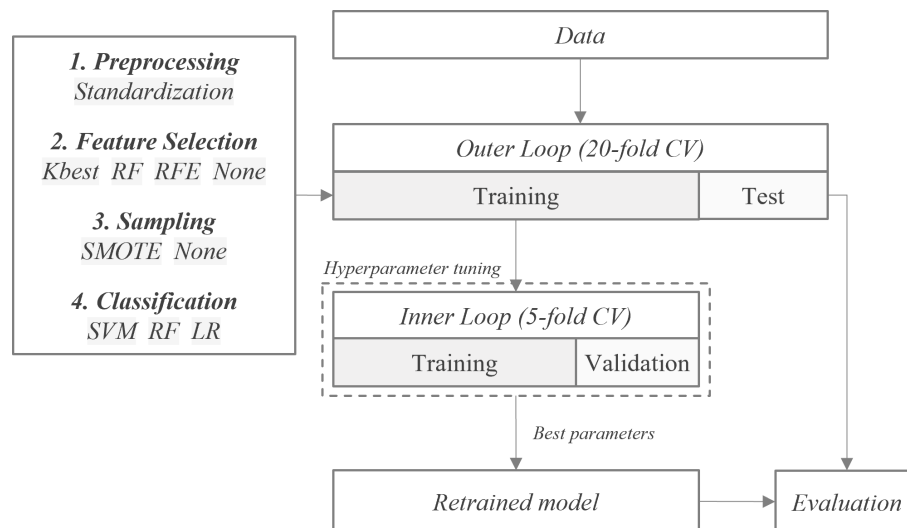


Figure 9.1: Machine learning pipeline with nested cross-validation.

As illustrated in Figure 9.1, the machine learning pipeline consists of 1) preprocessing, 2) feature selection, 3) sampling, and 4) classification. The preprocessing step consisted of a z-score standardization that scaled the features to zero mean and unit standard deviation. For the feature selection, we tested different methods: we optimized the number of features to select (1, 2, 3, 5, or 8) using the highest-scored features (KBest) according to their mutual information (or mutual dependence) with the target [404], or using recursive feature elimination (RFE) based on the feature ranking provided by a Support Vector Machine (SVM) with a linear kernel; we also tested feature selection based on feature importance as provided by Random Forest (RF) [437] with 50 estimators, selecting all those features with importance above the mean. We also tested the pipeline without any feature selection. In the sampling step, we used the Synthetic Minority Oversampling Technique (SMOTE) algorithm [405]. Since the incidence of cognitively normal participants in the dataset was only 38.3%, SMOTE was used to create new synthetic examples of the minority class. According to [405] SMOTE provides better results when applied after feature selection. We also evaluated the performance of the classifiers without applying any resampling strategy, as all the classifiers tested in the subsequent stage were able to adjust for dataset imbalance by using class weighting. In the classification stage, we optimized the following hyperparameters in three different classifiers: the regularization strength, C (0.001, 0.01, 0.1, 1.0, 10, or 100), and kernel (linear or rbf) of the Support Vector Machine (SVM); the number of estimators (10, 50, or 100) of the Random Forest (RF); and the regularization strength, C (0.001, 0.01, 0.1, 1.0, 10, or 100), of the Logistic Regression (LR) classifier.

The total number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were obtained for all test partitions in the outer loop and used to calculate the global metrics of performance, where the cognitively impaired group was considered the positive class. To evaluate and compare the performance of the machine learning models we calculated balanced accuracy (Equation 2.7), sensitivity (Equation 2.3), specificity (Equation 2.4), F1 score (Equation 2.6), and the area under the receiver operating characteristic (ROC) curve (AUC)—as described in Chapter 2. Since we were dealing with imbalanced data, balanced accuracy was used as the scoring criteria to select the best set of hyperparameters in the grid search. Balanced accuracy was also used to compare multiple machine-learning experiments.

Experiments were performed using Python 3 and the package scikit-learn v1.1.3 [438].

9.2.4 Classification using MoCA Scores

We compared the performance of the machine learning experiments with MoCA. Two different experiments were conducted: 1) we used the classical threshold of 26 to identify the groups and 2) we used a ROC analysis to determine the optimal MoCA cut-point that maximized sensitivity and specificity in the current dataset. The optimal cut-point was considered the point of the ROC curve closest to the upper left corner (0,1), defined as that yielding the minimum value of $(1 - sensitivity)^2 + (1 - specificity)^2$ [369]. A third experiment combined the classifications obtained with MoCA with those obtained using gait metrics in two of the best-performing conditions. The

classifications obtained by the three different approaches were combined using a majority voting scheme [439].

9.2.5 Statistical Analysis

Demographic and medical characteristics were summarized using either means and standard deviations, or frequencies and percentages, as appropriate, to characterize both groups. The comparisons between groups were performed using the Independent Samples T-test or the Mann-Whitney U test for continuous variables, and Person's chi-square test (χ^2) for categorical variables. The Mann-Whitney U test was used when data were not normally distributed or when outliers were present outside ± 1.5 of the interquartile range. To test for normality, we used the Shapiro-Wilk test.

To compare the results of the machine learning experiments, we used Paired Samples T-Tests or Wilcoxon signed-rank tests, that compared the average performance of the validation sets in all inner cross-validation (CV) runs. Wilcoxon signed-rank tests were used when paired data differences were not normally distributed or when outliers were present. Bonferroni-adjusted p-values were reported for the multiple comparisons [440].

McNemar's Test was used to compare the diagnosis provided by MoCA, and the classifications provided by machine learning experiments, as recommended in [409].

The statistical analysis was performed using Python 3 and the statistical tools from SciPy v1.9.3, NumPy v1.23.5, and statsmodels v0.13.5 [441]. A p-value of less than 0.05 indicated statistical significance.

9.3 Results

9.3.1 Descriptive Statistics

Of the 161 participants with CVD available in the ONDRI dataset, only 47 had gait features measured by both the accelerometer and the electronic walkway. Of the 47 participants, 29 (61.7%) met the criteria for CI, whereas 18 (38.3%) were classified as cognitively normal. The descriptive statistics of the two groups are provided in Table 9.2.

Significant differences between groups were obtained on MoCA scores and iADLs percent, indicating lower global cognitive performance in the impaired group and less independence on iADLs. The difference in ADLs was not significant, denoting that both groups were still equally capable of performing these tasks. Additionally, there were no significant differences in depression scores, age, and education between the two groups.

The group with CI had a slower gait speed, as indicated by significant differences in multiple gait tests. The electronic walkway revealed significant differences in gait speed in more tests than the accelerometer. No significant differences between groups were obtained for gait speed while walking fast, as measured either by the accelerometer or the electronic walkway. Although the

Table 9.2: Descriptive statistics (n = 47).

Characteristic	Normal (n = 18)	Impaired (n = 29)	p-value
Age [years]	67.3 ± 5.1	69.6 ± 7.3	0.220
Height [cm]	174.7 ± 6.5	169.1 ± 10.7	0.033*
Leg length [cm]	93.1 ± 6.2	91.8 ± 8.1	0.571
Weight [Kg]	85.7 ± 12.5	81.8 ± 14.2	0.242
Female, n (%)	4 (22.2)	8 (27.6)	0.098
Education [years]	15.3 ± 2.9	15.4 ± 2.7	0.781
Clinical tests			
QIDS Score	4.0 ± 2.5	4.1 ± 3.0	0.833
iADLs [%]	95.9 ± 6.4	85.2 ± 18.1	0.036*
ADLs [%]	99.5 ± 1.3	99.3 ± 2.5	0.949
MoCA Score	27.6 ± 2.0	24.3 ± 2.8	<0.001*
Gait using the accelerometer			
SS Speed [m/s]	1.18 ± 0.28	0.98 ± 0.27	0.024*
DAn Speed [m/s]	1.05 ± 0.26	0.84 ± 0.26	0.012*
DS1 Speed [m/s]	1.05 ± 0.29	0.91 ± 0.28	0.112
DS7 Speed [m/s]	0.96 ± 0.32	0.79 ± 0.25	0.106
F Speed [m/s]	1.73 ± 0.49	1.45 ± 0.45	0.060
Gait using the electronic walkway			
SS Speed [m/s]	1.25 ± 0.15	1.11 ± 0.20	0.021*
DAn Speed [m/s]	1.10 ± 0.24	0.92 ± 0.21	0.007*
DS1 Speed [m/s]	1.17 ± 0.21	1.02 ± 0.23	0.024*
DS7 Speed [m/s]	1.15 ± 0.32	0.86 ± 0.19	0.002*
F Speed [m/s]	1.67 ± 0.41	1.52 ± 0.27	0.088

Data are mean values ± standard deviation or the number of participants per category (absolute and relative frequency) when indicated. Group differences were evaluated using the Independent Samples T-test, Mann-Whitney U test, or Pearson's chi-square. * $p < 0.05$, p-values are two-tailed significance and bold values indicate significance. QIDS: Quick Inventory of Depressive Symptomatology; iADLs: Percent of independence in instrumental activities of daily living; ADLs: Percent of independence in activities of daily living; SS: Single-task; DAn: Walking while naming animals; DS1: Serial subtraction by 1s; DS7: Serial subtraction by 7s; F: Walking Fast.

impaired group had significantly shorter height, leg length presented no significant differences between groups, which should not contribute to differences in gait performance [59].

9.3.2 Classification using Accelerometer Metrics

The best-performing classifiers, feature selectors, and sampling methods are shown in Table 9.3 for each tested gait condition using metrics from the accelerometer. The average number (and standard deviation) of features selected (within each outer training fold) are also shown in this table. Performance metrics are reported for the test set.

Balanced accuracy spanned from 60.4% (using F walking features) to 70.6% (using DS1+SS walking features). The features extracted from the SS condition provided better performance than DAn, DS1, DS7, and F conditions, with 72.4% of sensitivity, 66.7% of specificity, and 66.3% of AUC. The performances of dual-task conditions (and F) improved when dual-task costs (or capacity indexes) were included (e.g., the performance of DAn+SS was higher than DAn alone). The

Table 9.3: Performance on test sets using gait metrics from the accelerometer (in %).

Condition	SS	DAn	DS1	DS7	F	DAn+SS ^{a)}	DS1+SS ^{b)}	DS7+SS ^{c)}	F+SS ^{b)}
Classifier	LR	LR	SVM	LR	LR	RF	LR	LR	RF
Feature Sel.	None	RF	None	None	None	RFE	RFE	KBest	KBest
Sampling	None	SMOTE	SMOTE	None	None	None	None	SMOTE	None
# Features	3.0±0.0	2.0±0.0	3.0±0.0	3.0±0.0	3.0±0.0	2.0±0.0	4.6±1.0	4.5±1.0	3.4±1.7
Bal. acc.	69.5	64.5	62.6	61.2	60.4	66.9	70.6	68.1	66.1
Sensitivity	72.4	64.3	58.6	51.7	62.1	75.0	69.0	65.5	79.3
Specificity	66.7	64.7	66.7	70.6	58.8	58.8	72.2	70.6	52.9
F1 score	75.0	69.2	65.4	61.2	66.7	75.0	74.1	71.7	76.7
AUC	66.3	65.5	65.7	57.4	63.1	63.4	75.5	66.5	57.8

SS: Single-task; DAn: Walking while naming animals; DS1: Serial subtraction by 1s; DS7: Serial subtraction by 7s; F: Walking Fast; LR: Logistic Regression; RF: Random Forest; SVM: Support Vector Machine; RFE: Recursive feature elimination; AUC: Area under the curve. ^{a)}using dual-task costs; ^{b)}using dual-task costs (or capacity indexes) and SS metrics; ^{c)}using dual-task costs and dual-task metrics.

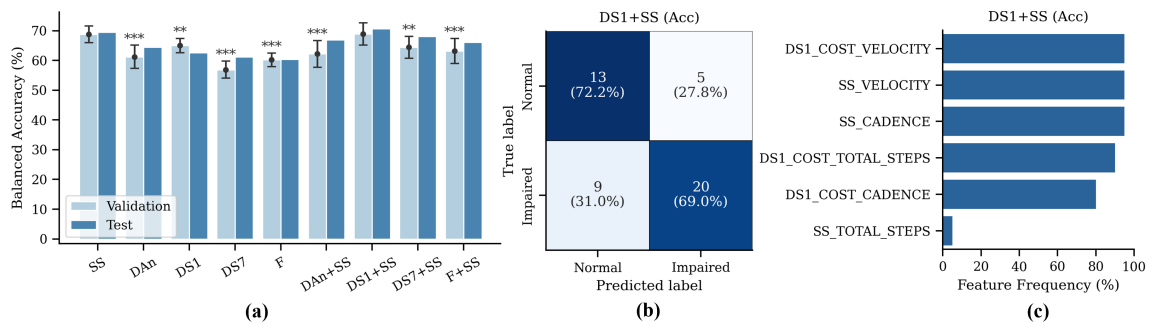


Figure 9.2: Performance using features extracted from the accelerometer. **(a)** Performance on validation and test sets, where vertical lines indicate the standard deviation of the 20 inner CVs. The statistical analysis compares the best-performing experiment (DS1+SS) with the results achieved in the other conditions, using Paired Samples T-tests or Wilcoxon signed-rank tests with Bonferroni correction, where $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$ indicate statistical significance. **(b)** Confusion matrix of the DS1+SS condition. **(c)** Feature selection frequency for the DS1+SS condition.

best-performing combination overall was with DS1+SS, with a sensitivity of 69.0%, a specificity of 72.2%, and an AUC of 75.5% (Table 9.3).

The performance (balanced accuracy) on validation and test sets using gait metrics from the accelerometer is visually compared in Figure 9.2–a. Statistical results are provided in this figure comparing DS1+SS with the other conditions.

From the analysis of Figure 9.2–a we verify that classification performance (measured on validation folds) of the best-performing condition, DS1+SS, although superior to the SS condition, does not differ significantly. All the remaining conditions had a lower average performance, differing significantly from the DS1+SS condition.

We note that in the SS condition only 3.0 ± 0.0 features were included, whereas in the DS1+SS condition an average of 4.6 ± 1.0 were included (Table 9.3). In the case of the accelerometer, as only three metrics were measured (i.e., number of steps, cadence, and speed), the maximum number of features selected was either three or six in the case of the experiments including dual-task costs. The features that were most commonly used in the DS1+SS condition are shown in

Table 9.4: Performance on test sets using gait metrics from the electronic walkway (in %).

Condition	SS	DAn	DS1	DS7	F	DAn+SS ^{a)}	DS1+SS ^{a)}	DS7+SS ^{b)}	F+SS ^{b)}
Classifier	LR	SVM	LR	SVM	RF	SVM	SVM	SVM	LR
Feature Sel.	RFE	RFE	RFE	None	RF	None	KBest	RF	RFE
Sampling	SMOTE	SMOTE	None	SMOTE	None	None	None	SMOTE	SMOTE
# Features	2.0±2.2	3.2±2.2	2.2±1.8	12.0±0.0	6.0±0.8	12.0±0.0	1.7±2.1	8.6±1.3	2.8±2.1
Bal. acc.	76.8	75.1	71.6	77.5	60.5	74.0	78.8	75.1	69.5
Sensitivity	75.9	72.4	65.5	82.8	65.5	75.9	96.6	72.4	72.4
Specificity	77.8	77.8	77.8	72.2	55.6	72.2	61.1	77.8	66.7
F1 score	80.0	77.8	73.1	82.8	67.9	78.6	87.5	77.8	75.0
AUC	75.3	75.9	69.2	79.5	54.8	73.9	68.6	72.4	67.2

SS: Single-task; DAn: Walking while naming animals; DS1: Serial subtraction by 1s; DS7: Serial subtraction by 7s; F: Walking Fast; LR: Logistic Regression; RF: Random Forest; SVM: Support Vector Machine; RFE: Recursive feature elimination; AUC: Area under the curve. ^{a)}using dual-task costs; ^{b)}using dual-task costs (or capacity indexes) and SS metrics.

Figure 9.2–c. In this condition, the most relevant features were the cost in velocity, single-task velocity, single-task cadence, the cost in total steps, and the cost in cadence. The total number of steps in the single-task condition was the least relevant feature. Figure 9.2–b shows the confusion matrix of the DS1+SS condition.

9.3.3 Classification using Walkway Metrics

The results achieved using the metrics extracted from the electronic walkway are shown in Table 9.4. As with the accelerometer, the best performing condition was the combination DS1+SS, with 78.8% of balanced accuracy, 96.6% of sensitivity, 61.1% of specificity, and 68.6% of AUC. When using features from the gait test alone, DS7 (with 77.5% of balanced accuracy) revealed the highest performance in comparison with SS, DAn, DS1, and F conditions. Balanced accuracies spanned from 60.5% (with F) to 78.8% (with DS1+SS), being test scores higher than its equivalents with the accelerometer.

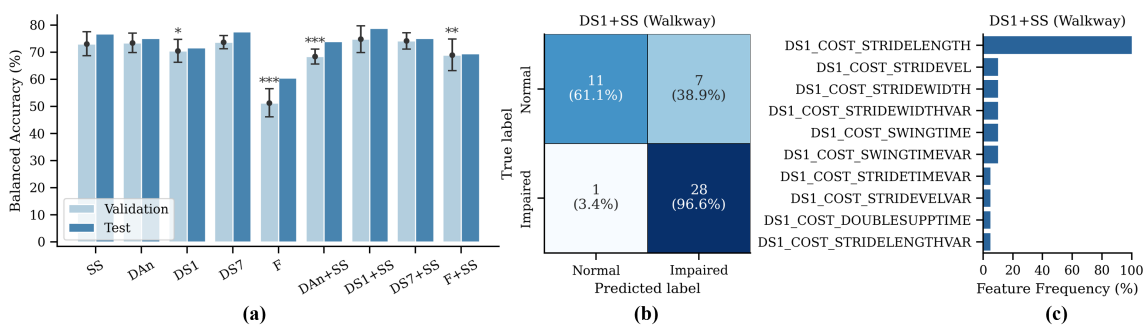


Figure 9.3: Performance using features extracted from the walkway. **(a)** Performance on validation and test sets, where vertical lines indicate the standard deviation of the 20 inner CVs. The statistical analysis compares the best-performing experiment (DS1+SS) with the results achieved in the other conditions, using Paired Samples T-tests or Wilcoxon signed-rank tests with Bonferroni correction, where $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$ indicate statistical significance. **(b)** Confusion matrix of the DS1+SS condition. **(c)** Feature selection frequency for the DS1+SS condition.

Although the electronic walkway provided an additional number of gait metrics than the accelerometer, the number of features used by machine learning models was not always higher. The SS condition, for instance, used only 2.0 ± 2.2 features, whereas in the case of the accelerometer an average number of 3.0 ± 0.0 features were used in this condition. However, the features included were different. In the SS condition, for instance, double support time was selected for all the 20 outer training folds and stride width variability was selected 5 times (results not shown). Double support time was one of the most relevant features of the electronic walkway in several conditions where feature selection has been applied, including, SS, DAn, DS1, F, DS7+SS, and F+SS. In the case of the DS1+SS condition, the best balanced accuracy of 78.8% was achieved with 1.7 ± 2.1 features, being stride length cost one of the most relevant features, selected in all folds of the outer training loop (Figure 9.3—c). Only 1 participant in the impaired group was misclassified using DS1+SS features (Figure 9.3—b), leading to a very high sensitivity of 96.6%.

Compared to the DS1+SS condition, only DS1, F, DAn+SS, and F+SS conditions revealed significantly lower performances (Figure 9.3—a). The performance of the DS1+SS was superior to the SS in validation and test folds, although the differences in validation performances were not statistically significant.

A Paired Samples T-test revealed that the performance of the machine learning models in the DS1+SS condition using walkway features was significantly higher than the performance of the DS1+SS condition using the features extracted from the accelerometer ($T = -4.446$, $p < 0.001$).

9.3.4 Performance using MoCA

Table 9.5 and Figure 9.4 summarize the results of the classification when a single cut-point was employed to MoCA. Using the traditional cut-point of 26, the MoCA's sensitivity was 62.1% and the specificity was 77.8% in the present sample (Table 9.5). The ROC analysis (ROC curve shown in Figure 9.4—a) indicated that in the present sample the cutoff of 27 would optimize sensitivity and specificity, which was according to the results obtained in [428]. The cutoff of 27 yielded a sensitivity of 75.9%, a specificity of 77.8%, and an AUC of 66.0%.

Since the DS1+SS condition measured by the walkway yielded the best classification performance when using gait metrics (Table 9.4), and because the DS1+SS condition requires a SS test in order to measure the dual-task costs, we combined the classification results using a majority voting approach that considered the classifications provided by the SS condition, the DS1+SS condition and $\text{MoCA} < 27$. The majority voting resulted in an improved balanced accuracy of 82.7%, a sensitivity of 93.1%, and a specificity of 72.2%. The resulting confusion matrix is shown in Figure 9.4—b.

An exact McNemar's Test showed that there were no statistically significant differences in the disagreements between the classifications provided by $\text{MoCA} < 27$ and the classifications provided by the machine learning models using the best performing condition of DS1+SS (walkway), with $p = 0.549$. The observed differences in classification results were also not statistically significant when the traditional threshold (< 26) was employed, with $p = 0.092$. Although the majority voting scheme revealed the highest balanced accuracy overall, the observed differences in classification

results were not significantly different from DS1+SS ($p = 1.0$), SS ($p = 0.289$), or MoCA<27 ($p = 0.219$), but differed significantly from MoCA<26 ($p < 0.05$).

Table 9.5: Performance using MoCA and a combination with gait.

Classifier	MoCA < 26	MoCA < 27	Majority voting
Bal. accuracy	69.9	76.8	82.7
Sensitivity	62.1	75.9	93.1
Specificity	77.8	77.8	72.2
F1 score	70.6	80.0	88.5

MoCA: Montreal Cognitive Assessment.

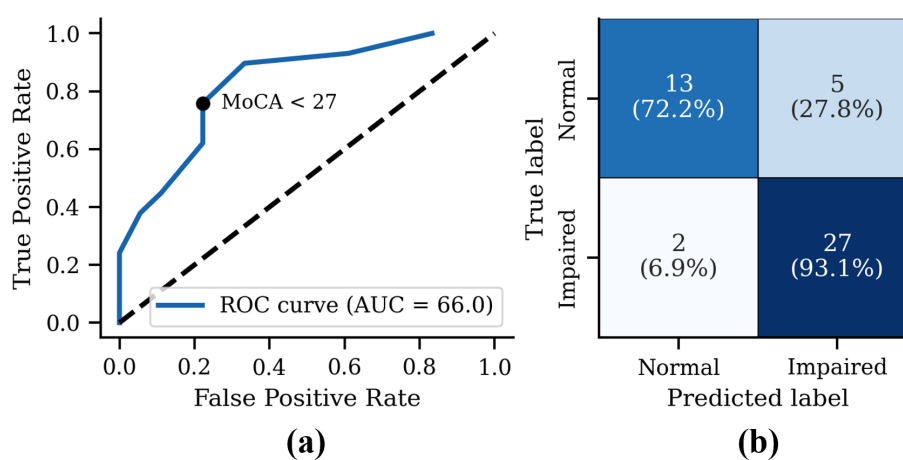


Figure 9.4: Classification performance using MoCA and a majority voting approach. **(a)** ROC analysis used to determine the optimal MoCA cut-point and **(b)** Confusion matrix resulting from the majority voting approach, which combined the SS condition, the DS1+SS condition (walkway) and MoCA<27. AUC: Area under the curve; ROC: receiver operating characteristic.

9.4 Discussion

This study was the first to investigate the use of machine learning and gait characteristics for detecting CI in post-stroke patients using a gold-standard neuropsychological battery as a reference. We used a nested cross-validation approach to evaluate the performance of multiple classifiers and feature selection methods on different walking conditions, including single-task (SS), dual-task (DAn, DS1, and DS7), and fast walking (F). Moreover, we tested the inclusion of dual-task costs and capacity indexes resulting from the combination of SS with DAn, DS1, DS7, or F. In line with our initial hypothesis, gait characteristics were able to distinguish between post-stroke individuals with and without CI. Moreover, a comprehensive set of gait metrics and dual-task costs allowed better differentiation between the two groups, with classification performances that were comparable to or even superior to those achieved with MoCA. Combining gait metrics with MoCA improved overall classification performance.

9.4.1 Performance using Gait Metrics

The machine learning experiment leading to the best classification overall was achieved using DS1+SS gait metrics extracted from the electronic walkway. Although disagreements in classification results were not significantly different from MoCA (using cutoffs of < 26 or < 27), the balanced accuracy achieved with gait metrics was higher (78.8% versus 69.9% or 76.8%, respectively), which is a positive indication that gait analysis can be a useful tool for the screening of CI in post-stroke patients. Compared to MoCA, gait tests are simpler and allow more objective and faster evaluations, which justifies their relevance in the context of cognitive screening [212].

The DS1+SS condition was consistently identified as the best-performing condition, either using the accelerometer or the walkway. However, in both cases, the differences between the DS1+SS condition and the SS condition were not statistically significant. This suggests that even a simple single-task gait analysis could provide acceptable classification results, without the need for participants to complete additional tests. Despite this, the sensitivity obtained in the DS1+SS condition using the electronic walkway was much higher at 96.6%. This suggests that the DS1+SS condition should be a preferred choice for cognitive screening, where the main goal is to detect all potential cases for further assessment [237].

The lower classification performances obtained with the accelerometer may be in part due to the lower level of detail it provided compared to the electronic walkway. While the accelerometer could only provide general gait parameters, like, velocity, cadence, and total steps, the electronic walkway provided additional spatiotemporal metrics that characterized the strides (e.g., length, speed, width) and gait phases (e.g. swing and double support time). We should note that the limitations of the accelerometer in this regard are not due to the device itself, but rather to the algorithms used to process its data [268, 442, 443]. It's also worth noting that in the analysis of the descriptive statistics, the electronic walkway detected significant differences in gait speeds between the two groups in more tests than the accelerometer (Table 9.2). This may suggest that there are differences in the performance of the two technologies. In fact, different algorithms can be used to process data from inertial sensors, and the inclusion of a gyroscope can often lead to improved results [268, 442, 443]. Therefore, the choice of algorithm and sensors used may have an impact on the results obtained. Thus, a possible lack of instrument accuracy (and/or precision) could partially justify the obtained results. Yet, including more detailed evaluations of gait seems to be beneficial, which is according to previous results on the detection of CI. According to previous studies, gait speed may be associated with several adverse health outcomes [444, 445], and therefore, it is not specific for the detection of CI [53, 197]. Since no other pathologies were included in our study, these results could not be confirmed.

Previous studies showed that different cognitive tasks can impact walking in different ways. In general, more cognitively demanding tasks tend to lead to greater interference effects, although the results are not always consistent [52]. The serial subtraction by sevens is particularly cognitively demanding and has been shown to be the most sensitive in distinguishing patients with CI in non-stroke populations [52]. In this study, the classification performance achieved with the

DS7 and DS7+SS conditions was significantly lower than the DS1+SS condition when using the accelerometer (Figure 9.2–a), but did not differ from the DS1+SS condition when using the walkway (Figure 9.3–a). However, the DS1+SS condition led to higher performances in both cases, indicating better discrimination ability. It's worth noting that post-stroke patients often have other disabilities, particularly physical ones [418], and the mechanisms underlying motor-cognitive interferences may be different in this population compared to patients with other neurocognitive disorders such as Alzheimer's disease. Therefore, further research is needed to understand the effects of different cognitive tasks on post-stroke patients.

Previous research has also shown that combining multiple gait metrics can provide better discrimination ability than using a single gait metric [59, 60]. However, some gait metrics may be highly related and their simultaneous inclusion may not provide additional discrimination ability [375]. In this study, we used feature selection methods to identify the most relevant features based on different criteria. In the best-performing condition (DS1+SS using the walkway), features were selected based on feature importance as provided by the Random Forest. In this experiment, only dual-task costs were selected for classification. According to previous studies, dual-task costs represent only the effect of adding the cognitive task and are not influenced by individual baseline characteristics (e.g., leg length) that could otherwise be considered covariates [377, 422]. For this reason, dual-task costs constitute an interesting evaluation metric for the detection of cognitive disorders (consistent with the results reported in Chapter 7).

The performances achieved in this study are at the level of the performances reported in the literature for the detection of CI in non-stroke patients. For instance, using dual-task gait data from an electronic walkway, Boettcher et al. [225] reported sensitivities of 82.0% and specificities of 67.7% on the detection of mild CI. Also with an electronic walkway, Ghoraani et al. [229] differentiated cognitively impaired groups from the healthy subjects with an accuracy of 86.0% and an F1 score of 88.0%. Using a detailed evaluation of gait provided by an inertial sensor-based gait analysis solution, Shahzad et al. [232] could achieve classification accuracies of 70.0% and sensitivities of 83.3% for the detection of mild CI, which proves the capacity of inertial sensor-based solutions.

9.4.2 Performance using MoCA

Similarly to the study by Zaidi et al. [428], the threshold of 27 in MoCA was the one achieving the best combination of sensitivity and specificity (Table 9.5). Although this study used the same dataset, they included all 161 participants and employed different neuropsychological criteria to characterize CI. In [428] the threshold of 27 resulted in a maximized sensitivity of 79% and specificity of 67% in detecting CI post-stroke. As in our study, increasing the cutoff from 26 to 27 allowed an increase in sensitivity.

To increase specificity, some studies recommend lowering the threshold of MoCA [237, 419]. A pooled analysis showed that MoCA scores below 26 are able to detect post-stroke CI with a sensitivity of 95% and a specificity of 45%, but lowering the threshold to 22 results in a lower sensitivity of 84% and a better specificity of 78% [237]. The differences in sensitivity and specificity

reported by the different studies are mostly justified by the different neuropsychological criteria used to characterize CI [421].

Although disagreements in classification results obtained with $\text{MoCA} < 27$ and with the best-performing gait condition did not differ significantly when compared to the majority voting approach, the combination of these two modalities resulted in an overall improvement in classification performance (Table 9.5). Compared to $\text{MoCA} < 26$, the disagreements in classifications were significantly different, with the majority voting scheme achieving better sensitivity (93.1% versus 62.1%) for detecting CI. The sensitivity also improved compared to $\text{MoCA} < 27$ (93.1% versus 75.9%), making the majority voting approach more suitable for screening purposes [237]. Although its sensitivity was lower than that of the DS1+SS condition (93.1% versus 96.6%), it still remained very high, with a large improvement in specificity (72.2% versus 61.1%), which should favor its choice. The majority voting approach has received much attention in the literature for its simplicity and good performance on real data and has been applied in various fields [439]. To the best of our knowledge, this is the first study to propose the combination of MoCA with gait metrics for classifying cognitive status in post-stroke patients.

9.4.3 Study Limitations

Several factors constitute the limitations of our study. First, the sample size was relatively small, which may affect the robustness of the models and the generalizability of our findings. Second, the criteria for classifying participants as impaired or cognitively normal based on neuropsychological assessments are not universally accepted and may vary depending on the cognitive domains evaluated, the tests used, their grouping criteria, the cutoffs demarcating impairment (most commonly 1 SD, 1.5 SD, or 2 SD), and the number of tests required to declare a deficit in a certain domain [421,433]. In our study, we employed the comprehensive criteria recommended by Jak et al. [433] and the cognitive domains proposed by Dillio et al. [432], which resulted in 61.7% of the patients meeting the criteria for CI. This frequency is above the pooled prevalence rates reported in the literature for the first year after stroke [421], but it is important to note that different criteria could lead to different results. Third, the participants in this study may have higher functioning levels than those typically seen in the general clinic, as those with severe functional disabilities were excluded from the study. Moreover, only individuals with $\text{MoCA} > 18$ were included, which reduced the range of cognitive disorders associated with stroke that could be tested in this study. However, it is worth noting that our approach was able to detect CI in this group, which included more subtle symptoms that are usually harder to detect. Fourth, we did not evaluate the impact of the different study sites on the results. Although efforts were made to standardize assessments and ensure data quality across sites [425,431,435], differences could still be observed due to discrepancies in acquisition. This effect should be explored in future research. Finally, it is important to note that we did not evaluate the impact of the walking task on cognitive performance, although Plummer et al. [200] argue that the correct interpretation of the dual-task interference requires an objective evaluation of both tasks. Motor-cognitive interference may be observed in just one of the tasks, or in both simultaneously, and, for this reason, the performance on both tasks should

be taken into account [200]. Additionally, it is worth noting that the order of the gait tests was maintained for all participants, which could potentially affect the results.

9.4.4 Future Work

Further research is needed to determine the feasibility and effectiveness of gait assessments and machine learning models in detecting CI in post-stroke patients within a clinical setting. To achieve this, it would be necessary to include a larger and more diverse group of participants, including those with lower functioning levels and a wider range of cognitive abilities. Additionally, we should explore the detection of CI in the context of chronic CVD without a previous stroke [418].

Utilizing additional gait metrics as measured by inertial sensors could also be a valuable area of investigation. Inertial sensors have gained popularity in gait assessment due to their flexibility, low cost, and performance [58], and have the potential to extract gait metrics that have not yet been explored in this context (for instance, using the gait metrics according to Chapters 3, 4, and 7). As suggested by Jung et al. [230], using sequential gait parameters extracted from inertial sensors could improve classification results and should, thus, be explored in the future. Furthermore, it is important to include an analysis of the performance on the cognitive task to provide a complete assessment of motor-cognitive interference in these patients [200].

9.5 Conclusion

Cognitive impairment (CI) after stroke is a common but often undiagnosed condition. To better manage cognitive disorders resulting from cardiovascular disease, it is necessary to develop new techniques to improve the screening of CI. In this study, we examined the use of gait, dual tasks, and machine learning to detect CI in post-stroke patients. Our results showed that a machine learning approach combined with dual-task gait metrics can effectively differentiate impaired and cognitively normal groups, with a performance that is not inferior to commonly used screening tests like MoCA. We also found that when using gait to assess cognitive status, it is important to include gait features that evaluate individual strides and gait phases, as opposed to just overall gait metrics like speed, cadence, or number of steps, as these result in poorer classification results. Using a majority voting approach that combines MoCA and dual-task gait metrics resulted in improved classification results, with better combinations of sensitivity and specificity for the detection CI. This study demonstrates the potential of machine learning and gait assessments as objective tools for cognitive screening, offering a good alternative or complement to MoCA in identifying CI in post-stroke patients. Further research is required to evaluate the practical and effective performance of this approach in a clinical setting.

Part IV

FINAL REMARKS

Chapter 10

Conclusion

Despite recent efforts to improve screening, cognitive impairment is still detected too late, a problem that affects our increasingly aging society. Traditional cognitive screening methods rely on paper-and-pencil tests administered by trained and qualified individuals. Since their administration and interpretation are done by humans, they are susceptible to some subjectivity. Moreover, these tests are time-consuming and unsuitable for recurrent use. Because of these and other constraints, the detection of early stages of dementia occurs very late, compromising treatment and/or the discovery of new therapies.

Although cognitive decline is the most apparent indicator of a potential neurocognitive disorder, recent studies indicate that changes in motor skills can also be observed in affected patients. In particular, the simultaneous execution of motor and cognitive tasks, known as dual-task, exposes performance changes that are more evident in older adults with cognitive impairment. Objective assessment of performance in this type of task requires the use of instruments, but conventional instruments such as optical motion capture systems are not suitable for screening. In this sense, we proposed to explore the use of wearable alternatives for the objective evaluation of movement and study if they are capable of detecting motor problems associated with cognitive impairment.

Our contributions were thus structured in two parts: **i) in the first part**, we explored the use of inertial sensors for the objective evaluation of gait and stepping parameters, thus making significant contributions in the areas of time-series signal processing, movement analysis, and data-driven movement analysis; **ii) in the second part**, we explored the use of these metrics to detect cognitive impairment through motor, cognitive, and dual tasks, using statistical and machine learning approaches, making significant contributions to the fields of biomedical and health informatics.

In this chapter, we present an overview of our conclusions, reflecting on the main contributions, implications, significance, and prospects for future work.

10.1 Overview of the Contributions

The first contributions of this thesis were in the area of **movement analysis**: novel tools were developed for the evaluation of gait and stepping relying on the analysis of foot-worn inertial sensor data, resulting in 3 journal publications and 1 conference paper. In this respect, the contributions can be summarized as follows:

- We developed a gait analysis solution that uses information from foot-worn inertial sensors to extract a comprehensive and clinically relevant set of gait parameters, including, stride length, gait speed, swing and stance phases duration, swing width, and minimum toe clearance (**Chapter 3**). The proposed approach is inspired by existing algorithms for movement analysis but introduces some specificities that allow gait metrics to be extracted irrespective of the sensor orientation on the shoes. This invariance to sensor orientation was demonstrated by simulating multiple rotations of the sensor. The proposed approach was developed and tested in a group of young adults, walking at multiple walking speeds, and using an optical motion capture system as a reference. Results showed a good agreement between both systems, particularly in straight walking conditions, which denotes its potential ability to detect clinically relevant differences in gait.
- Besides more traditional spatiotemporal gait parameters, also heel and toe trajectories can offer relevant information to assess individuals' walking performance. However, traditional approaches in the literature show a limited ability to evaluate three-dimensional trajectories resorting to a single foot-worn inertial sensor. To tackle this problem, we proposed a data-driven approach based on a deep recurrent network to estimate heel and toe trajectories from raw inertial sensor data on a stride-by-stride basis (**Chapter 4**). To support this formulation, stride trajectories were represented in a new coordinate frame that ensured inter-stride independence. Predicted trajectories were used to extract gait events and an extensive set of gait metrics, including speed, stride length, swing and stance phases duration, swing width, and minimum toe clearance. Using the same dataset as in Chapter 3, we showed that bi-directional LSTM networks can be applied to the context of gait analysis to obtain reliable estimates of three-dimensional heel and toe trajectories. In contrast to other deep learning approaches in the literature, the proposed approach focus on the estimation of trajectories that allow then the extraction of a comprehensive set of gait parameters. The approaches in the literature focus on the extraction of gait metrics directly from the deep learning models, which imply new training each time a new parameter needs to be included. The proposed solution is robust to imperfect sensor-shoe alignment conditions, due to the application of a data augmentation approach, that not only increased the total number of sequences in the training dataset but also allowed better generalization performance with robustness to different sensor orientations. Results obtained in this work support the application of this solution in a real scenario.

- We developed a novel foot-worn inertial sensor-based solution for the detection and evaluation of multidirectional steps while interacting with exergames (**Chapter 5**). Stepping exergames combine movement with challenging cognitive tasks and are, thus, a potential solution to assist cognitive screening at home. As in the gait analysis proposed in Chapters 3 and 4, inertial sensors were placed on the shoes, allowing objective evaluation of stepping movements. Inspired by previous works on gait analysis, we developed a multi-stage signal processing pipeline that allowed objective evaluation of stepping features. These features were then used in a supervised machine learning approach to classify the directions of the steps. Three novel mechanisms were proposed in this study: i) a novel foot sagittal reflection method that allowed the analysis of both feet without any distinction; ii) a better discrimination of similar step displacements due to the inclusion of the last step position as an input to the classifier; and iii) a periodic recalibration of sensor orientation estimations to better deal with orientation estimation drift over time. The solution was tested using data from young and older adults, supporting its application in the context of exergames as an alternative to the commonly used floor platforms. Movement characteristics like step duration, length, and speed, could be extracted from the steps, showing significantly lower values in older adults, consistent with the expected declines in movement as we age. This study provided the first insights into the potential of inertial sensors to detect and evaluate foot motions while interacting with stepping exergames.

Although wearable inertial sensors constitute a promising approach for the evaluation of movement, their applications to improve cognitive screening still lack validation for use and/or interpretation by clinicians. Therefore, in this thesis, we have also explored how inertial sensors and gait, stepping, and dual-tasks could support the **screening for cognitive impairment** in older adults. This area of research resulted in 5 journal publications. In this respect, our contributions can be summarized as follows:

- Following our contributions to the detection and evaluation of multidirectional steps, an exergame was developed that integrated the previously developed movement analysis solution. The exergame was comprised of several mini-games for balance training (focused on rapid and well-directed steps) and for cognitive-motor training (games combining stepping interaction with a cognitive task). Stepping and game parameters were recorded from a group of older adults with mobility limitations that played the exergame multiple times. The parameters measured on the first gameplay were used in a correlational study to investigate their association with standard cognitive and mobility performance indicators. We found that better stepping performance (with faster, longer, and higher steps) was associated with better mobility scores, while better cognitive games' scores and reaction times, and longer and faster steps were associated with better cognitive performance. People with lower cognitive scores also required additional time to advance to the next game difficulty level, presenting slower reaction times and stepping speeds as compared to people with higher cognitive

scores. This study thus provided some evidence that metrics extracted from stepping exergames, as measured by inertial sensors, may be useful for the assessment of the cognitive and motor status of older adults, potentially allowing assessments to be more frequent, affordable, and enjoyable (**Chapter 6**). A solution like the one proposed in this study may be useful for unsupervised cognitive screening, with interesting applications in the context of remote monitoring.

- In **Chapter 7** we investigated the potential of the instrumented gait assessments proposed in Chapters 3 and 4 to evaluate motor declines associated with cognitive impairment. We extracted gait metrics that characterized gait phases (e.g., stance, swing, foot flat, pushing, and loading ratios), temporal metrics (e.g., stride, stance, and swing duration), and spatial metrics (e.g., strike and liftoff angles, stride length, and minimum toe clearance). Additionally, we included metrics of variability (incl. the variability of heel and toe trajectories), metrics of asymmetry (comparing left and right gait parameters), and dual-task effects. For that purpose, older adults with mobility limitations walked at their preferred speed (single task) and repeated the assessment by including a simultaneous cognitive task (dual task). Older adults with lower cognitive scores were considered to have a cognitive impairment, and their gait parameters were compared with those of older adults without cognitive impairment (control group) using statistical tests. We explored differences between these two groups and investigated which gait parameters measured by the foot-worn inertial sensors would be more relevant to assess cognitive status. For the first time, we investigated these topics in a group of older adults with mobility limitations, a very common condition usually concomitant with cognitive decline in older adults. The results of the study showed that the gait of the group with cognitive impairment was more influenced by the simultaneous execution of the cognitive task than the control group, which led to significant differences between groups on several dual-task effects and some dual-task variability and asymmetry metrics. Several of these metrics also provided acceptable discrimination ability and had a significant association with cognitive scores as measured by standard paper-and-pencil tests. As hypothesized, dual-task gait metrics and dual-task costs had better discrimination ability than single-task gait metrics. Some of the metrics considered in this study had never been explored in the past and proved useful to assess cognitive impairment. Results thus confirm that the newly developed inertial sensor-based gait analysis solution is useful to evaluate gait metrics affected by the cognitive status of older adults based on single- and dual-task gait assessments, which may find important applications in the context of cognitive screening in clinical settings.
- Besides gait metrics, other task performance measures may also be relevant for the detection of MCI. In **Chapter 8**, we investigated the use of various measures, including gait, fingertapping, cognitive tasks, and dual tasks, for detecting Mild Cognitive Impairment (MCI) using a machine learning approach. We collected data from two groups: older adults with confirmed MCI diagnosis (MCI group), and older adults without MCI diagnosis (control group).

The data included inertial sensor data from foot-worn inertial sensors while walking, smart-phone touch events from fingertapping, cognitive task responses (serial subtraction by 7s), and dual-task data. Gait metrics were extracted using the methods described in Chapters 3 and 4. For the first time, and since the dataset was highly imbalanced, we employed one-class classification, or novelty detection, to detect MCI. To account for the minority class (MCI patients), we modeled the majority group (the control group) and evaluated whether the data from MCI patients were correctly classified as anomalies. We compared the performance of the different tasks, as well as the fusion of features (early fusion) and models (late fusion). Additionally, we introduced feature adjustment based on confound control during pipeline training in cross-validation. Despite its common use in statistics, the incorporation of confound control in machine learning experiments is still uncommon and understudied. Our results showed that one-class classification has a higher predictive accuracy for MCI detection, whereas two-class classification performs better in identifying healthy controls. The best performance was observed when using gait metrics alone, followed by the combination of fingertapping and cognitive task features. Dual-task metrics and effects were favored over single-task metrics, consistent with previous studies and our earlier findings. Combining individual test models (gait, fingertapping, and cognitive tasks) yielded better performance than combining the features themselves. Moreover, employing a majority voting approach with one-class classification models resulted in superior classification performance compared to MoCA. In contrast, the combination of two-class classification models failed to recover the low sensitivities attained by the individual models due to the limited representation of the MCI group. Using oversampling techniques like SMOTE to address the imbalanced dataset did not effectively compensate for the low sensitivity in two-class classification. Therefore, our results support the use of one-class classification with confound control to detect abnormal patterns in gait, fingertapping, cognitive, and dual tasks, aiming to improve the early detection of cognitive impairment. Compared to MoCA, the movement analysis and task evaluation tools we employed are simpler, faster, and more objective, making them suitable for screening as they don't require expert scoring.

- Finally, we investigated whether gait and dual tasks could also be useful to detect cognitive impairment in post-stroke patients (**Chapter 9**). Stroke care has evolved substantially over the past years, however, interventions are usually focused on physical disabilities leaving cognitive impairment after stroke undetected and without any intervention. This study used data from a public dataset to explore the detection of cognitive impairment in these patients using a machine learning approach. The dataset included comprehensive neuropsychological assessments that allowed objective categorization of study participants as cognitively impaired or cognitively normal. Moreover, the dataset included assessments of single-task walking, fast walking, and multiple dual-tasks, including walking while performing serial subtractions by 1s, serial subtractions by 7s, and animal nomination. Gait parameters were assessed by two different instruments: an accelerometer that allowed the assessment of sim-

ple gait metrics like speed, cadence, and number of steps, and an electronic walkway that allowed a more complete evaluation, including metrics that characterized the strides and the different gait phases. We used a nested cross-validation approach to evaluate the performance of multiple classifiers and feature selection methods and assessed their performance separately for each considered walking condition. In line with our initial hypothesis, gait characteristics were able to distinguish between post-stroke individuals with and without cognitive impairment. The best classification performance was achieved using the comprehensive set of metrics measured by the electronic walkway, which included dual-task costs of the serial subtractions by 1s. The classification performances were not inferior to commonly used paper-and-pencil tests. Moreover, combining gait metrics with the scores of a paper test improved overall classification performance, achieving better combinations of sensitivity and specificity for the detection of cognitive impairment. Although the performances achieved with the accelerometer were inferior, we note that in this study the accelerometer only provided simple gait evaluation metrics that do not fully characterize the walking performance. As seen from ours and other previous studies, this is not a limitation of the accelerometer itself, but rather a limitation of the algorithms employed to process the data. Our results suggest that gait analysis can be a useful tool for detecting cognitive impairment in patients with cerebrovascular disease, serving as a suitable alternative or complement to paper tests in the screening for cognitive impairment.

Overall, this thesis strengthens the current evidence that motor task performance evaluations can be used for the assessment of cognitive impairment in older adults, which is demonstrated using different groups with different origins, namely i) Swiss older adults with mobility limitations, ii) Portuguese older adults with and without cognitive impairment, and iii) Canadian post-stroke older adults with and without cognitive impairment. Moreover, it shows that inertial sensors can be used to effectively and reliably measure gait, stepping, and dual tasks, being appropriate to support cognitive screening in multiple contexts. Contrary to existing paper-based cognitive screening tests, the approaches proposed in this work allow for a more objective and language-independent evaluation. Because they are easy and quick to perform, dual-task tests may be suitable for screening in a clinical setting, contributing to improving early diagnosis of cognitive impairment. Moreover, the evaluation of stepping and game performance may contribute to more continuous and unsupervised monitoring of the cognitive status of older adults, with interesting applications in the context of remote monitoring. The inertial sensors used allow for the evaluation of movement in both contexts, simultaneously responding to the requirements of the two application scenarios.

10.2 Limitations and Future Work

The field of movement analysis for cognitive screening has traditionally relied on statistical hypothesis testing. However, with the increasing popularity of wearable sensors, researchers are collecting more extensive and higher-dimensional data on human movement, which requires more

advanced analytical techniques to extract relevant features and enable complex modeling. Machine learning is thus emerging as a valuable tool for developing new cognitive screening tools based on movement analysis. To successfully implement these models, however, we need greater access to relevant data. Unfortunately, we found only one public dataset that included gait and dual tasks in neurocognitive disorders, which we investigated in Chapter 9. The remaining datasets used in this work were acquired by us in the context of public-funded projects. However, due to the impact of the Covid-19 pandemic, as well as the fact that data collection in some cases had a primary focus different from our study objectives (as described in Chapters 6 and 7), these datasets did not encompass all the required data dimensions and data size necessary to thoroughly explore all the aspects of this work. Furthermore, the dataset detailed in Chapter 8 was still being collected at the time of writing.

To improve our understanding of movement performance and its relationship with cognition, we would need to include larger and more heterogeneous groups of older adults, including individuals with a range of functioning levels, cognitive abilities, and neurocognitive disorders. Moreover, to ensure the widespread applicability of our findings, we should investigate these topics in a broader range of settings, including individuals from different countries, educational backgrounds, and cultural contexts.

In addition, there is a need to better define and characterize the metrics used to evaluate movement analysis tools. Currently, studies use different evaluation metrics, protocols, and participant groups, making it challenging to compare the performance of different methods. This lack of standardized methodologies and publicly available datasets hinders fair and robust comparisons between methods. Despite a vast and growing body of literature on inertial sensor-based gait analysis, it remains difficult to summarize findings due to the lack of study quality and standardized protocols. Additionally, making more public datasets available would enable researchers to test their methods on a wider range of populations and improve the generalizability of their findings.

Another major limitation in current research is the difficulty in establishing a reliable ground truth for distinguishing normal from abnormal cognitive decline, especially in its earliest stages. Although some studies (including ours) have used paper-and-pencil screening tests to categorize older adults with and without cognitive impairment (as in Chapters 6 and 7), these tools cannot be considered a substitute for more comprehensive neuropsychological assessments (as performed in Chapters 8 and 9). However, even these assessments are not without subjectivity, and no universally accepted criteria exist nowadays for objectively assessing cognitive impairment. Various factors can vary across studies, including the cognitive domains being evaluated, the specific tests used, their grouping criteria, the cutoffs used to demarcate impairment, and the number of tests required to declare a deficit in a certain domain. These factors affect the classification and should be taken into account when designing or comparing studies in this area.

Concerning data analysis and machine learning, there are still many opportunities to explore. For example, incorporating sequential gait parameters extracted from inertial sensors could potentially improve classification results and should be explored in future studies. Additionally, deep learning models could be used to unveil more complex relationships and patterns within the data.

However, deep learning models are often viewed as black boxes, making it difficult to understand how they reach their decisions. This lack of transparency can undermine trust and pose ethical concerns. Therefore, working on explainable, trustworthy, and responsible AI should be prioritized to foster greater trust in these systems and ensure their safe and responsible use.

While we have taken the first steps in analyzing the impact of cognitive tasks on motor performance, future research should also explore the potential for automatic analysis of cognitive tasks, such as through speech analysis, to determine the impact of motor tasks on cognitive performance. Observing hesitations during the cognitive task due to walking, or hesitations during walking due to the cognitive task, can provide additional insight into motor-cognitive interference. Objective assessments of both tasks and their interactions can provide valuable information about an individual's performance.

It is also important to investigate alternative cognitive tasks that can be performed simultaneously with a motor task to maximize dual-task interference and fully understand its effects on motor-cognitive performance. It is possible that different tasks may suit different users in varying ways, but protocols and analysis must be standardized to ensure the reliability of the results. Commonly used cognitive tasks, such as serial subtractions, may heavily rely on individuals' calculation skills, leading to anxiety that can affect test results. To avoid this, cognitive tasks that depend less on speech, such as pressing buttons to answer specific tasks, may be easier to evaluate and create less anxiety, leading to a more objective and reliable assessment of motor-cognitive interference.

Further exploration of movement tests and metrics is also crucial for a comprehensive understanding of motor-cognitive interference. For instance, past studies have not fully explored the utility of dual-task effects to distinguish between older adults with and without cognitive impairment. However, our results have shown that dual-task effects offer great potential for such discrimination. By comparing dual-task performance with baseline (single-task) performance, the effect of individual baseline characteristics can be eliminated. Thus, dual-task effects could provide a more reliable and accurate measure of motor-cognitive performance. Exploring additional movement parameters, beyond spatiotemporal parameters, could also be a valuable area for future research.

In addition to analyzing gait, stepping, fingertapping, and dual tasks, future research should aim to integrate data from multiple sources, including daily living activities, social engagement, and other relevant factors. Longitudinal data, in particular, may provide valuable insights into individual performance over time and aid in distinguishing between normal and abnormal declines. By examining how performance changes over the course of months or years, we may be able to detect personalized patterns of change that are indicative of underlying neurocognitive disorders. A longitudinal analysis of the data may provide more relevant information than a cross-sectional analysis because it can help distinguish between momentary constraints that may affect cognitive performance (e.g., poor sleep or momentary medication effects) and the development of a neurocognitive disorder.

Motor, gait, and cognitive performance evaluations are highly relevant not only in the context

of neurocognitive disorders but also in other scenarios, such as fall prevention. Exploring a wider range of applications could help justify the introduction of wearable sensors and gait analysis in the clinical context. For instance, the cost of acquiring wearables needs to be weighed against the benefits and costs of routinely applying gait analysis in clinical practice. Although science recognizes the importance of gait analysis as a means to improve patients' mobility and independence, its role in clinical care remains controversial. Poor gait speed, for instance, often goes undiagnosed, making impaired mobility invisible. Including gait assessments in clinical practice could stimulate studies of treatable causes of poor gait, improve clinical care, and facilitate the regulatory approval of treatments to improve mobility. Moreover, the potential of mobile gait analysis, both in clinics and beyond, to monitor patients' gait in their daily lives has not been fully exploited.

10.3 Implications and Significance

The rapid development of technology has significantly contributed to advancements in health-related fields, leading to better disease management and care. This progress is particularly relevant for mobility and cognitive impairments, which continue to pose a significant burden on an increasingly aging society. However, despite their impact, these conditions are often underdiagnosed and undertreated.

In geriatric medicine, efforts are being made to mitigate the disparity between life expectancy and the number of years a person can live without experiencing disabilities. As cognitive testing becomes increasingly common, it is important to ensure responsible and accurate testing to prevent inappropriate diagnoses, which can lead to stigma, demoralization, or false reassurance. Neglecting this responsibility can also hinder policymakers' and public health officials' ability to monitor cognitive health, estimate the prevalence and severity of cognitive impairment, and implement appropriate interventions to address them.

The solutions presented in this thesis can have a positive impact in these areas. We anticipate that standardized dual-task assessments and movement analysis based on foot-worn inertial sensors could serve as a quick screening tool for the early detection of cognitive impairment in primary care settings. To evaluate and monitor motor-cognitive function in older adults without supervision, stepping exergames could be employed, making assessments more frequent, affordable, and enjoyable. A pervasive, non-expensive, non-invasive, rapidly administered, and user-friendly solution could be widely used in multiple studies or clinical practices to support clinical decision-making. The technology proposed in this thesis may contribute to the early identification of individuals at risk of cognitive decline, better management of geriatric syndromes, development of new therapies and interventions, selection of individuals who benefit most from interventions, reduced healthcare costs, and promotion of disability-free life expectancy. An integrated approach to evaluating motion and cognition may improve our understanding of cognitive-motor interactions which may contribute to supporting better care and treatment of cognitive impairment.

References

- [1] M. H. Park *et al.*, “Mini-Mental Status Examination as predictors of mortality in the elderly,” *Acta Psychiatrica Scandinavica*, vol. 127, no. 4, pp. 298–304, Apr. 2013. doi:10.1111/j.1600-0447.2012.01918.x
- [2] R. Roberts and D. S. Knopman, “Classification and Epidemiology of MCI,” *Clinics in geriatric medicine*, vol. 29, no. 4, Nov. 2013. doi:10.1016/j.cger.2013.07.003
- [3] W. Zhang *et al.*, “Review of Gait, Cognition, and Fall Risks with Implications for Fall Prevention in Older Adults with Dementia,” *Dementia and Geriatric Cognitive Disorders*, vol. 48, no. 1-2, pp. 17–29, 2019. doi:10.1159/000504340
- [4] M. Sabbagh *et al.*, “Rationale for Early Diagnosis of Mild Cognitive Impairment (MCI) supported by Emerging Digital Technologies,” *The Journal of Prevention of Alzheimer’s Disease*, pp. 1–7, 2020. doi:10.14283/jpad.2020.19
- [5] J. Rasmussen and H. Langerman, “Alzheimer’s Disease – Why We Need Early Diagnosis,” *Degenerative Neurological and Neuromuscular Disease*, vol. Volume 9, pp. 123–130, Dec. 2019. doi:10.2147/DNND.S228939
- [6] L. Lang *et al.*, “Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis,” *BMJ Open*, vol. 7, no. 2, p. e011146, Feb. 2017. doi:10.1136/bmjopen-2016-011146
- [7] P. Maresova *et al.*, “Consequences of chronic diseases and other limitations associated with old age – a scoping review,” *BMC Public Health*, vol. 19, no. 1, p. 1431, Dec. 2019. doi:10.1186/s12889-019-7762-5
- [8] J. Cylus, J. Figueras, and C. Normand, *Will Population Ageing Spell the End of the Welfare State? A review of evidence and policy options*, ser. European Observatory Policy Briefs, A. Sagan, E. Richardson, J. North, and C. White, Eds. Copenhagen (Denmark): European Observatory on Health Systems and Policies, 2019.
- [9] R. Pais, L. Ruano, O. P. Carvalho, and H. Barros, “Global Cognitive Impairment Prevalence and Incidence in Community Dwelling Older Adults—A Systematic Review,” *Geriatrics*, vol. 5, no. 4, p. 84, Oct. 2020. doi:10.3390/geriatrics5040084
- [10] R. C. Petersen and S. Negash, “Mild cognitive impairment: an overview,” *CNS spectrums*, vol. 13, no. 1, pp. 45–53, Jan. 2008. doi:10.1017/s1092852900016151
- [11] J. Hugo and M. Ganguli, “Dementia and Cognitive Impairment: Epidemiology, Diagnosis, and Treatment,” *Clinics in geriatric medicine*, vol. 30, no. 3, pp. 421–442, Aug. 2014. doi:10.1016/j.cger.2014.04.001

- [12] P. S. Sachdev *et al.*, “Classifying neurocognitive disorders: the DSM-5 approach,” *Nature Reviews. Neurology*, vol. 10, no. 11, pp. 634–642, Nov. 2014. doi:10.1038/nrneurol.2014.181
- [13] R. C. Petersen, “Mild Cognitive Impairment,” *Continuum (Minneapolis, Minn.)*, vol. 22, no. 2 Dementia, pp. 404–418, Apr. 2016. doi:10.1212/CON.0000000000000313
- [14] E. Nichols *et al.*, “Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019,” *The Lancet Public Health*, vol. 7, no. 2, pp. 105–125, Feb. 2022. doi:10.1016/S2468-2667(21)00249-8
- [15] I. Santana, F. Farinha, S. Freitas, V. Rodrigues, and A. Carvalho, “The Epidemiology of Dementia and Alzheimer Disease in Portugal: Estimations of Prevalence and Treatment-Costs,” *Acta Medica Portuguesa*, vol. 28, no. 2, pp. 182–188, Apr. 2015. doi:10.20344/amp.6025
- [16] World Health Organization, “Global action plan on the public health response to dementia 2017–2025,” 2017. ISBN 978-92-4-151348-7
- [17] Z. Arvanitakis, R. C. Shah, and D. A. Bennett, “Diagnosis and Management of Dementia: Review,” *JAMA*, vol. 322, no. 16, pp. 1589–1599, 2019. doi:10.1001/jama.2019.4782
- [18] O. Beauchet *et al.*, “Gait analysis in demented subjects: Interests and perspectives,” *Neuropsychiatric Disease and Treatment*, vol. 4, no. 1, pp. 155–160, Feb. 2008. doi:10.2147/ndt.s2070
- [19] M. Montero-Odasso *et al.*, “Consensus on Shared Measures of Mobility and Cognition: From the Canadian Consortium on Neurodegeneration in Aging (CCNA),” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, Jun. 2018. doi:10.1093/gerona/gly148
- [20] L. Boise, M. B. Neal, and J. Kaye, “Dementia assessment in primary care: results from a study in three managed care systems,” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 59, no. 6, pp. M621–626, Jun. 2004. doi:10.1093/gerona/59.6.m621
- [21] P. Maruff *et al.*, “Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer’s disease,” *BMC psychology*, vol. 1, no. 1, p. 30, 2013. doi:10.1186/2050-7283-1-30
- [22] M. Abd Razak *et al.*, “Validity of screening tools for dementia and mild cognitive impairment among the elderly in primary health care: a systematic review,” *Public Health*, vol. 169, pp. 84–92, Apr. 2019. doi:10.1016/j.puhe.2019.01.001
- [23] T. Matsuura *et al.*, “Statistical Analysis of Dual-task Gait Characteristics for Cognitive Score Estimation,” *Scientific Reports*, vol. 9, no. 1, p. 19927, Dec. 2019. doi:10.1038/s41598-019-56485-w
- [24] A. H. Snijders, B. P. van de Warrenburg, N. Giladi, and B. R. Bloem, “Neurological gait disorders in elderly people: clinical approach and classification,” *The Lancet Neurology*, vol. 6, no. 1, pp. 63–74, 2007. doi:10.1016/S1474-4422(06)70678-0

- [25] B. R. Bloem *et al.*, “Idiopathic senile gait disorders are signs of subclinical disease,” *Journal of the American Geriatrics Society*, vol. 48, no. 9, pp. 1098–1101, Sep. 2000. doi:10.1111/j.1532-5415.2000.tb04786.x
- [26] G. Waldemar *et al.*, “Access to diagnostic evaluation and treatment for dementia in Europe,” *International Journal of Geriatric Psychiatry*, vol. 22, no. 1, pp. 47–54, 2007. doi:10.1002/gps.1652
- [27] L. Ruano *et al.*, “Development of a self-administered web-based test for longitudinal cognitive assessment,” *Scientific Reports*, vol. 6, p. 19114, Jan. 2016. doi:10.1038/srep19114
- [28] T. Tong, M. Chignell, M. C. Tierney, and J. Lee, “A Serious Game for Clinical Assessment of Cognitive Status: Validation Study,” *JMIR serious games*, vol. 4, no. 1, p. e7, May 2016. doi:10.2196/games.5006
- [29] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, “Spoken Language Derived Measures for Detecting Mild Cognitive Impairment,” *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 7, pp. 2081–2090, Sep. 2011. doi:10.1109/TASL.2011.2112351
- [30] A. König *et al.*, “Automatic speech analysis for the assessment of patients with predementia and Alzheimer’s disease,” *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112–124, Mar. 2015. doi:10.1016/j.dadm.2014.11.012
- [31] A. Akl, J. Snoek, and A. Mihailidis, “Unobtrusive Detection of Mild Cognitive Impairment in Older Adults Through Home Monitoring,” *IEEE journal of biomedical and health informatics*, vol. 21, no. 2, pp. 339–348, 2017. doi:10.1109/JBHI.2015.2512273
- [32] P. Dagum, “Digital biomarkers of cognitive function,” *npj Digital Medicine*, vol. 1, no. 1, pp. 1–3, Mar. 2018. doi:10.1038/s41746-018-0018-4
- [33] B. M. Koo and L. M. Vizer, “Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review,” *Innovation in Aging*, vol. 3, no. 1, Jan. 2019. doi:10.1093/geroni/igy038
- [34] Alzheimer’s Association, “2019 Alzheimer’s disease facts and figures,” *Alzheimer’s & Dementia*, vol. 15, no. 3, pp. 321–387, Mar. 2019. doi:10.1016/j.jalz.2019.01.010
- [35] W. Pirker and R. Katzenschlager, “Gait disorders in adults and the elderly,” *Wiener Klinische Wochenschrift*, vol. 129, no. 3, pp. 81–95, 2017. doi:10.1007/s00508-016-1096-4
- [36] M. Montero-Odasso, J. Verghese, O. Beauchet, and J. M. Hausdorff, “Gait and Cognition: A Complementary Approach to Understanding Brain Function and the Risk of Falling,” *Journal of the American Geriatrics Society*, vol. 60, no. 11, pp. 2127–2136, Nov. 2012. doi:10.1111/j.1532-5415.2012.04209.x
- [37] T. Buracchio, H. Dodge, D. Howieson, D. Wasserman, and J. Kaye, “The trajectory of gait speed preceding MCI,” *Archives of neurology*, vol. 67, no. 8, pp. 980–986, Aug. 2010. doi:10.1001/archneurol.2010.159
- [38] Q. Tian *et al.*, “The brain map of gait variability in aging, cognitive impairment and dementia-A systematic review,” *Neuroscience and Biobehavioral Reviews*, vol. 74, no. Pt A, pp. 149–162, Mar. 2017. doi:10.1016/j.neubiorev.2017.01.020

- [39] O. Beauchet *et al.*, “Motor phenotype of decline in cognitive performance among community-dwellers without dementia: population-based study and meta-analysis,” *PloS One*, vol. 9, no. 6, p. e99318, 2014. doi:10.1371/journal.pone.0099318
- [40] —, “Spatiotemporal Gait Characteristics Associated with Cognitive Impairment: A Multicenter Cross-Sectional Study, the Intercontinental “Gait, cOgnitiOn & Decline” Initiative,” *Current Alzheimer Research*, vol. 15, no. 3, pp. 273–282, 2018. doi:10.2174/1567205014666170725125621
- [41] I. Mulas *et al.*, “Clinical assessment of gait and functional mobility in Italian healthy and cognitively impaired older persons using wearable inertial sensors,” *Aging Clinical and Experimental Research*, vol. 33, no. 7, pp. 1853–1864, Jul. 2021. doi:10.1007/s40520-020-01715-9
- [42] J. Wilson, L. Allcock, R. Mc Ardle, J.-P. Taylor, and L. Rochester, “The neural correlates of discrete gait characteristics in ageing: A Structured Review,” *Neuroscience & Biobehavioral Reviews*, vol. 100, Dec. 2018. doi:10.1016/j.neubiorev.2018.12.017
- [43] P. L. Sheridan and J. M. Hausdorff, “The role of higher-level cognitive function in gait: executive dysfunction contributes to fall risk in Alzheimer’s disease,” *Dementia and Geriatric Cognitive Disorders*, vol. 24, no. 2, pp. 125–137, 2007. doi:10.1159/000105126
- [44] O. Beauchet *et al.*, “Gait control: a specific subdomain of executive function?” *Journal of NeuroEngineering and Rehabilitation*, vol. 9, p. 12, Feb. 2012. doi:10.1186/1743-0003-9-12
- [45] R. Morris, S. Lord, J. Bunce, D. Burn, and L. Rochester, “Gait and cognition: Mapping the global and discrete relationships in ageing and neurodegenerative disease,” *Neuroscience & Biobehavioral Reviews*, vol. 64, pp. 326–345, May 2016. doi:10.1016/j.neubiorev.2016.02.012
- [46] V. Valkanova and K. P. Ebmeier, “What can gait tell us about dementia? Review of epidemiological and neuropsychological evidence,” *Gait & Posture*, vol. 53, pp. 215–223, 2017. doi:10.1016/j.gaitpost.2017.01.024
- [47] V. Valkanova *et al.*, “Association between gait and cognition in an elderly population based sample,” *Gait & Posture*, vol. 65, pp. 240–245, Sep. 2018. doi:10.1016/j.gaitpost.2018.07.178
- [48] R. Parihar, J. R. Mahoney, and J. Verghese, “Relationship of gait and cognition in the elderly,” *Current translational geriatrics and experimental gerontology reports*, vol. 2, no. 3, Sep. 2013. doi:10.1007/s13670-013-0052-7
- [49] J. K. Chhetri, P. Chan, B. Vellas, and M. Cesari, “Motoric Cognitive Risk Syndrome: Predictor of Dementia and Age-Related Negative Outcomes,” *Frontiers in Medicine*, vol. 4, Oct. 2017. doi:10.3389/fmed.2017.00166
- [50] H. Sekhon *et al.*, “Motoric cognitive risk syndrome, incident cognitive impairment and morphological brain abnormalities: Systematic review and meta-analysis,” *Maturitas*, vol. 123, pp. 45–54, May 2019. doi:10.1016/j.maturitas.2019.02.006
- [51] M. Montero-Odasso *et al.*, “The motor signature of mild cognitive impairment: results from the gait and brain study,” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 69, no. 11, pp. 1415–1421, Nov. 2014. doi:10.1093/gerona/glu155

- [52] Q. Yang *et al.*, “Gait Change in Dual Task as a Behavioral Marker to Detect Mild Cognitive Impairment in Elderly Persons: A Systematic Review and Meta-analysis,” *Archives of Physical Medicine and Rehabilitation*, vol. 101, no. 10, pp. 1813–1821, Oct. 2020. doi:10.1016/j.apmr.2020.05.020
- [53] L. Bahureksa *et al.*, “The Impact of Mild Cognitive Impairment on Gait and Balance: A Systematic Review and Meta-Analysis of Studies Using Instrumented Assessment,” *Gerontology*, vol. 63, no. 1, pp. 67–83, 2017. doi:10.1159/000445831
- [54] M. Bayot *et al.*, “The interaction between cognition and motor control: A theoretical framework for dual-task interference effects on posture, gait initiation, gait and turning,” *Neurophysiologie Clinique*, vol. 48, no. 6, pp. 361–375, Dec. 2018. doi:10.1016/j.neucli.2018.10.003
- [55] D. A. Lowe, R. K. MacAulay, D. M. Szeles, N. J. Milano, and M. T. Wagner, “Dual-Task Gait Assessment in a Clinical Sample: Implications for Improved Detection of Mild Cognitive Impairment,” *The Journals of Gerontology: Series B, Psychological sciences and social sciences*, vol. 75, pp. 1372–1381, 2019. doi:10.1093/geronb/gbz119
- [56] J. Lee and S. Park, “Effects of a priority-based dual task on gait velocity and variability in older adults with mild cognitive impairment,” *Journal of Exercise Rehabilitation*, vol. 14, no. 6, pp. 993–997, Dec. 2018. doi:10.12965/jer.1836402.201
- [57] 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,” *Annals of Physical and Rehabilitation Medicine*, vol. 52, no. 6, pp. 453–474, Jul. 2009. doi:10.1016/j.rehab.2008.10.004
- [58] C. Buckley *et al.*, “The Role of Movement Analysis in Diagnosing and Monitoring Neurodegenerative Conditions: Insights from Gait and Postural Control,” *Brain Sciences*, vol. 9, no. 2, p. 34, Feb. 2019. doi:10.3390/brainsci9020034
- [59] A.-M. De Cock *et al.*, “Gait characteristics under different walking conditions: Association with the presence of cognitive impairment in community-dwelling older people,” *PloS One*, vol. 12, no. 6, p. e0178566, 2017. doi:10.1371/journal.pone.0178566
- [60] W. Hao *et al.*, “Association of gait with global cognitive function and cognitive domains detected by MoCA-J among community-dwelling older adults: a cross-sectional study,” *BMC Geriatrics*, vol. 21, no. 1, p. 523, Dec. 2021. doi:10.1186/s12877-021-02467-5
- [61] A. Godfrey *et al.*, “Inertial wearables as pragmatic tools in dementia,” *Maturitas*, vol. 127, pp. 12–17, Sep. 2019. doi:10.1016/j.maturitas.2019.05.010
- [62] F. Petraglia *et al.*, “Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis,” *European Journal of Physical and Rehabilitation Medicine*, Oct. 2018. doi:10.23736/S1973-9087.18.05306-6
- [63] L. C. Kourtis, O. B. Regele, J. M. Wright, and G. B. Jones, “Digital biomarkers for Alzheimer’s disease: the mobile/wearable devices opportunity,” *npj Digital Medicine*, vol. 2, no. 1, p. 9, Feb. 2019. doi:10.1038/s41746-019-0084-2
- [64] R. Mc Ardle, S. Del Din, B. Galna, A. Thomas, and L. Rochester, “Differentiating dementia disease subtypes with gait analysis: feasibility of wearable sensors?” *Gait & Posture*, vol. 76, pp. 372–376, Feb. 2020. doi:10.1016/j.gaitpost.2019.12.028

- [65] D. K. Ravi *et al.*, “Revealing the optimal thresholds for movement performance: A systematic review and meta-analysis to benchmark pathological walking behaviour,” *Neuroscience & Biobehavioral Reviews*, vol. 108, pp. 24–33, Jan. 2020. doi:10.1016/j.neubiorev.2019.10.008
- [66] M. M. Chiudo *et al.*, “Mapping features and patterns of accelerometry data on human movement in different age groups and associated health problems: A cross-sectional study,” *Experimental Gerontology*, vol. 168, p. 111949, Oct. 2022. doi:10.1016/j.exger.2022.111949
- [67] S. Musich, S. S. Wang, J. Ruiz, K. Hawkins, and E. Wicker, “The impact of mobility limitations on health outcomes among older adults,” *Geriatric Nursing*, vol. 39, no. 2, pp. 162–169, Mar. 2018. doi:10.1016/j.gerinurse.2017.08.002
- [68] G. Netuveli and D. Blane, “Quality of life in older ages,” *British Medical Bulletin*, vol. 85, no. 1, pp. 113–126, Feb. 2008. doi:10.1093/bmb/ldn003
- [69] J. C. Cornman, D. A. Gleib, and M. Weinstein, “Change in Mobility: Consistency of Estimates and Predictors Across Studies of Older Adults,” *The Journals of Gerontology: Series B*, vol. 76, no. 1, pp. 209–218, Jan. 2021. doi:10.1093/geronb/gbz091
- [70] Y. Celik, S. Stuart, W. Woo, and A. Godfrey, “Gait analysis in neurological populations: Progression in the use of wearables,” *Medical Engineering & Physics*, vol. 87, pp. 9–29, Jan. 2021. doi:10.1016/j.medengphy.2020.11.005
- [71] C. R. Kovacs, “Age-Related Changes in Gait and Obstacle Avoidance Capabilities in Older Adults: A Review,” *Journal of Applied Gerontology*, vol. 24, no. 1, pp. 21–34, Feb. 2005. doi:10.1177/0733464804271279
- [72] P. Mahlknecht *et al.*, “Prevalence and burden of gait disorders in elderly men and women aged 60-97 years: a population-based study,” *PloS One*, vol. 8, no. 7, p. e69627, 2013. doi:10.1371/journal.pone.0069627
- [73] A. Fasano and B. R. Bloem, “Gait Disorders,” *CONTINUUM: Lifelong Learning in Neurology*, vol. 19, pp. 1344–1382, Oct. 2013. doi:10.1212/01.CON.0000436159.33447.69
- [74] J. A. Cohen, J. Verghese, and J. L. Zwerling, “Cognition and gait in older people,” *Maturitas*, vol. 93, pp. 73–77, Nov. 2016. doi:10.1016/j.maturitas.2016.05.005
- [75] H. Shimada *et al.*, “Relationship between Age-Associated Changes of Gait and Falls and Life-Space in Elderly People,” *Journal of Physical Therapy Science*, vol. 22, no. 4, pp. 419–424, 2010. doi:10.1589/jpts.22.419
- [76] B. Salzman, “Gait and balance disorders in older adults,” *American Family Physician*, vol. 82, no. 1, pp. 61–68, Jul. 2010.
- [77] K. A. Shorter, A. Wu, and A. D. Kuo, “The high cost of swing leg circumduction during human walking,” *Gait & Posture*, vol. 54, pp. 265–270, May 2017. doi:10.1016/j.gaitpost.2017.03.021
- [78] D. Jarchi *et al.*, “A Review on Accelerometry-Based Gait Analysis and Emerging Clinical Applications,” *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 177–194, 2018. doi:10.1109/RBME.2018.2807182

- [79] F. Dadashi *et al.*, “Gait and Foot Clearance Parameters Obtained Using Shoe-Worn Inertial Sensors in a Large-Population Sample of Older Adults,” *Sensors (Basel, Switzerland)*, vol. 14, no. 1, pp. 443–457, Dec. 2013. doi:10.3390/s140100443
- [80] U. Lindemann, “Spatiotemporal gait analysis of older persons in clinical practice and research : Which parameters are relevant?” *Zeitschrift Fur Gerontologie Und Geriatrie*, Feb. 2019. doi:10.1007/s00391-019-01520-8
- [81] U. Lindemann *et al.*, “Distance to achieve steady state walking speed in frail elderly persons,” *Gait & Posture*, vol. 27, no. 1, pp. 91–96, Jan. 2008. doi:10.1016/j.gaitpost.2007.02.005
- [82] M. Błażkiewicz, I. Wiszomirska, and A. Wit, “Comparison of four methods of calculating the symmetry of spatial-temporal parameters of gait,” *Acta of Bioengineering and Biomechanics*, vol. 16, no. 1, pp. 29–35, 2014.
- [83] S. Studenski *et al.*, “Gait speed and survival in older adults,” *JAMA*, vol. 305, no. 1, pp. 50–58, Jan. 2011. doi:10.1001/jama.2010.1923
- [84] G. Grande *et al.*, “Measuring gait speed to better identify prodromal dementia,” *Experimental Gerontology*, vol. 124, p. 110625, Sep. 2019. doi:10.1016/j.exger.2019.05.014
- [85] R. W. Bohannon and A. Williams Andrews, “Normal walking speed: a descriptive meta-analysis,” *Physiotherapy*, vol. 97, no. 3, pp. 182–189, Sep. 2011. doi:10.1016/j.physio.2010.12.004
- [86] J. A. Schrack, V. Zipunnikov, E. M. Simonsick, S. Studenski, and L. Ferrucci, “Rising Energetic Cost of Walking Predicts Gait Speed Decline With Aging,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 71, no. 7, pp. 947–953, Jul. 2016. doi:10.1093/gerona/glw002
- [87] R. W. Bohannon and S. S. Glenney, “Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review,” *Journal of Evaluation in Clinical Practice*, vol. 20, no. 4, pp. 295–300, Aug. 2014. doi:10.1111/jep.12158
- [88] S. Studenski *et al.*, “Physical Performance Measures in the Clinical Setting,” *Journal of the American Geriatrics Society*, vol. 51, no. 3, pp. 314–322, Mar. 2003. doi:10.1046/j.1532-5415.2003.51104.x
- [89] L. Ferrucci *et al.*, “Disease severity and health-related quality of life across different chronic conditions,” *Journal of the American Geriatrics Society*, vol. 48, no. 11, pp. 1490–1495, Nov. 2000. doi:10.1111/jgs.2000.48.11.1490
- [90] Y. Schoon, K. Bongers, J. Van Kempen, R. Melis, and M. Olde Rikkert, “Gait speed as a test for monitoring frailty in community-dwelling older people has the highest diagnostic value compared to step length and chair rise time,” *European Journal of Physical and Rehabilitation Medicine*, vol. 50, no. 6, pp. 693–701, Dec. 2014.
- [91] B. E. Maki, “Gait changes in older adults: predictors of falls or indicators of fear,” *Journal of the American Geriatrics Society*, vol. 45, no. 3, pp. 313–320, Mar. 1997. doi:10.1111/j.1532-5415.1997.tb00946.x

- [92] N. R. Marques *et al.*, “Is it possible to predict falls in older adults using gait kinematics?” *Clinical Biomechanics*, vol. 59, pp. 15–18, Nov. 2018. doi:10.1016/j.clinbiomech.2018.08.006
- [93] M. M. Mielke *et al.*, “Assessing the Temporal Relationship Between Cognition and Gait: Slow Gait Predicts Cognitive Decline in the Mayo Clinic Study of Aging,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 68, no. 8, pp. 929–937, Aug. 2013. doi:10.1093/gerona/gls256
- [94] S. R. Cummings, S. Studenski, and L. Ferrucci, “A Diagnosis of Dismobility—Giving Mobility Clinical Visibility,” *JAMA*, vol. 311, no. 20, pp. 2061–2062, May 2014. doi:10.1001/jama.2014.3033
- [95] B. Bogen, R. Moe-Nilssen, A. H. Ranhoff, and M. K. Aaslund, “The walk ratio: Investigation of invariance across walking conditions and gender in community-dwelling older people,” *Gait & Posture*, vol. 61, pp. 479–482, Mar. 2018. doi:10.1016/j.gaitpost.2018.02.019
- [96] E. Nordin, R. Moe-Nilssen, A. Ramnemark, and L. Lundin-Olsson, “Changes in step-width during dual-task walking predicts falls,” *Gait & Posture*, vol. 32, no. 1, pp. 92–97, May 2010. doi:10.1016/j.gaitpost.2010.03.012
- [97] D. Hamacher, D. Hamacher, and L. Schega, “Towards the importance of minimum toe clearance in level ground walking in a healthy elderly population,” *Gait & Posture*, vol. 40, no. 4, pp. 727–729, Sep. 2014. doi:10.1016/j.gaitpost.2014.07.016
- [98] J. Verghese, R. Holtzer, R. B. Lipton, and C. Wang, “Quantitative Gait Markers and Incident Fall Risk in Older Adults,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 64A, no. 8, pp. 896–901, Aug. 2009. doi:10.1093/gerona/glp033
- [99] J. M. Hausdorff, D. A. Rios, and H. K. Edelberg, “Gait variability and fall risk in community-living older adults: a 1-year prospective study,” *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 8, pp. 1050–1056, Aug. 2001. doi:10.1053/apmr.2001.24893
- [100] S. A. Bridenbaugh and R. W. Kressig, “Laboratory Review: The Role of Gait Analysis in Seniors’ Mobility and Fall Prevention,” *Gerontology*, vol. 57, no. 3, pp. 256–264, 2011. doi:10.1159/000322194
- [101] M. Saleh and G. Murdoch, “In defence of gait analysis. Observation and measurement in gait assessment,” *The Journal of Bone and Joint Surgery. British Volume*, vol. 67, no. 2, pp. 237–241, Mar. 1985. doi:10.1302/0301-620X.67B2.3980533
- [102] D. E. Krebs, J. E. Edelstein, and S. Fishman, “Reliability of Observational Kinematic Gait Analysis,” *Physical Therapy*, vol. 65, no. 7, pp. 1027–1033, Jul. 1985. doi:10.1093/ptj/65.7.1027
- [103] M. Akhtaruzzaman, A. A. Shafie, and M. R. Khan, “Gait analysis: systems, technologies, and importance,” *Journal of Mechanics in Medicine and Biology*, vol. 16, no. 07, p. 1630003, Mar. 2016. doi:10.1142/S0219519416300039
- [104] N. B. Alexander and A. Goldberg, “Clinical gait and stepping performance measures in older adults,” *European Review of Aging and Physical Activity*, vol. 3, no. 1, p. 20, May 2006. doi:10.1007/s11556-006-0001-y

- [105] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, "Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1521–1537, Nov. 2016. doi:10.1109/JBHI.2016.2608720
- [106] O. Beauchet *et al.*, "Guidelines for Assessment of Gait and Reference Values for Spatiotemporal Gait Parameters in Older Adults: The Biomathics and Canadian Gait Consortiums Initiative," *Frontiers in Human Neuroscience*, vol. 11, 2017. doi:10.3389/fnhum.2017.00353
- [107] R. Długosz, J. Pauk, and P.-A. Farine, "New Trends in Motion Capture Systems for Human Gait Analysis," *Machine Graphics & Vision International Journal*, vol. 20, no. 3, pp. 319–331, Mar. 2011.
- [108] K. Aminian, B. Najafi, C. Büla, P. F. Leyvraz, and P. Robert, "Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes," *Journal of Biomechanics*, vol. 35, no. 5, pp. 689–699, May 2002. doi:10.1016/S0021-9290(02)00008-8
- [109] A. Muro-de-la Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications," *Sensors (Basel, Switzerland)*, vol. 14, no. 2, pp. 3362–3394, Feb. 2014. doi:10.3390/s140203362
- [110] J. R. Rebula, L. V. Ojeda, P. G. Adamczyk, and A. D. Kuo, "Measurement of foot placement and its variability with inertial sensors," *Gait & Posture*, vol. 38, no. 4, pp. 974–980, Sep. 2013. doi:10.1016/j.gaitpost.2013.05.012
- [111] Z. Jamali and S. Behzadipour, "Quantitative evaluation of parameters affecting the accuracy of Microsoft Kinect in GAIT analysis," in *2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME)*. Tehran, Iran: IEEE, 2016. doi:10.1109/ICBME.2016.7890977. ISBN 978-1-5090-3452-9 pp. 306–311.
- [112] S. Springer and G. Yogev Seligmann, "Validity of the Kinect for Gait Assessment: A Focused Review," *Sensors*, vol. 16, no. 2, p. 194, Feb. 2016. doi:10.3390/s16020194
- [113] R. Lun and W. Zhao, "A Survey of Applications and Human Motion Recognition with Microsoft Kinect," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 05, p. 1555008, Aug. 2015. doi:10.1142/S0218001415550083
- [114] G. Faity, D. Mottet, and J. Froger, "Validity and Reliability of Kinect v2 for Quantifying Upper Body Kinematics during Seated Reaching," *Sensors*, vol. 22, no. 7, p. 2735, Apr. 2022. doi:10.3390/s22072735
- [115] A. L. McDonough, M. Batavia, F. C. Chen, S. Kwon, and J. Ziai, "The validity and reliability of the GAITRite system's measurements: A preliminary evaluation," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 3, pp. 419–425, Mar. 2001. doi:10.1053/apmr.2001.19778
- [116] C. J. Shanahan *et al.*, "Technologies for Advanced Gait and Balance Assessments in People with Multiple Sclerosis," *Frontiers in Neurology*, vol. 8, p. 708, Feb. 2018. doi:10.3389/fneur.2017.00708
- [117] T. Karcnik, "Using motion analysis data for foot-floor contact detection," *Medical & Biological Engineering & Computing*, vol. 41, no. 5, pp. 509–512, Sep. 2003. doi:10.1007/BF02345310

- [118] F. Kluge *et al.*, “Towards Mobile Gait Analysis: Concurrent Validity and Test-Retest Reliability of an Inertial Measurement System for the Assessment of Spatio-Temporal Gait Parameters,” *Sensors*, vol. 17, no. 7, p. 1522, Jun. 2017. doi:10.3390/s17071522
- [119] I. H. Lopez-Nava and A. Munoz-Melendez, “Wearable Inertial Sensors for Human Motion Analysis: A Review,” *IEEE Sensors Journal*, vol. 16, no. 22, pp. 7821–7834, Nov. 2016. doi:10.1109/JSEN.2016.2609392
- [120] F. M. Rast and R. Labruyère, “Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments,” *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 148, Dec. 2020. doi:10.1186/s12984-020-00779-y
- [121] D. Trojaniello *et al.*, “Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait,” *Journal of Neuroengineering and Rehabilitation*, vol. 11, p. 152, Nov. 2014. doi:10.1186/1743-0003-11-152
- [122] M. Bertoli *et al.*, “Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: multicenter validation among Parkinson, mildly cognitively impaired and healthy older adults,” *Biomedical Engineering Online*, vol. 17, no. 1, p. 58, May 2018. doi:10.1186/s12938-018-0488-2
- [123] A. Vienne, R. P. Barrois, S. Buffat, D. Ricard, and P.-P. Vidal, “Inertial Sensors to Assess Gait Quality in Patients with Neurological Disorders: A Systematic Review of Technical and Analytical Challenges,” *Frontiers in Psychology*, vol. 8, May 2017. doi:10.3389/fpsyg.2017.00817
- [124] P. Esser, H. Dawes, J. Collett, M. G. Feltham, and K. Howells, “Assessment of spatio-temporal gait parameters using inertial measurement units in neurological populations,” *Gait & Posture*, vol. 34, no. 4, pp. 558–560, Oct. 2011. doi:10.1016/j.gaitpost.2011.06.018
- [125] A. Ferrari *et al.*, “A Mobile Kalman-Filter Based Solution for the Real-Time Estimation of Spatio-Temporal Gait Parameters,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 7, pp. 764–773, Jul. 2016. doi:10.1109/TNSRE.2015.2457511
- [126] R. W. Motl *et al.*, “Accelerometry as a measure of walking behavior in multiple sclerosis,” *Acta Neurologica Scandinavica*, vol. 127, no. 6, pp. 384–390, Jun. 2013. doi:10.1111/ane.12036
- [127] H. Lau and K. Tong, “The reliability of using accelerometer and gyroscope for gait event identification on persons with dropped foot,” *Gait & Posture*, vol. 27, no. 2, pp. 248–257, Feb. 2008. doi:10.1016/j.gaitpost.2007.03.018
- [128] B. Vigna *et al.*, “Low Cost Silicon Coriolis’ Gyroscope Paves the Way to Consumer IMU,” in *Advanced Materials and Technologies for Micro/Nano-Devices, Sensors and Actuators*. Dordrecht: Springer Netherlands, 2010, pp. 67–74. ISBN 978-90-481-3805-0 978-90-481-3807-4
- [129] G. Krishnan, C. U Kshirsagar, G. K Ananthasuresh, and N. Bhat, “Micromachined High-Resolution Accelerometers,” *Journal of the Indian Institute of Science*, vol. 87, no. 3, 2007.

- [130] W. A. Gill, I. Howard, I. Mazhar, and K. McKee, "A Review of MEMS Vibrating Gyroscopes and Their Reliability Issues in Harsh Environments," *Sensors*, vol. 22, no. 19, p. 7405, Sep. 2022. doi:10.3390/s22197405
- [131] B. Mariani *et al.*, "3D gait assessment in young and elderly subjects using foot-worn inertial sensors," *Journal of Biomechanics*, vol. 43, no. 15, pp. 2999–3006, Nov. 2010. doi:10.1016/j.jbiomech.2010.07.003
- [132] B. Mariani, M. C. Jiménez, F. J. G. Vingerhoets, and K. Aminian, "On-Shoe Wearable Sensors for Gait and Turning Assessment of Patients With Parkinson's Disease," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 155–158, Jan. 2013. doi:10.1109/TBME.2012.2227317
- [133] B. Mariani, S. Rochat, C. J. Büla, and K. Aminian, "Heel and Toe Clearance Estimation for Gait Analysis Using Wireless Inertial Sensors," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 11, pp. 3162–3168, Nov. 2012. doi:10.1109/TBME.2012.2216263
- [134] W. Zijlstra and A. L. Hof, "Assessment of spatio-temporal gait parameters from trunk accelerations during human walking," *Gait & Posture*, vol. 18, no. 2, pp. 1–10, Oct. 2003. doi:10.1016/S0966-6362(02)00190-X
- [135] Z. Wang, S. Qiu, Z. Cao, and M. Jiang, "Quantitative assessment of dual gait analysis based on inertial sensors with body sensor network," *Sensor Review*, vol. 33, Jan. 2013. doi:10.1108/02602281311294342
- [136] A. M. Sabatini, "Quaternion-based strap-down integration method for applications of inertial sensing to gait analysis," *Medical and Biological Engineering and Computing*, vol. 43, no. 1, pp. 94–101, Feb. 2005. doi:10.1007/BF02345128
- [137] J. J. Kavanagh and H. B. Menz, "Accelerometry: a technique for quantifying movement patterns during walking," *Gait & Posture*, vol. 28, no. 1, pp. 1–15, Jul. 2008. doi:10.1016/j.gaitpost.2007.10.010
- [138] R. Caldas, M. Mundt, W. Potthast, F. Buarque de Lima Neto, and B. Markert, "A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms," *Gait & Posture*, vol. 57, pp. 204–210, Sep. 2017. doi:10.1016/j.gaitpost.2017.06.019
- [139] M. S. H. Aung *et al.*, "Automated detection of instantaneous gait events using time frequency analysis and manifold embedding," *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 21, no. 6, pp. 908–916, Nov. 2013. doi:10.1109/TNSRE.2013.2239313
- [140] J. Taborri, S. Rossi, E. Palermo, F. Patanè, and P. Cappa, "A novel HMM distributed classifier for the detection of gait phases by means of a wearable inertial sensor network," *Sensors (Basel, Switzerland)*, vol. 14, no. 9, pp. 16 212–16 234, Sep. 2014. doi:10.3390/s140916212
- [141] U. Lindemann, "Spatiotemporal gait analysis of older persons in clinical practice and research: Which parameters are relevant?" *Zeitschrift für Gerontologie und Geriatrie*, vol. 53, no. 2, pp. 171–178, Mar. 2020. doi:10.1007/s00391-019-01520-8
- [142] A. Rampp *et al.*, "Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients," *IEEE transactions on bio-medical engineering*, vol. 62, no. 4, pp. 1089–1097, Apr. 2015. doi:10.1109/TBME.2014.2368211

- [143] K. Hori *et al.*, “Inertial Measurement Unit-Based Estimation of Foot Trajectory for Clinical Gait Analysis,” *Frontiers in Physiology*, vol. 10, p. 1530, Jan. 2020. doi:10.3389/fphys.2019.01530
- [144] J. Hannink *et al.*, “Mobile Stride Length Estimation With Deep Convolutional Neural Networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 354–362, Mar. 2018. doi:10.1109/JBHI.2017.2679486
- [145] A. R. Anwary, H. Yu, and M. Vassallo, “Optimal Foot Location for Placing Wearable IMU Sensors and Automatic Feature Extraction for Gait Analysis,” *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2555–2567, Mar. 2018. doi:10.1109/JSEN.2017.2786587
- [146] A. Peruzzi, U. Della Croce, and A. Cereatti, “Estimation of stride length in level walking using an inertial measurement unit attached to the foot: A validation of the zero velocity assumption during stance,” *Journal of Biomechanics*, vol. 44, no. 10, pp. 1991–1994, Jul. 2011. doi:10.1016/j.jbiomech.2011.04.035
- [147] J. C. Pérez-Ibarra, H. Williams, A. A. G. Siqueira, and H. I. Krebs, “Real-Time Identification of Impaired Gait Phases Using a Single Foot-Mounted Inertial Sensor: Review and Feasibility Study,” in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Aug. 2018. doi:10.1109/BIOROB.2018.8487694 pp. 1157–1162.
- [148] B. M. Ben-David, G. Malkin, and H. Erel, “Ageism and Neuropsychological Tests,” in *Contemporary Perspectives on Ageism*, ser. International Perspectives on Aging, L. Ayalon and C. Tesch-Römer, Eds. Cham: Springer International Publishing, 2018, pp. 277–297. ISBN 978-3-319-73820-8
- [149] I. E. M. Evans *et al.*, “Social isolation, cognitive reserve, and cognition in healthy older people,” *PLoS ONE*, vol. 13, no. 8, Aug. 2018. doi:10.1371/journal.pone.0201008
- [150] Committee on the Public Health Dimensions of Cognitive Aging, Board on Health Sciences Policy, and Institute of Medicine, *Cognitive Aging: Progress in Understanding and Opportunities for Action*, ser. The National Academies Collection: Reports funded by National Institutes of Health, D. G. Blazer, K. Yaffe, and C. T. Liverman, Eds. Washington (DC): National Academies Press (US), 2015.
- [151] D. L. Murman, “The Impact of Age on Cognition,” *Seminars in Hearing*, vol. 36, no. 3, pp. 111–121, Aug. 2015. doi:10.1055/s-0035-1555115
- [152] R. Peters, “Ageing and the brain,” *Postgraduate Medical Journal*, vol. 82, no. 964, pp. 84–88, Feb. 2006. doi:10.1136/pgmj.2005.036665
- [153] A. M. Gorman, “Neuronal cell death in neurodegenerative diseases: recurring themes around protein handling,” *Journal of Cellular and Molecular Medicine*, vol. 12, no. 6a, pp. 2263–2280, Dec. 2008. doi:10.1111/j.1582-4934.2008.00402.x
- [154] W. M. McDonald, “Overview of Neurocognitive Disorders,” *Focus*, vol. 15, no. 1, pp. 4–12, Jan. 2017. doi:10.1176/appi.focus.20160030
- [155] B. Winblad *et al.*, “Mild cognitive impairment - beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment,” *Journal of Internal Medicine*, vol. 256, no. 3, pp. 240–246, Sep. 2004. doi:10.1111/j.1365-2796.2004.01380.x

- [156] T. Foley, A. Jennings, and G. Swanwick, “Dementia: Diagnosis & Management in General Practice,” Irish College of General Practitioners: Quality and Safety in Practice Committee, Tech. Rep., 2019.
- [157] N. Al-Qazzaz, S. Ali, S. A. Ahmad, S. Islam, and K. Mohamad, “Cognitive impairment and memory dysfunction after a stroke diagnosis: a post-stroke memory assessment,” *Neuropsychiatric Disease and Treatment*, p. 1677, Sep. 2014. doi:10.2147/NDT.S67184
- [158] M. Dichgans and D. Leys, “Vascular Cognitive Impairment,” *Circulation Research*, vol. 120, no. 3, pp. 573–591, Feb. 2017. doi:10.1161/CIRCRESAHA.116.308426
- [159] R. N. Kalaria *et al.*, “Alzheimer’s disease and vascular dementia in developing countries: prevalence, management, and risk factors,” *The Lancet Neurology*, vol. 7, no. 9, pp. 812–826, Sep. 2008. doi:10.1016/S1474-4422(08)70169-8
- [160] L. Gjøra *et al.*, “Current and Future Prevalence Estimates of Mild Cognitive Impairment, Dementia, and Its Subtypes in a Population-Based Sample of People 70 Years and Older in Norway: The HUNT Study,” *Journal of Alzheimer’s Disease*, vol. 79, no. 3, pp. 1213–1226, Feb. 2021. doi:10.3233/JAD-201275
- [161] A. Hategan, J. A. Bourgeois, T. Cheng, and J. Young, “Common Major and Mild Neurocognitive Disorders: Alzheimer Disease, Frontotemporal, Lewy Body, and Vascular Types,” in *Geriatric Psychiatry Study Guide*. Cham: Springer International Publishing, 2018, pp. 203–242. ISBN 978-3-319-77127-4 978-3-319-77128-1
- [162] R. Li *et al.*, “Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions,” *Journal of Biomedical Informatics*, vol. 127, p. 104030, Mar. 2022. doi:10.1016/j.jbi.2022.104030
- [163] Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, Board on the Health of Select Populations, and Institute of Medicine, *Cognitive Tests and Performance Validity Tests*. National Academies Press (US), Jun. 2015.
- [164] X. Li, J. Dai, S. Zhao, W. Liu, and H. Li, “Comparison of the value of Mini-Cog and MMSE screening in the rapid identification of Chinese outpatients with mild cognitive impairment,” *Medicine*, vol. 97, no. 22, p. e10966, Jun. 2018. doi:10.1097/MD.00000000000010966
- [165] A. J. Mitchell, “A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment,” *Journal of Psychiatric Research*, vol. 43, no. 4, pp. 411–431, Jan. 2009. doi:10.1016/j.jpsychires.2008.04.014
- [166] Z. S. Nasreddine *et al.*, “The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment,” *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, Apr. 2005. doi:10.1111/j.1532-5415.2005.53221.x
- [167] S. Freitas, M. R. Simões, J. Marôco, L. Alves, and I. Santana, “Construct Validity of the Montreal Cognitive Assessment (MoCA),” *Journal of the International Neuropsychological Society: JINS*, vol. 18, no. 2, pp. 242–250, Mar. 2012. doi:10.1017/S1355617711001573
- [168] A. E. Thomann, M. Berres, N. Goettel, L. A. Steiner, and A. U. Monsch, “Enhanced diagnostic accuracy for neurocognitive disorders: a revised cut-off approach for the Montreal Cognitive Assessment,” *Alzheimer’s Research & Therapy*, vol. 12, no. 39, Dec. 2020. doi:10.1186/s13195-020-00603-8

- [169] J. W. Shega *et al.*, “Measuring cognition: the Chicago Cognitive Function Measure in the National Social Life, Health and Aging Project, Wave 2,” *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, vol. 69 Suppl 2, pp. S166–176, Nov. 2014. doi:10.1093/geronb/gbu106
- [170] L. C. Beishon *et al.*, “Addenbrooke’s Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment,” *Cochrane Database of Systematic Reviews*, Dec. 2019. doi:10.1002/14651858.CD013282.pub2
- [171] D. Bruno and S. Schurmann Vignaga, “Addenbrooke’s cognitive examination III in the diagnosis of dementia: a critical review,” *Neuropsychiatric Disease and Treatment*, vol. Volume 15, pp. 441–447, Feb. 2019. doi:10.2147/NDT.S151253
- [172] C. Potts *et al.*, “Reliability of Addenbrooke’s Cognitive Examination III in differentiating between dementia, mild cognitive impairment and older adults who have not reported cognitive problems,” *European Journal of Ageing*, vol. 19, no. 3, pp. 495–507, Sep. 2022. doi:10.1007/s10433-021-00652-4
- [173] S. Freitas, M. R. Simões, L. Alves, and I. Santana, “Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease,” *Alzheimer Disease and Associated Disorders*, vol. 27, no. 1, pp. 37–43, Mar. 2013. doi:10.1097/WAD.0b013e3182420bfe
- [174] J. R. Hodges and A. J. Larner, “Addenbrooke’s Cognitive Examinations: ACE, ACE-R, ACE-III, ACEapp, and M-ACE,” in *Cognitive Screening Instruments*, A. J. Larner, Ed. Cham: Springer International Publishing, 2017, pp. 109–137. ISBN 978-3-319-44774-2 978-3-319-44775-9
- [175] E. E. De Roeck, P. P. De Deyn, E. Dierckx, and S. Engelborghs, “Brief cognitive screening instruments for early detection of Alzheimer’s disease: a systematic review,” *Alzheimer’s Research & Therapy*, vol. 11, no. 1, p. 21, Dec. 2019. doi:10.1186/s13195-019-0474-3
- [176] D. Jones, R. Wilkinson, C. Jackson, and P. Drew, “Variation and Interactional Non-Standardization in Neuropsychological Tests: The Case of the Addenbrooke’s Cognitive Examination,” *Qualitative Health Research*, vol. 30, no. 3, pp. 458–470, Feb. 2020. doi:10.1177/1049732319873052
- [177] T. Ijmker and C. J. C. Lamoth, “Gait and cognition: the relationship between gait stability and variability with executive function in persons with and without dementia,” *Gait & Posture*, vol. 35, no. 1, pp. 126–130, Jan. 2012. doi:10.1016/j.gaitpost.2011.08.022
- [178] R. Holtzer, J. Verghese, X. Xue, and R. B. Lipton, “Cognitive processes related to gait velocity: results from the Einstein Aging Study,” *Neuropsychology*, vol. 20, no. 2, pp. 215–223, Mar. 2006. doi:10.1037/0894-4105.20.2.215
- [179] V. J. A. Verlinden, J. N. van der Geest, A. Hofman, and M. A. Ikram, “Cognition and gait show a distinct pattern of association in the general population,” *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, vol. 10, no. 3, pp. 328–335, May 2014. doi:10.1016/j.jalz.2013.03.009
- [180] J. M. Hausdorff, G. Yogev, S. Springer, E. S. Simon, and N. Giladi, “Walking is more like catching than tapping: gait in the elderly as a complex cognitive task,” *Experimental Brain Research*, vol. 164, no. 4, pp. 541–548, Aug. 2005. doi:10.1007/s00221-005-2280-3

- [181] M. Amboni, P. Barone, and J. M. Hausdorff, "Cognitive Contributions to Gait and Falls: Evidence and Implications," *Movement disorders: official journal of the Movement Disorder Society*, vol. 28, no. 11, pp. 1520–1533, Sep. 2013. doi:10.1002/mds.25674
- [182] A. Mirelman *et al.*, "Executive function and falls in older adults: new findings from a five-year prospective study link fall risk to cognition," *PloS One*, vol. 7, no. 6, p. e40297, 2012. doi:10.1371/journal.pone.0040297
- [183] N. L. Watson *et al.*, "Executive function, memory, and gait speed decline in well-functioning older adults," *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 65, no. 10, pp. 1093–1100, Oct. 2010. doi:10.1093/gerona/glq111
- [184] R. Holtzer, C. Wang, and J. Verghese, "The relationship between attention and gait in aging: facts and fallacies," *Motor Control*, vol. 16, no. 1, pp. 64–80, Jan. 2012. doi:10.1123/mcj.16.1.64
- [185] K. L. Martin *et al.*, "Cognitive function, gait, and gait variability in older people: a population-based study," *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 68, no. 6, pp. 726–732, Jun. 2013. doi:10.1093/gerona/gls224
- [186] K. Duff, J. W. Mold, and M. M. Roberts, "Walking speed and global cognition: results from the OKLAHOMA Study," *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, vol. 15, no. 1, pp. 31–39, Jan. 2008. doi:10.1080/13825580701531904
- [187] S. Lord *et al.*, "Independent domains of gait in older adults and associated motor and non-motor attributes: validation of a factor analysis approach," *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 68, no. 7, pp. 820–827, Jul. 2013. doi:10.1093/gerona/gls255
- [188] G. Yogev-Seligmann, J. M. Hausdorff, and N. Giladi, "The role of executive function and attention in gait," *Movement Disorders: Official Journal of the Movement Disorder Society*, vol. 23, no. 3, pp. 329–342, Feb. 2008. doi:10.1002/mds.21720
- [189] P. Sunderaraman *et al.*, "Differential Associations Between Distinct Components of Cognitive Function and Mobility: Implications for Understanding Aging, Turning and Dual-Task Walking," *Frontiers in Aging Neuroscience*, vol. 11, Jul. 2019. doi:10.3389/fnagi.2019.00166
- [190] E. Sasson, G. M. Doniger, O. Pasternak, and Y. Assaf, "Structural correlates of memory performance with diffusion tensor imaging," *NeuroImage*, vol. 50, no. 3, pp. 1231–1242, Apr. 2010. doi:10.1016/j.neuroimage.2009.12.079
- [191] R. Holtzer, N. Epstein, J. R. Mahoney, M. Izzetoglu, and H. M. Blumen, "Neuroimaging of Mobility in Aging: A Targeted Review," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 69, no. 11, pp. 1375–1388, Nov. 2014. doi:10.1093/gerona/glu052
- [192] Z. Meiner, E. Ayers, and J. Verghese, "Motoric Cognitive Risk Syndrome: A Risk Factor for Cognitive Impairment and Dementia in Different Populations," *Annals of Geriatric Medicine and Research*, vol. 24, no. 1, pp. 3–14, Mar. 2020. doi:10.4235/agmr.20.0001

- [193] M. B. van Iersel, W. Hoefsloot, M. Munneke, B. R. Bloem, and M. G. M. Olde Rikkert, "Systematic review of quantitative clinical gait analysis in patients with dementia," *Zeitschrift Fur Gerontologie Und Geriatrie*, vol. 37, no. 1, pp. 27–32, Feb. 2004. doi:10.1007/s00391-004-0176-7
- [194] A.-M. De Cock *et al.*, "Comprehensive Quantitative Spatiotemporal Gait Analysis Identifies Gait Characteristics for Early Dementia Subtyping in Community Dwelling Older Adults," *Frontiers in Neurology*, vol. 10, p. 313, Apr. 2019. doi:10.3389/fneur.2019.00313
- [195] J. Dumurgier *et al.*, "Gait Speed and Decline in Gait Speed as Predictors of Incident Dementia," *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 72, no. 5, pp. 655–661, May 2017. doi:10.1093/gerona/glw110
- [196] T. Skillbäck *et al.*, "Slowing gait speed precedes cognitive decline by several years," *Alzheimer's & Dementia*, vol. 18, no. 9, pp. 1667–1676, Sep. 2022. doi:10.1002/alz.12537
- [197] L. H. J. Kikkert, N. Vuillerme, J. P. van Campen, T. Hortobágyi, and C. J. Lamoth, "Walking ability to predict future cognitive decline in old adults: A scoping review," *Ageing Research Reviews*, vol. 27, pp. 1–14, May 2016. doi:10.1016/j.arr.2016.02.001
- [198] O. Beauchet *et al.*, "Poor Gait Performance and Prediction of Dementia: Results From a Meta-Analysis," *Journal of the American Medical Directors Association*, vol. 17, no. 6, pp. 482–490, 2016. doi:10.1016/j.jamda.2015.12.092
- [199] D. J. Clark, "Automaticity of walking: functional significance, mechanisms, measurement and rehabilitation strategies," *Frontiers in Human Neuroscience*, vol. 9, May 2015. doi:10.3389/fnhum.2015.00246
- [200] P. Plummer and G. Eskes, "Measuring treatment effects on dual-task performance: a framework for research and clinical practice," *Frontiers in Human Neuroscience*, vol. 9, Apr. 2015. doi:10.3389/fnhum.2015.00225
- [201] C. Leone *et al.*, "Cognitive-motor dual-task interference: A systematic review of neural correlates," *Neuroscience & Biobehavioral Reviews*, vol. 75, pp. 348–360, Apr. 2017. doi:10.1016/j.neubiorev.2017.01.010
- [202] A. Nascimbeni *et al.*, "Dual task-related gait changes in patients with mild cognitive impairment," *Functional Neurology*, vol. 30, no. 1, pp. 59–65, Mar. 2015.
- [203] S. W. Muir *et al.*, "Gait assessment in mild cognitive impairment and Alzheimer's disease: the effect of dual-task challenges across the cognitive spectrum," *Gait & Posture*, vol. 35, no. 1, pp. 96–100, Jan. 2012. doi:10.1016/j.gaitpost.2011.08.014
- [204] M. Montero-Odasso, S. W. Muir, and M. Speechley, "Dual-task complexity affects gait in people with mild cognitive impairment: the interplay between gait variability, dual tasking, and risk of falls," *Archives of Physical Medicine and Rehabilitation*, vol. 93, no. 2, pp. 293–299, Feb. 2012. doi:10.1016/j.apmr.2011.08.026
- [205] S. W. Hunter, A. Divine, C. Frengopoulos, and M. Montero Odasso, "A framework for secondary cognitive and motor tasks in dual-task gait testing in people with mild cognitive impairment," *BMC Geriatrics*, vol. 18, no. 1, p. 202, Sep. 2018. doi:10.1186/s12877-018-0894-0

- [206] D. Maquet *et al.*, “Gait analysis in elderly adult patients with mild cognitive impairment and patients with mild Alzheimer’s disease: simple versus dual task: a preliminary report,” *Clinical Physiology and Functional Imaging*, vol. 30, no. 1, pp. 51–56, Jan. 2010. doi:10.1111/j.1475-097X.2009.00903.x
- [207] M. M. Montero-Odasso *et al.*, “Association of Dual-Task Gait With Incident Dementia in Mild Cognitive Impairment: Results From the Gait and Brain Study,” *JAMA neurology*, vol. 74, no. 7, pp. 857–865, 2017. doi:10.1001/jamaneurol.2017.0643
- [208] J. H. Ansai *et al.*, “Gait, dual task and history of falls in elderly with preserved cognition, mild cognitive impairment, and mild Alzheimer’s disease,” *Brazilian Journal of Physical Therapy*, vol. 21, no. 2, pp. 144–151, Mar. 2017. doi:10.1016/j.bjpt.2017.03.010
- [209] A. F. Pettersson, E. Olsson, and L.-O. Wahlund, “Effect of divided attention on gait in subjects with and without cognitive impairment,” *Journal of Geriatric Psychiatry and Neurology*, vol. 20, no. 1, pp. 58–62, Mar. 2007. doi:10.1177/0891988706293528
- [210] O. Beauchet, V. Dubost, K. Aminian, R. Gonthier, and R. W. Kressig, “Dual-task-related gait changes in the elderly: does the type of cognitive task matter?” *Journal of Motor Behavior*, vol. 37, no. 4, pp. 259–264, Jul. 2005.
- [211] H. Ehsani *et al.*, “The association between cognition and dual-tasking among older adults: the effect of motor function type and cognition task difficulty,” *Clinical Interventions in Aging*, vol. 14, pp. 659–669, 2019. doi:10.2147/CIA.S198697
- [212] K. Aoki *et al.*, “Early Detection of Lower MMSE Scores in Elderly Based on Dual-Task Gait,” *IEEE Access*, vol. 7, pp. 40 085–40 094, 2019. doi:10.1109/ACCESS.2019.2906908
- [213] S. Wu *et al.*, “Detecting Lower MMSE Scores in Older Adults Using Cross-Trial Features From a Dual-Task With Gait and Arithmetic,” *IEEE Access*, vol. 9, pp. 150 268–150 282, 2021. doi:10.1109/ACCESS.2021.3126067
- [214] —, “Early Detection of Low Cognitive Scores from Dual-task Performance Data Using a Spatio-temporal Graph Convolutional Neural Network,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Mexico: IEEE, Nov. 2021. doi:10.1109/EMBC46164.2021.9630304. ISBN 978-1-72811-179-7 pp. 1895–1901.
- [215] G. Mancioppi *et al.*, “Innovative motor and cognitive dual-task approaches combining upper and lower limbs may improve dementia early detection,” *Scientific Reports*, vol. 11, no. 1, p. 7449, Apr. 2021. doi:10.1038/s41598-021-86579-3
- [216] Y. Cedervall, K. Halvorsen, and A. C. Aberg, “A longitudinal study of gait function and characteristics of gait disturbance in individuals with Alzheimer’s disease,” *Gait & Posture*, vol. 39, no. 4, pp. 1022–1027, Apr. 2014. doi:10.1016/j.gaitpost.2013.12.026
- [217] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [218] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Machine learning: a review of classification and combining techniques,” *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, Nov. 2006. doi:10.1007/s10462-007-9052-3

- [219] Z. Wu, J. Zhang, and S. Hu, “Review on Classification Algorithm and Evaluation System of Machine Learning,” in *2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA)*. Xi’an, China: IEEE, Oct. 2020. doi:10.1109/ICICTA51737.2020.00052. ISBN 978-1-72818-666-5 pp. 214–218.
- [220] F. Ahamed, S. Shahrestani, and H. Cheung, “Internet of Things and Machine Learning for Healthy Ageing: Identifying the Early Signs of Dementia,” *Sensors*, vol. 20, no. 21, p. 6031, Oct. 2020. doi:10.3390/s20216031
- [221] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” in *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. NLD: IOS Press, Jun. 2007. ISBN 978-1-58603-780-2 pp. 3–24.
- [222] P. Xu, “Review on Studies of Machine Learning Algorithms,” *Journal of Physics: Conference Series*, vol. 1187, no. 5, p. 052103, Apr. 2019. doi:10.1088/1742-6596/1187/5/052103
- [223] M. Gwak, M. Sarrafzadeh, and E. Woo, “Support for a Clinical Diagnosis of Mild Cognitive Impairment Using Photoplethysmography and Gait Sensors,” in *Proceedings of the 2018 APSIPA Annual Summit and Conference*, Hawaii, USA, Nov. 2018.
- [224] J.-A. Martín-Gonzalo *et al.*, “Permutation Entropy and Irreversibility in Gait Kinematic Time Series from Patients with Mild Cognitive Decline and Early Alzheimer’s Dementia,” *Entropy*, vol. 21, no. 9, p. 868, Sep. 2019. doi:10.3390/e21090868
- [225] L. N. Boettcher *et al.*, “Dual-Task Gait Assessment and Machine Learning for Early-detection of Cognitive Decline,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Montreal, QC, Canada: IEEE, Jul. 2020. doi:10.1109/EMBC44109.2020.9175955. ISBN 978-1-72811-990-8 pp. 3204–3207.
- [226] C. Ricciardi *et al.*, “Machine learning can detect the presence of Mild cognitive impairment in patients affected by Parkinson’s Disease,” in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Bari, Italy: IEEE, Jun. 2020. doi:10.1109/MeMeA49120.2020.9137301. ISBN 978-1-72815-386-5 pp. 1–6.
- [227] Z. You *et al.*, “Alzheimer’s Disease Classification With a Cascade Neural Network,” *Frontiers in Public Health*, vol. 8, p. 584387, Nov. 2020. doi:10.3389/fpubh.2020.584387
- [228] —, “Alzheimer’s Disease Distinction Based On Gait Feature Analysis,” in *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*. Shenzhen, China: IEEE, Mar. 2021. doi:10.1109/HEALTHCOM49281.2021.9398984. ISBN 978-1-72816-267-6 pp. 1–6.
- [229] B. Ghoraani *et al.*, “Detection of mild cognitive impairment and Alzheimer’s disease using dual-task gait assessments and machine learning,” *Biomedical Signal Processing and Control*, vol. 64, p. 102249, Feb. 2021. doi:10.1016/j.bspc.2020.102249
- [230] D. Jung, J. Kim, M. Kim, C. W. Won, and K.-R. Mun, “Classifying the Risk of Cognitive Impairment Using Sequential Gait Characteristics and Long Short-Term Memory Networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 4029–4040, Oct. 2021. doi:10.1109/JBHI.2021.3073372

- [231] Y. Yamada *et al.*, “Combining Multimodal Behavioral Data of Gait, Speech, and Drawing for Classification of Alzheimer’s Disease and Mild Cognitive Impairment,” *Journal of Alzheimer’s Disease*, vol. 84, no. 1, pp. 315–327, Oct. 2021. doi:10.3233/JAD-210684
- [232] A. Shahzad, A. Dadlani, H. Lee, and K. Kim, “Automated Prescreening of Mild Cognitive Impairment Using Shank-Mounted Inertial Sensors Based Gait Biomarkers,” *IEEE Access*, vol. 10, pp. 15 835–15 844, 2022. doi:10.1109/ACCESS.2022.3149100
- [233] M. Seifollahi *et al.*, “Detection of mild cognitive Impairment from gait using Adaptive Neuro-Fuzzy Inference system,” *Biomedical Signal Processing and Control*, vol. 71, p. 103195, Jan. 2022. doi:10.1016/j.bspc.2021.103195
- [234] N. Toosizadeh *et al.*, “Screening older adults for amnesic mild cognitive impairment and early-stage Alzheimer’s disease using upper-extremity dual-tasking,” *Scientific Reports*, vol. 9, no. 1, pp. 1–11, Jul. 2019. doi:10.1038/s41598-019-46925-y
- [235] H. Ehsani *et al.*, “Can motor function uncertainty and local instability within upper-extremity dual-tasking predict amnesic mild cognitive impairment and early-stage Alzheimer’s disease?” *Computers in Biology and Medicine*, vol. 120, p. 103705, May 2020. doi:10.1016/j.combiomed.2020.103705
- [236] E. I. Konstantinidis *et al.*, “Physical Training In-Game Metrics for Cognitive Assessment: Evidence from Extended Trials with the Fitforall Exergaming Platform,” *Sensors*, vol. 21, no. 17, p. 5756, Aug. 2021. doi:10.3390/s21175756
- [237] R. Lees *et al.*, “Test Accuracy of Cognitive Screening Tests for Diagnosis of Dementia and Multidomain Cognitive Impairment in Stroke,” *Stroke*, vol. 45, no. 10, pp. 3008–3018, Oct. 2014. doi:10.1161/STROKEAHA.114.005842
- [238] K. Hajian-Tilaki, “Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation,” *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627–635, 2013.
- [239] W. Feldman, “How serious are the adverse effects of screening?” *Journal of General Internal Medicine*, vol. 5, no. S2, pp. S50–S53, Sep. 1990. doi:10.1007/BF02600842
- [240] C. Tunca *et al.*, “Inertial Sensor-Based Robust Gait Analysis in Non-Hospital Settings for Neurological Disorders,” *Sensors*, vol. 17, no. 4, p. 825, Apr. 2017. doi:10.3390/s17040825
- [241] J. Hannink *et al.*, “Benchmarking Foot Trajectory Estimation Methods for Mobile Gait Analysis,” *Sensors*, vol. 17, no. 9, p. 1940, Aug. 2017. doi:10.3390/s17091940
- [242] A. M. Sabatini, C. Martelloni, S. Scapellato, and F. Cavallo, “Assessment of walking features from foot inertial sensing,” *IEEE transactions on bio-medical engineering*, vol. 52, no. 3, pp. 486–494, Mar. 2005. doi:10.1109/TBME.2004.840727
- [243] K. Hori *et al.*, “Inertial Measurement Unit-Based Estimation of Foot Trajectory for Clinical Gait Analysis,” *Frontiers in Physiology*, vol. 10, p. 1530, 2019. doi:10.3389/fphys.2019.01530
- [244] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, “Estimation of IMU and MARG orientation using a gradient descent algorithm,” in *2011 IEEE International Conference on Rehabilitation Robotics*. Zurich: IEEE, Jun. 2011.

- doi:10.1109/ICORR.2011.5975346. ISBN 978-1-4244-9862-8 978-1-4244-9863-5 978-1-4244-9861-1 pp. 1–7.
- [245] M. Euston, P. Coote, R. Mahony, Jonghyuk Kim, and T. Hamel, “A complementary filter for attitude estimation of a fixed-wing UAV,” in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Nice: IEEE, Sep. 2008. doi:10.1109/IROS.2008.4650766. ISBN 978-1-4244-2057-5 978-1-4244-2058-2 pp. 340–345.
- [246] M. Basso, A. Martinelli, S. Morosi, and F. Sera, “A Real-Time GNSS/PDR Navigation System for Mobile Devices,” *Remote Sensing*, vol. 13, no. 8, p. 1567, Jan. 2021. doi:10.3390/rs13081567
- [247] R. Leonardo *et al.*, “Determination of the Walking Direction of a Pedestrian from Acceleration Data,” in *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sep. 2019. doi:10.1109/IPIN.2019.8911801 pp. 1–6.
- [248] M. Falbriard, F. Meyer, B. Mariani, G. P. Millet, and K. Aminian, “Accurate Estimation of Running Temporal Parameters Using Foot-Worn Inertial Sensors,” *Frontiers in Physiology*, vol. 9, 2018. doi:10.3389/fphys.2018.00610
- [249] C. M. Kanzler *et al.*, “Inertial sensor based and shoe size independent gait analysis including heel and toe clearance estimation,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan: IEEE, Aug. 2015. doi:10.1109/EMBC.2015.7319618. ISBN 978-1-4244-9271-8 pp. 5424–5427.
- [250] S. Byun *et al.*, “Walking-speed estimation using a single inertial measurement unit for the older adults,” *PLOS ONE*, vol. 14, no. 12, p. e0227075, Dec. 2019. doi:10.1371/journal.pone.0227075
- [251] S. Lambrecht *et al.*, “Real-Time Gait Event Detection Based on Kinematic Data Coupled to a Biomechanical Model,” *Sensors*, vol. 17, no. 4, p. 671, Mar. 2017. doi:10.3390/s17040671
- [252] A. Hreljac and R. N. Marshall, “Algorithms to determine event timing during normal walking using kinematic data,” *Journal of Biomechanics*, vol. 33, no. 6, pp. 783–786, Jun. 2000. doi:10.1016/S0021-9290(00)00014-2
- [253] V. Guimarães, I. Sousa, and M. V. Correia, “Gait events detection from heel and toe trajectories: comparison of methods using multiple datasets,” in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Lausanne, Switzerland: IEEE, Jun. 2021. doi:10.1109/MeMeA52024.2021.9478606. ISBN 978-1-66541-914-7 pp. 1–6.
- [254] F. Huxham, J. Gong, R. Baker, M. Morris, and R. Ianssek, “Defining spatial parameters for non-linear walking,” *Gait & Posture*, vol. 23, no. 2, pp. 159–163, Feb. 2006. doi:10.1016/j.gaitpost.2005.01.001
- [255] I. Skog, J.-O. Nilsson, and P. Handel, “Evaluation of zero-velocity detectors for foot-mounted inertial navigation systems,” in *2010 International Conference on Indoor Positioning and Indoor Navigation*. Zurich, Switzerland: IEEE, Sep. 2010. doi:10.1109/IPIN.2010.5646936. ISBN 978-1-4244-5862-2 pp. 1–6.
- [256] M. D. Shuster and S. D. Oh, “Three-axis attitude determination from vector observations,” *Journal of Guidance and Control*, vol. 4, no. 1, pp. 70–77, Jan. 1981. doi:10.2514/3.19717

- [257] S. O. H. Madgwick, “An efficient orientation filter for inertial and inertial/magnetic sensor arrays,” Report x-io and University of Bristol (UK), Technical Report, 2010.
- [258] D. G. Altman and J. M. Bland, “Measurement in Medicine: The Analysis of Method Comparison Studies,” *The Statistician*, vol. 32, no. 3, p. 307, Sep. 1983. doi:10.2307/2987937
- [259] K. Shoemake, “Uniform Random Rotations,” in *Graphics Gems III (IBM Version)*. Elsevier, 1992, pp. 124–132. ISBN 978-0-12-409673-8
- [260] M. Mukaka, “A guide to appropriate use of Correlation coefficient in medical research,” *Malawi Medical Journal : The Journal of Medical Association of Malawi*, vol. 24, no. 3, pp. 69–71, Sep. 2012.
- [261] B. Mariani, H. Rouhani, X. Crevoisier, and K. Aminian, “Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors,” *Gait & Posture*, vol. 37, no. 2, pp. 229–234, Feb. 2013. doi:10.1016/j.gaitpost.2012.07.012
- [262] S. Han, Z. Meng, O. Omisore, T. Akinyemi, and Y. Yan, “Random Error Reduction Algorithms for MEMS Inertial Sensor Accuracy Improvement—A Review,” *Micromachines*, vol. 11, no. 11, p. 1021, Nov. 2020. doi:10.3390/mi11111021
- [263] J. Hannink *et al.*, “Sensor-Based Gait Parameter Extraction With Deep Convolutional Neural Networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 85–93, Jan. 2017. doi:10.1109/JBHI.2016.2636456
- [264] J. P. Silva do Monte Lima, H. Uchiyama, and R.-i. Taniguchi, “End-to-End Learning Framework for IMU-Based 6-DOF Odometry,” *Sensors*, vol. 19, no. 17, p. 3777, Aug. 2019. doi:10.3390/s19173777
- [265] C. Chen, C. X. Lu, J. Wahlstrom, A. Markham, and N. Trigoni, “Deep Neural Network Based Inertial Odometry Using Low-Cost Inertial Measurement Units,” *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1351–1364, Apr. 2021. doi:10.1109/TMC.2019.2960780
- [266] Q. Wang *et al.*, “Pedestrian Stride-Length Estimation Based on LSTM and Denoising Autoencoders,” *Sensors*, vol. 19, no. 4, p. 840, Feb. 2019. doi:10.3390/s19040840
- [267] O. Asraf, F. Shama, and I. Klein, “PDRNet: A Deep-Learning Pedestrian Dead Reckoning Framework,” *IEEE Sensors Journal*, pp. 1–1, 2021. doi:10.1109/JSEN.2021.3066840
- [268] V. Guimarães, I. Sousa, and M. V. Correia, “Orientation-Invariant Spatio-Temporal Gait Analysis Using Foot-Worn Inertial Sensors,” *Sensors*, vol. 21, no. 11, p. 3940, Jun. 2021. doi:10.3390/s21113940
- [269] C. Tunca, G. Salur, and C. Ersoy, “Deep Learning for Fall Risk Assessment With Inertial Sensors: Utilizing Domain Knowledge in Spatio-Temporal Gait Parameters,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1994–2005, Jul. 2020. doi:10.1109/JBHI.2019.2958879
- [270] M. Mundt *et al.*, “Prediction of lower limb joint angles and moments during gait using artificial neural networks,” *Medical & Biological Engineering & Computing*, vol. 58, no. 1, pp. 211–225, Jan. 2020. doi:10.1007/s11517-019-02061-3

- [271] D. Weber, C. Gühmann, and T. Seel, “Neural Networks Versus Conventional Filters for Inertial-Sensor-based Attitude Estimation,” *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pp. 1–8, Jul. 2020. doi:10.23919/FUSION45008.2020.9190634
- [272] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019. doi:10.1162/neco_a_01199
- [273] J. Conte Alcaraz, S. Moghaddamnia, and J. Peissig, “Efficiency of deep neural networks for joint angle modeling in digital gait assessment,” *EURASIP Journal on Advances in Signal Processing*, vol. 2021, no. 1, p. 10, Dec. 2021. doi:10.1186/s13634-020-00715-1
- [274] M. Mundt *et al.*, “A Comparison of Three Neural Network Approaches for Estimating Joint Angles and Moments from Inertial Measurement Units,” *Sensors*, vol. 21, no. 13, p. 4535, Jul. 2021. doi:10.3390/s21134535
- [275] M. Mundt, A. Koeppe, F. Bamer, S. David, and B. Markert, “Artificial Neural Networks in Motion Analysis—Applications of Unsupervised and Heuristic Feature Selection Techniques,” *Sensors*, vol. 20, no. 16, p. 4581, Aug. 2020. doi:10.3390/s20164581
- [276] M. A. Esfahani, H. Wang, K. Wu, and S. Yuan, “OriNet: Robust 3-D Orientation Estimation With a Single Particular IMU,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 399–406, Apr. 2020. doi:10.1109/LRA.2019.2959507
- [277] K. Xia, J. Huang, and H. Wang, “LSTM-CNN Architecture for Human Activity Recognition,” *IEEE Access*, vol. 8, pp. 56 855–56 866, 2020. doi:10.1109/ACCESS.2020.2982225
- [278] T. Zebin, M. Sperrin, N. Peek, and A. J. Casson, “Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Honolulu, HI: IEEE, Jul. 2018. doi:10.1109/EMBC.2018.8513115. ISBN 978-1-5386-3646-6 pp. 1–4.
- [279] L. Alawneh, B. Mohsen, M. Al-Zinati, A. Shatnawi, and M. Al-Ayyoub, “A Comparison of Unidirectional and Bidirectional LSTM Networks for Human Activity Recognition,” in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. Austin, TX, USA: IEEE, Mar. 2020. doi:10.1109/PerComWorkshops48775.2020.9156264. ISBN 978-1-72814-716-1 pp. 1–6.
- [280] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017.
- [281] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, “Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization,” *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6765–6816, 2017.
- [282] T. O’Malley *et al.*, “Keras tuner,” <https://github.com/keras-team/keras-tuner>, 2019.
- [283] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

- [284] C. Ornelas-Vences, L. P. Sanchez-Fernandez, L. A. Sanchez-Perez, A. Garza-Rodriguez, and A. Villegas-Bastida, “Fuzzy inference model evaluating turn for Parkinson’s disease patients,” *Computers in Biology and Medicine*, vol. 89, pp. 379–388, Oct. 2017. doi:10.1016/j.combiomed.2017.08.026
- [285] R. Delgado-Escano, F. M. Castro, J. R. Cozar, M. J. Marin-Jimenez, and N. Guil, “An End-to-End Multi-Task and Fusion CNN for Inertial-Based Gait Recognition,” *IEEE Access*, vol. 7, pp. 1897–1908, 2019. doi:10.1109/ACCESS.2018.2886899
- [286] B. K. Santhiranayagam, D. T. Lai, W. Sparrow, and R. K. Begg, “A machine learning approach to estimate Minimum Toe Clearance using Inertial Measurement Units,” *Journal of Biomechanics*, vol. 48, no. 16, pp. 4309–4316, Dec. 2015. doi:10.1016/j.jbiomech.2015.10.040
- [287] K. E. Peyer, C. A. Brassey, K. A. Rose, and W. I. Sellers, “Locomotion pattern and foot pressure adjustments during gentle turns in healthy subjects,” *Journal of Biomechanics*, vol. 60, pp. 65–71, Jul. 2017. doi:10.1016/j.jbiomech.2017.06.025
- [288] C. Bonnyaud, D. Pradon, D. Bensmail, and N. Roche, “Dynamic Stability and Risk of Tripping during the Timed Up and Go Test in Hemiparetic and Healthy Subjects,” *PLOS ONE*, vol. 10, no. 10, p. e0140317, Oct. 2015. doi:10.1371/journal.pone.0140317
- [289] X. Bai *et al.*, “Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments,” *Pattern Recognition*, vol. 120, p. 108102, Dec. 2021. doi:10.1016/j.patcog.2021.108102
- [290] L. Tran and D. Choi, “Data Augmentation for Inertial Sensor-Based Gait Deep Neural Network,” *IEEE Access*, vol. 8, pp. 12 364–12 378, 2020. doi:10.1109/ACCESS.2020.2966142
- [291] J. Camargo, A. Ramanathan, N. Csomay-Shanklin, and A. Young, “Automated gap-filling for marker-based biomechanical motion capture data,” *Computer Methods in Biomechanics and Biomedical Engineering*, pp. 1–10, Jul. 2020. doi:10.1080/10255842.2020.1789971
- [292] J. Meyer, M. Kuderer, J. Muller, and W. Burgard, “Online marker labeling for fully automatic skeleton tracking in optical motion capture,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong, China: IEEE, May 2014. doi:10.1109/ICRA.2014.6907690. ISBN 978-1-4799-3685-4 pp. 5652–5657.
- [293] E. Giannouli, T. Morat, and W. Zijlstra, “A Novel Square-Stepping Exercise Program for Older Adults (StepIt): Rationale and Implications for Falls Prevention,” *Frontiers in Medicine*, vol. 6, p. 318, Jan. 2020. doi:10.3389/fmed.2019.00318
- [294] Y. Okubo, D. Schoene, and S. R. Lord, “Step training improves reaction time, gait and balance and reduces falls in older people: a systematic review and meta-analysis,” *British Journal of Sports Medicine*, vol. 51, no. 7, pp. 586–593, Apr. 2017. doi:10.1136/bjsports-2015-095452
- [295] K. Hauer *et al.*, “Effectiveness and sustainability of a motor-cognitive stepping exergame training on stepping performance in older adults: a randomized controlled trial,” *European Review of Aging and Physical Activity*, vol. 17, no. 17, Dec. 2020. doi:10.1186/s11556-020-00248-4

- [296] C. Di Lorito *et al.*, “A systematic literature review and meta-analysis on digital health interventions for people living with dementia and Mild Cognitive Impairment,” *International Journal of Geriatric Psychiatry*, vol. 37, no. 6, Jun. 2022. doi:10.1002/gps.5730
- [297] J. Swanenburg, K. Wild, D. Straumann, and E. D. de Bruin, “Exergaming in a Moving Virtual World to Train Vestibular Functions and Gait; a Proof-of-Concept-Study With Older Adults,” *Frontiers in Physiology*, vol. 9, Jul. 2018. doi:10.3389/fphys.2018.00988
- [298] D. Schoene, T. Valenzuela, S. R. Lord, and E. D. de Bruin, “The effect of interactive cognitive-motor training in reducing fall risk in older people: a systematic review,” *BMC geriatrics*, vol. 14, p. 107, Sep. 2014. doi:10.1186/1471-2318-14-107
- [299] M. van Diest, C. J. Lamothe, J. Stegenga, G. J. Verkerke, and K. Postema, “Exergaming for balance training of elderly: state of the art and future developments,” *Journal of Neuro-Engineering and Rehabilitation*, vol. 10, no. 101, 2013. doi:10.1186/1743-0003-10-101
- [300] Q. Zhou, H. Zhang, Z. Lari, Z. Liu, and N. El-Sheimy, “Design and Implementation of Foot-Mounted Inertial Sensor Based Wearable Electronic Device for Game Play Application,” *Sensors (Basel, Switzerland)*, vol. 16, no. 10, Oct. 2016. doi:10.3390/s16101752
- [301] X. Hou and J. Bergmann, “Pedestrian Dead Reckoning With Wearable Sensors: A Systematic Review,” *IEEE Sensors Journal*, vol. 21, no. 1, pp. 143–152, Jan. 2021. doi:10.1109/JSEN.2020.3014955
- [302] Y. Sakuma and M. Fujii, “A Study on Direction Estimation of Movement by Multiple Sensors for Pedestrian Dead-Reckoning,” in *2017 Fifth International Symposium on Computing and Networking (CANDAR)*. Aomori: IEEE, Nov. 2017. doi:10.1109/CANDAR.2017.64. ISBN 978-1-5386-2087-8 pp. 603–605.
- [303] T. Matsumoto and M. Fujii, “A study on estimation of traveling direction for pedestrian dead reckoning by using inertial sensors,” in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*. Kobe, Japan: IEEE, Oct. 2020. doi:10.1109/GCCE50665.2020.9291953. ISBN 978-1-72819-802-6 pp. 465–466.
- [304] B. Wagstaff, V. Peretroukhin, and J. Kelly, “Robust Data-Driven Zero-Velocity Detection for Foot-Mounted Inertial Navigation,” *IEEE Sensors Journal*, vol. 20, no. 2, pp. 957–967, Jan. 2020. doi:10.1109/JSEN.2019.2944412
- [305] Web Bluetooth Community Group. Web Bluetooth API. [Online]. Available: <https://webbluetoothcg.github.io/web-bluetooth/>
- [306] R. Mahony, T. Hamel, and J.-M. Pflimlin, “Nonlinear Complementary Filters on the Special Orthogonal Group,” *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203–1218, Jun. 2008. doi:10.1109/TAC.2008.923738
- [307] S. F. Fux, “Development of a planar low cost Inertial Measurement Unit for UAVs and MAVs,” Master’s thesis, Swiss Federal Institute of Technology Zurich, 2008.
- [308] K. Manal and W. Rose, “A general solution for the time delay introduced by a low-pass Butterworth digital filter: An application to musculoskeletal modeling,” *Journal of Biomechanics*, vol. 40, no. 3, pp. 678–681, Jan. 2007. doi:10.1016/j.jbiomech.2006.02.001

- [309] A. Mansour *et al.*, “Drift Control of Pedestrian Dead Reckoning (PDR) for Long Period Navigation under Different Smartphone Poses,” *Engineering Proceedings*, vol. 10, no. 1, p. 21, 2021. doi:10.3390/ecsa-8-11302
- [310] V. Niennattrakul, D. Srisai, and C. A. Ratanamahatana, “Shape-based template matching for time series data,” *Knowledge-Based Systems*, vol. 26, pp. 1–8, Feb. 2012. doi:10.1016/j.knosys.2011.04.015
- [311] L. Dommershuijsen *et al.*, “Gait speed reference values in community-dwelling older adults – Cross-sectional analysis from the Rotterdam Study,” *Experimental Gerontology*, vol. 158, p. 111646, Feb. 2022. doi:10.1016/j.exger.2021.111646
- [312] S. Lambrecht and A. J. del Ama, “Human Movement Analysis with Inertial Sensors,” in *Emerging Therapies in Neurorehabilitation*, J. L. Pons and D. Torricelli, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, vol. 4, pp. 305–328. ISBN 978-3-642-38555-1 978-3-642-38556-8
- [313] M. Caruso *et al.*, “Analysis of the Accuracy of Ten Algorithms for Orientation Estimation Using Inertial and Magnetic Sensing under Optimal Conditions: One Size Does Not Fit All,” *Sensors*, vol. 21, no. 7, p. 2543, Apr. 2021. doi:10.3390/s21072543
- [314] S. Qiu *et al.*, “Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges,” *Information Fusion*, vol. 80, pp. 241–265, Apr. 2022. doi:10.1016/j.inffus.2021.11.006
- [315] T. T. Pham and Y. S. Suh, “Conditional Generative Adversarial Network-Based Regression Approach for Walking Distance Estimation Using Waist-Mounted Inertial Sensors,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022. doi:10.1109/TIM.2022.3177730
- [316] S. Choy, B. Wong, G. Simon, and C. Rosenberg, “A hybrid edge-cloud architecture for reducing on-demand gaming latency,” *Multimedia Systems*, vol. 20, no. 5, pp. 503–519, Oct. 2014. doi:10.1007/s00530-014-0367-z
- [317] Y. Celik, S. Stuart, W. Woo, E. Sejdic, and A. Godfrey, “Multi-modal gait: A wearable, algorithm and data fusion approach for clinical and free-living assessment,” *Information Fusion*, vol. 78, pp. 57–70, Feb. 2022. doi:10.1016/j.inffus.2021.09.016
- [318] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò, “Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy,” *JMIR Serious Games*, vol. 4, no. 2, p. e11, Jul. 2016. doi:10.2196/games.5888
- [319] R. L. Mandryk and M. V. Birk, “The Potential of Game-Based Digital Biomarkers for Modeling Mental Health,” *JMIR Mental Health*, vol. 6, no. 4, p. e13485, 2019. doi:10.2196/13485
- [320] C. Boletsis and S. McCallum, “Smartkuber: A Serious Game for Cognitive Health Screening of Elderly Players,” *Games for Health Journal*, vol. 5, no. 4, pp. 241–251, Aug. 2016. doi:10.1089/g4h.2015.0107
- [321] Y.-T. Chen *et al.*, “Evaluation of the Reaction Time and Accuracy Rate in Normal Subjects, MCI, and Dementia Using Serious Games,” *Applied Sciences*, vol. 11, no. 2, p. 628, Jan. 2021. doi:10.3390/app11020628

- [322] S. Wiloth, N. Lemke, C. Werner, and K. Hauer, "Validation of a Computerized, Game-based Assessment Strategy to Measure Training Effects on Motor-Cognitive Functions in People With Dementia," *JMIR Serious Games*, vol. 4, no. 2, Jul. 2016. doi:10.2196/games.5696
- [323] E. Litz *et al.*, "Validation of a Motor-Cognitive Assessment for a Stepping Exergame in Older Adults: Use of Game-Specific, Internal Data Stream," *Games for Health Journal*, Oct. 2019. doi:10.1089/g4h.2019.0081
- [324] J. L. Medell and N. B. Alexander, "A Clinical Measure of Maximal and Rapid Stepping in Older Women," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 55, no. 8, pp. M429–M433, Aug. 2000. doi:10.1093/gerona/55.8.M429
- [325] V. Guimarães *et al.*, "An Exergame Solution for Personalized Multicomponent Training in Older Adults," *Applied Sciences*, vol. 11, no. 17, p. 7986, Aug. 2021. doi:10.3390/app11177986
- [326] S. Vasunilashorn *et al.*, "Use of the Short Physical Performance Battery Score to predict loss of ability to walk 400 meters: analysis from the InCHIANTI study," *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, vol. 64, no. 2, pp. 223–229, Feb. 2009. doi:10.1093/gerona/gln022
- [327] R. Bernabeu-Mora *et al.*, "The Short Physical Performance Battery is a discriminative tool for identifying patients with COPD at risk of disability," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 10, pp. 2619–2626, 2015. doi:10.2147/COPD.S94377
- [328] A. E. Thomann *et al.*, "The Montreal Cognitive Assessment: Normative Data from a German-Speaking Cohort and Comparison with International Normative Samples," *Journal of Alzheimer's Disease*, vol. 64, no. 2, pp. 643–655, Jun. 2018. doi:10.3233/JAD-180080
- [329] J. M. Guralnik *et al.*, "A Short Physical Performance Battery Assessing Lower Extremity Function: Association With Self-Reported Disability and Prediction of Mortality and Nursing Home Admission," *Journal of Gerontology*, vol. 49, no. 2, pp. M85–M94, Mar. 1994. doi:10.1093/geronj/49.2.M85
- [330] J. M. Guralnik, L. Ferrucci, E. M. Simonsick, M. E. Salive, and R. B. Wallace, "Lower-Extremity Function in Persons over the Age of 70 Years as a Predictor of Subsequent Disability," *New England Journal of Medicine*, vol. 332, no. 9, pp. 556–562, Mar. 1995. doi:10.1056/NEJM199503023320902
- [331] A. Middleton, S. L. Fritz, and M. Lusardi, "Walking Speed: The Functional Vital Sign," *Journal of Aging and Physical Activity*, vol. 23, no. 2, pp. 314–322, Apr. 2015. doi:10.1123/japa.2013-0236
- [332] M. G. Borda *et al.*, "Educational level and its Association with the domains of the Montreal Cognitive Assessment Test," *Aging & Mental Health*, vol. 23, no. 10, pp. 1300–1306, Oct. 2019. doi:10.1080/13607863.2018.1488940
- [333] B. R. Overholser and K. M. Sowinski, "Biostatistics primer: part 2," *Nutrition in Clinical Practice: Official Publication of the American Society for Parenteral and Enteral Nutrition*, vol. 23, no. 1, pp. 76–84, Feb. 2008. doi:10.1177/011542650802300176
- [334] R. O'Caoimh, S. Timmons, and D. W. Molloy, "Screening for Mild Cognitive Impairment: Comparison of "MCI Specific" Screening Instruments," *Journal of Alzheimer's Disease*, vol. 51, no. 2, pp. 619–629, Mar. 2016. doi:10.3233/JAD-150881

- [335] N. Skjæret-Maroni *et al.*, “Exergaming in Older Adults: Movement Characteristics While Playing Stepping Games,” *Frontiers in Psychology*, vol. 7, 2016. doi:10.3389/fpsyg.2016.00964
- [336] B.-I. Cho, D. Scarpace, and N. B. Alexander, “Tests of Stepping as Indicators of Mobility, Balance, and Fall Risk in Balance-Impaired Older Adults: Maximal Step Length in Impaired Older Adults,” *Journal of the American Geriatrics Society*, vol. 52, no. 7, pp. 1168–1173, Jul. 2004. doi:10.1111/j.1532-5415.2004.52317.x
- [337] C. W. Luchies *et al.*, “Effects of Age, Step Direction, and Reaction Condition on the Ability to Step Quickly,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 57, no. 4, pp. M246–M249, Apr. 2002. doi:10.1093/gerona/57.4.M246
- [338] S. T. Smith, C. Sherrington, S. Studenski, D. Schoene, and S. R. Lord, “A novel Dance Dance Revolution (DDR) system for in-home training of stepping ability: basic parameters of system use by older adults,” *British Journal of Sports Medicine*, vol. 45, no. 5, pp. 441–445, Apr. 2011. doi:10.1136/bjsm.2009.066845
- [339] D. J. Plebanek and V. M. Sloutsky, “Selective attention, filtering, and the development of working memory,” *Developmental Science*, vol. 22, no. 1, p. e12727, Jan. 2019. doi:10.1111/desc.12727
- [340] A. Diamond, “Executive Functions,” *Annual Review of Psychology*, vol. 64, no. 1, pp. 135–168, Jan. 2013. doi:10.1146/annurev-psych-113011-143750
- [341] A. Bäck, “The Concept of Abstraction,” *The Society for Ancient Greek Philosophy Newsletter*, no. 376, Mar. 2006.
- [342] P. Julayanont, N. Phillips, H. Chertkow, and Z. S. Nasreddine, “Montreal Cognitive Assessment (MoCA): Concept and Clinical Review,” in *Cognitive Screening Instruments*, A. J. Larner, Ed. London: Springer London, 2013, pp. 111–151. ISBN 978-1-4471-2451-1 978-1-4471-2452-8
- [343] K. B. Kortte, M. D. Horner, and W. K. Windham, “The Trail Making Test, Part B: Cognitive Flexibility or Ability to Maintain Set?” *Applied Neuropsychology*, vol. 9, no. 2, pp. 106–109, Jun. 2002. doi:10.1207/S15324826AN0902_5
- [344] I. Sánchez-Cubillo *et al.*, “Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities,” *Journal of the International Neuropsychological Society*, vol. 15, no. 3, pp. 438–450, May 2009. doi:10.1017/S1355617709090626
- [345] S. Mori *et al.*, “Possibility of Using Quantitative Assessment with the Cube Copying Test for Evaluation of Visuo-spatial Function in Patients with Alzheimer’s Disease,” *Progress in Rehabilitation Medicine*, vol. 6, no. 0, p. n/a, 2021. doi:10.2490/prm.20210021
- [346] A. E. Patla, “Understanding the roles of vision in the control of human locomotion,” *Gait & Posture*, vol. 5, no. 1, pp. 54–69, Feb. 1997. doi:10.1016/S0966-6362(96)01109-5
- [347] W. R. Young and M. A. Hollands, “Evidence for age-related decline in visuomotor function and reactive stepping adjustments,” *Gait & Posture*, vol. 36, no. 3, pp. 477–481, Jul. 2012. doi:10.1016/j.gaitpost.2012.04.009

- [348] J. J. d. Paula, D. M. d. Miranda, E. N. d. Moraes, and L. F. Malloy-Diniz, "Mapping the clockworks: what does the Clock Drawing Test assess in normal and pathological aging?" *Arquivos de Neuro-Psiquiatria*, vol. 71, no. 10, pp. 763–768, Oct. 2013. doi:10.1590/0004-282X20130118
- [349] M. Soffer *et al.*, "Time setting errors in the Clock Drawing Test are associated with both semantic and executive deficits," *Applied Neuropsychology: Adult*, pp. 1–10, Jan. 2022. doi:10.1080/23279095.2021.2023154
- [350] C. Rothmayr *et al.*, "Dissociation of neural correlates of verbal and non-verbal visual working memory with different delays," *Behavioral and Brain Functions*, vol. 3, no. 1, p. 56, Dec. 2007. doi:10.1186/1744-9081-3-56
- [351] C. C. Morey and N. Cowan, "When Do Visual and Verbal Memories Conflict? The Importance of Working-Memory Load and Retrieval." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 31, no. 4, pp. 703–713, 2005. doi:10.1037/0278-7393.31.4.703
- [352] R. F. Coen, D. A. Robertson, R. A. Kenny, and B. L. King-Kallimanis, "Strengths and Limitations of the MoCA for Assessing Cognitive Functioning: Findings From a Large Representative Sample of Irish Older Adults," *Journal of Geriatric Psychiatry and Neurology*, vol. 29, no. 1, pp. 18–24, Jan. 2016. doi:10.1177/0891988715598236
- [353] P. Julayanont, M. Brousseau, H. Chertkow, N. Phillips, and Z. S. Nasreddine, "Montreal Cognitive Assessment Memory Index Score (MoCA-MIS) as a Predictor of Conversion from Mild Cognitive Impairment to Alzheimer's Disease," *Journal of the American Geriatrics Society*, vol. 62, no. 4, pp. 679–684, Apr. 2014. doi:10.1111/jgs.12742
- [354] K. V. Papp, P. J. Snyder, P. Maruff, J. Bartkowiak, and R. H. Pietrzak, "Detecting Subtle Changes in Visuospatial Executive Function and Learning in the Amnestic Variant of Mild Cognitive Impairment," *PLoS ONE*, vol. 6, no. 7, p. e21688, Jul. 2011. doi:10.1371/journal.pone.0021688
- [355] B. Bonnechère *et al.*, "Age-Associated Capacity to Progress When Playing Cognitive Mobile Games: Ecological Retrospective Observational Study," *JMIR Serious Games*, vol. 8, no. 2, p. e17121, Jun. 2020. doi:10.2196/17121
- [356] L. H. Jakobsen, J. M. Sorensen, I. K. Rask, B. S. Jensen, and J. Kondrup, "Validation of reaction time as a measure of cognitive function and quality of life in healthy subjects and patients," *Nutrition*, vol. 27, no. 5, pp. 561–570, May 2011. doi:10.1016/j.nut.2010.08.003
- [357] D. Andriuta, M. Diouf, M. Roussel, and O. Godefroy, "Is Reaction Time Slowing an Early Sign of Alzheimer's Disease? A Meta-Analysis," *Dementia and Geriatric Cognitive Disorders*, vol. 47, no. 4-6, pp. 281–288, 2019. doi:10.1159/000500348
- [358] A. F. Healy, J. A. Kole, and L. E. Bourne, "Training principles to advance expertise," *Frontiers in Psychology*, vol. 5, p. 131, 2014. doi:10.3389/fpsyg.2014.00131
- [359] T. R. Hendershott, D. Zhu, S. Llanes, and K. L. Poston, "Domain-specific accuracy of the Montreal Cognitive Assessment subsections in Parkinson's disease," *Parkinsonism & Related Disorders*, vol. 38, pp. 31–34, May 2017. doi:10.1016/j.parkreldis.2017.02.008

- [360] E. von Elm *et al.*, “The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies,” *PLoS medicine*, vol. 4, no. 10, p. e296, Oct. 2007. doi:10.1371/journal.pmed.0040296
- [361] C. R. Bowie and P. D. Harvey, “Administration and interpretation of the Trail Making Test,” *Nature Protocols*, vol. 1, no. 5, pp. 2277–2281, Dec. 2006. doi:10.1038/nprot.2006.390
- [362] F. Scarpina and S. Tagini, “The Stroop Color and Word Test,” *Frontiers in Psychology*, vol. 8, Apr. 2017. doi:10.3389/fpsyg.2017.00557
- [363] R. W. Elwood, “The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application,” *Neuropsychology Review*, vol. 2, no. 2, pp. 179–201, Jun. 1991. doi:10.1007/BF01109053
- [364] S. L. Teruya, C. Dimino, K. D. Silverman, and T. Mielenz, “Poor Lower Extremity Functioning Is Associated with Modest Increased Incidence of Probable Dementia,” *Geriatrics*, vol. 6, no. 3, p. 77, Aug. 2021. doi:10.3390/geriatrics6030077
- [365] G. V. Ostir, S. Volpato, L. P. Fried, P. Chaves, and J. M. Guralnik, “Reliability and sensitivity to change assessed for a summary measure of lower body function,” *Journal of Clinical Epidemiology*, vol. 55, no. 9, pp. 916–921, Sep. 2002. doi:10.1016/S0895-4356(02)00436-5
- [366] R. Pavašini *et al.*, “Short Physical Performance Battery and all-cause mortality: systematic review and meta-analysis,” *BMC Medicine*, vol. 14, no. 1, p. 215, Dec. 2016. doi:10.1186/s12916-016-0763-7
- [367] A. D. Stookey, L. I. Katzel, G. Steinbrenner, M. Shaughnessy, and F. M. Ivey, “The Short Physical Performance Battery as a Predictor of Functional Capacity after Stroke,” *Journal of Stroke and Cerebrovascular Diseases*, vol. 23, no. 1, pp. 130–135, Jan. 2014. doi:10.1016/j.jstrokecerebrovasdis.2012.11.003
- [368] R. W. Bohannon and R. Crouch, “1-Minute Sit-to-Stand Test: Systematic Review of Procedures, Performance, and Clinimetric Properties,” *Journal of Cardiopulmonary Rehabilitation and Prevention*, vol. 39, no. 1, pp. 2–8, Jan. 2019. doi:10.1097/HCR.0000000000000336
- [369] N. J. Perkins and E. F. Schisterman, “The Inconsistency of “Optimal” Cut-points Using Two ROC Based Criteria,” *American journal of epidemiology*, vol. 163, no. 7, pp. 670–675, Apr. 2006. doi:10.1093/aje/kwj063
- [370] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, 3rd ed., ser. Wiley series in probability and statistics. Hoboken, New Jersey: Wiley, 2013, no. 398. ISBN 978-1-118-54835-6
- [371] R. Soangra and T. E. Lockhart, “Dual-Task Does Not Increase Slip and Fall Risk in Healthy Young and Older Adults during Walking,” *Applied Bionics and Biomechanics*, vol. 2017, pp. 1–12, 2017. doi:10.1155/2017/1014784
- [372] C. Li, J. Verghese, and R. Holtzer, “A comparison of two walking while talking paradigms in aging,” *Gait & posture*, vol. 40, no. 3, pp. 415–419, Jul. 2014. doi:10.1016/j.gaitpost.2014.05.062

- [373] M. Pau *et al.*, “Smoothness of Gait in Healthy and Cognitively Impaired Individuals: A Study on Italian Elderly Using Wearable Inertial Sensor,” *Sensors*, vol. 20, no. 12, p. 3577, Jun. 2020. doi:10.3390/s20123577
- [374] R. K. MacAulay, M. T. Wagner, D. Szeles, and N. J. Milano, “Improving Sensitivity to Detect Mild Cognitive Impairment: Cognitive Load Dual-Task Gait Speed Assessment,” *Journal of the International Neuropsychological Society*, vol. 23, no. 6, pp. 493–501, Jul. 2017. doi:10.1017/S1355617717000261
- [375] F. Pieruccini-Faria *et al.*, “Gait variability across neurodegenerative and cognitive disorders: Results from the Canadian Consortium of Neurodegeneration in Aging (CCNA) and the Gait and Brain Study,” *Alzheimer’s & Dementia*, vol. 17, no. 8, pp. 1317–1328, Aug. 2021. doi:10.1002/alz.12298
- [376] J. Choi *et al.*, “The Correlation between Cognition Screening Scores and Gait Status from Three-Dimensional Gait Analysis,” *Journal of Clinical Neurology*, vol. 15, no. 2, p. 152, 2019. doi:10.3988/jcn.2019.15.2.152
- [377] J. Verghese *et al.*, “Gait dysfunction in mild cognitive impairment syndromes,” *Journal of the American Geriatrics Society*, vol. 56, no. 7, pp. 1244–1251, Jul. 2008. doi:10.1111/j.1532-5415.2008.01758.x
- [378] T. Miyazaki *et al.*, “Relationships between Gait Regularity and Cognitive Function, including Cognitive Domains and Mild Cognitive Impairment, in Community-Dwelling Older People,” *Healthcare*, vol. 9, no. 11, p. 1571, Nov. 2021. doi:10.3390/healthcare9111571
- [379] H. G. Kang and J. B. Dingwell, “Separating the effects of age and walking speed on gait variability,” *Gait & Posture*, vol. 27, no. 4, pp. 572–577, May 2008. doi:10.1016/j.gaitpost.2007.07.009
- [380] E. G. James *et al.*, “Walking Speed Affects Gait Coordination and Variability Among Older Adults With and Without Mobility Limitations,” *Archives of Physical Medicine and Rehabilitation*, vol. 101, no. 8, pp. 1377–1382, Aug. 2020. doi:10.1016/j.apmr.2020.04.009
- [381] S.-K. Teh, I. Rawtaer, and H. P. Tan, “Predictive Accuracy of Digital Biomarker Technologies for Detection of Mild Cognitive Impairment and Pre-Frailty Amongst Older Adults: A Systematic Review and Meta-Analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3638–3648, Aug. 2022. doi:10.1109/JBHI.2022.3185798
- [382] C. Bellinger, S. Sharma, and N. Japkowicz, “One-Class versus Binary Classification: Which and When?” in *2012 11th International Conference on Machine Learning and Applications*. Boca Raton, FL, USA: IEEE, Dec. 2012. doi:10.1109/ICMLA.2012.212. ISBN 978-1-4673-4651-1 978-0-7695-4913-2 pp. 102–106.
- [383] N. Seliya, A. Abdollah Zadeh, and T. M. Khoshgoftaar, “A literature review on one-class classification and its potential applications in big data,” *Journal of Big Data*, vol. 8, no. 1, p. 122, Sep. 2021. doi:10.1186/s40537-021-00514-x
- [384] M. Hejazi and Y. P. Singh, “One-Class Support Vector Machines Approach to Anomaly Detection,” *Applied Artificial Intelligence*, vol. 27, no. 5, pp. 351–366, May 2013. doi:10.1080/08839514.2013.785791

- [385] A. Beghi *et al.*, “A One-Class SVM Based Tool for Machine Learning Novelty Detection in HVAC Chiller Systems,” *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 1953–1958, 2014. doi:10.3182/20140824-6-ZA-1003.02382
- [386] W. Rasheed and T. B. Tang, “Anomaly Detection of Moderate Traumatic Brain Injury Using Auto-Regularized Multi-Instance One-Class SVM,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 83–93, Jan. 2020. doi:10.1109/TNSRE.2019.2948798
- [387] S. Dreiseitl, M. Osl, C. Scheibböck, and M. Binder, “Outlier Detection with One-Class SVMs: An Application to Melanoma Prognosis,” *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2010, pp. 172–176, Nov. 2010.
- [388] J. Mourão-Miranda *et al.*, “Patient classification as an outlier detection problem: An application of the One-Class Support Vector Machine,” *NeuroImage*, vol. 58, no. 3, pp. 793–804, Oct. 2011. doi:10.1016/j.neuroimage.2011.06.042
- [389] R. Lang, R. Lu, C. Zhao, H. Qin, and G. Liu, “Graph-based semi-supervised one class support vector machine for detecting abnormal lung sounds,” *Applied Mathematics and Computation*, vol. 364, p. 124487, Jan. 2020. doi:10.1016/j.amc.2019.06.001
- [390] A. Retico, I. Gori, A. Giuliano, F. Muratori, and S. Calderoni, “One-Class Support Vector Machines Identify the Language and Default Mode Regions As Common Patterns of Structural Alterations in Young Children with Autism Spectrum Disorders,” *Frontiers in Neuroscience*, vol. 10, Jun. 2016. doi:10.3389/fnins.2016.00306
- [391] N. Naghavi and E. Wade, “Towards Real-Time Prediction of Freezing of Gait in Patients With Parkinson’s Disease: A Novel Deep One-Class Classifier,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 4, pp. 1726–1736, Apr. 2022. doi:10.1109/JBHI.2021.3103071
- [392] E. Manduchi, W. Fu, J. D. Romano, S. Ruberto, and J. H. Moore, “Embedding covariate adjustments in tree-based automated machine learning for biomedical big data analyses,” *BMC Bioinformatics*, vol. 21, no. 1, p. 430, Dec. 2020. doi:10.1186/s12859-020-03755-4
- [393] S. Freitas, M. R. Simões, L. Alves, and I. Santana, “Montreal Cognitive Assessment (MoCA): normative study for the Portuguese population,” *Journal of Clinical and Experimental Neuropsychology*, vol. 33, no. 9, pp. 989–996, Nov. 2011. doi:10.1080/13803395.2011.589374
- [394] T. Bristow, C.-S. Jih, A. Slabich, and J. Gunn, “Standardization and adult norms for the sequential subtracting tasks of serial 3’s and 7’s,” *Applied Neuropsychology. Adult*, vol. 23, no. 5, pp. 372–378, Oct. 2016. doi:10.1080/23279095.2016.1179504
- [395] K. P. Raghubar, M. A. Barnes, and S. A. Hecht, “Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches,” *Learning and Individual Differences*, vol. 20, no. 2, pp. 110–122, Apr. 2010. doi:10.1016/j.lindif.2009.10.005
- [396] S. Yamashita, K. Igarashi, and N. Ogihara, “Reducing the foot trajectory variabilities during walking through vibratory stimulation of the plantar surface of the foot,” *Scientific Reports*, vol. 11, no. 1, p. 7125, Mar. 2021. doi:10.1038/s41598-021-86583-7

- [397] J. Dukart, M. L. Schroeter, K. Mueller, and The Alzheimer's Disease Neuroimaging Initiative, "Age Correction in Dementia – Matching to a Healthy Brain," *PLoS ONE*, vol. 6, no. 7, p. e22193, Jul. 2011. doi:10.1371/journal.pone.0022193
- [398] M. Goni, S. B. Eickhoff, M. S. Far, K. R. Patil, and J. Dukart, "Smartphone-Based Digital Biomarkers for Parkinson's Disease in a Remotely-Administered Setting," *IEEE Access*, vol. 10, pp. 28 361–28 384, 2022. doi:10.1109/ACCESS.2022.3156659
- [399] K. A. Linn, B. Gaonkar, J. Doshi, C. Davatzikos, and R. T. Shinohara, "Addressing Confounding in Predictive Models with an Application to Neuroimaging," *The International Journal of Biostatistics*, vol. 12, no. 1, pp. 31–44, May 2016. doi:10.1515/ijb-2015-0030
- [400] R. Dinga, L. Schmaal, B. W. Penninx, D. J. Veltman, and A. F. Marquand, "Controlling for effects of confounding variables on machine learning predictions," *Bioinformatics*, preprint, Aug. 2020.
- [401] L. Snoek, S. Miletić, and H. S. Scholte, "How to control for confounds in decoding analyses of neuroimaging data," *NeuroImage*, vol. 184, pp. 741–760, Jan. 2019. doi:10.1016/j.neuroimage.2018.09.074
- [402] Y. Wang and J. Zheng, "An Evaluation of One-Class Feature Selection and Classification for Zero-Day Android Malware Detection," in *17th International Conference on Information Technology–New Generations (ITNG 2020)*, S. Latifi, Ed. Cham: Springer International Publishing, 2020, vol. 1134, pp. 105–111. ISBN 978-3-030-43019-1 978-3-030-43020-7
- [403] L. H. N. Lorena, A. C. P. L. F. Carvalho, and A. C. Lorena, "Filter Feature Selection for One-Class Classification," *Journal of Intelligent & Robotic Systems*, vol. 80, no. S1, pp. 227–243, Dec. 2015. doi:10.1007/s10846-014-0101-2
- [404] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, p. 066138, Jun. 2004. doi:10.1103/PhysRevE.69.066138
- [405] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, no. 1, p. 106, Dec. 2013. doi:10.1186/1471-2105-14-106
- [406] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class support vector machines for unsupervised anomaly detection," in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*. Chicago Illinois: ACM, Aug. 2013. doi:10.1145/2500853.2500857. ISBN 978-1-4503-2335-2 pp. 8–15.
- [407] B. Das, D. J. Cook, N. C. Krishnan, and M. Schmitter-Edgecombe, "One-Class Classification-Based Real-Time Activity Error Detection in Smart Homes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 5, pp. 914–923, Aug. 2016. doi:10.1109/JSTSP.2016.2535972
- [408] A. Airola, T. Pahikkala, W. Waegeman, B. De Baets, and T. Salakoski, "An experimental comparison of cross-validation techniques for estimating the area under the ROC curve," *Computational Statistics & Data Analysis*, vol. 55, no. 4, pp. 1828–1844, Apr. 2011. doi:10.1016/j.csda.2010.11.018
- [409] T. G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998. doi:10.1162/089976698300017197

- [410] K. M. Ishikawa, J. Davis, J. J. Chen, and E. Lim, “The prevalence of mild cognitive impairment by aspects of social isolation,” *PLOS ONE*, vol. 17, no. 6, p. e0269795, Jun. 2022. doi:10.1371/journal.pone.0269795
- [411] Q. Zhao, E. Adeli, and K. M. Pohl, “Training confounder-free deep learning models for medical applications,” *Nature Communications*, vol. 11, no. 1, p. 6010, Nov. 2020. doi:10.1038/s41467-020-19784-9
- [412] B. H. Zhang, B. Lemoine, and M. Mitchell, “Mitigating Unwanted Biases with Adversarial Learning,” in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. New Orleans LA USA: ACM, Dec. 2018. doi:10.1145/3278721.3278779. ISBN 978-1-4503-6012-8 pp. 335–340.
- [413] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, “Deep Learning for Anomaly Detection: A Review,” *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–38, Mar. 2022. doi:10.1145/3439950
- [414] S. Aam *et al.*, “Post-stroke Cognitive Impairment—Impact of Follow-Up Time and Stroke Subtype on Severity and Cognitive Profile: The Nor-COAST Study,” *Frontiers in Neurology*, vol. 11, p. 699, Jul. 2020. doi:10.3389/fneur.2020.00699
- [415] S. Melkas, H. Jokinen, M. Hietanen, and T. Erkinjuntti, “Poststroke cognitive impairment and dementia: prevalence, diagnosis, and treatment,” *Degenerative Neurological and Neuromuscular Disease*, p. 21, Feb. 2014. doi:10.2147/DNND.S37353
- [416] P. Calabrese *et al.*, “The assessment of cognitive and behavioural disturbances in vascular cognitive impairment (VCI) — recommendations of an expert working group,” *Neurologia i Neurochirurgia Polska*, vol. 55, no. 4, pp. 333–345, Aug. 2021. doi:10.5603/PJNNS.a2021.0035
- [417] L. Mellon *et al.*, “Cognitive impairment six months after ischaemic stroke: a profile from the ASPIRE-S study,” *BMC Neurology*, vol. 15, no. 1, p. 31, Dec. 2015. doi:10.1186/s12883-015-0288-2
- [418] A. Verdelho *et al.*, “Cognitive impairment in patients with cerebrovascular disease: A white paper from the links between stroke ESO Dementia Committee,” *European Stroke Journal*, vol. 6, no. 1, pp. 5–17, Mar. 2021. doi:10.1177/23969873211000258
- [419] M. D. Mijajlović *et al.*, “Post-stroke dementia – a comprehensive review,” *BMC Medicine*, vol. 15, no. 1, p. 11, Dec. 2017. doi:10.1186/s12916-017-0779-7
- [420] M. Barbay, M. Diouf, M. Roussel, O. Godefroy, and GRECOGVASC study group, “Systematic Review and Meta-Analysis of Prevalence in Post-Stroke Neurocognitive Disorders in Hospital-Based Studies,” *Dementia and Geriatric Cognitive Disorders*, vol. 46, no. 5-6, pp. 322–334, 2018. doi:10.1159/000492920
- [421] E. Sexton *et al.*, “Systematic review and meta-analysis of the prevalence of cognitive impairment no dementia in the first year post-stroke,” *European Stroke Journal*, vol. 4, no. 2, pp. 160–171, Jun. 2019. doi:10.1177/2396987318825484
- [422] D. A. Lowe, R. K. MacAulay, D. M. Szeles, N. J. Milano, and M. T. Wagner, “Dual-Task Gait Assessment in a Clinical Sample: Implications for Improved Detection of Mild Cognitive Impairment,” *The Journals of Gerontology: Series B*, vol. 75, no. 7, pp. 1372–1381, Aug. 2020. doi:10.1093/geronb/gbz119

- [423] K. Inasaridze, J. A. Foley, R. H. Logie, and S. D. Sala, “Dual Task Impairments in Vascular Dementia,” *Behavioural Neurology*, vol. 22, no. 1-2, pp. 45–52, 2010. doi:10.1155/2010/604919
- [424] M. Mateus, “Performance characterization of dual-task in community dwelling elderly people and mild vascular cognitive impairment,” Master’s thesis, Universidade de Lisboa, Faculdade de Medicina, Lisboa, Portugal, 2019. [Online]. Available: <http://hdl.handle.net/10451/39415>
- [425] S. M. K. Farhan *et al.*, “The Ontario Neurodegenerative Disease Research Initiative (ON-DRI),” *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*, vol. 44, no. 2, pp. 196–202, Mar. 2017. doi:10.1017/cjn.2016.415
- [426] K. M. Sunderland *et al.*, “The Ontario Neurodegenerative Disease Research Initiative,” *Neurology*, preprint, Aug. 2020.
- [427] S. M. H. Haddad *et al.*, “Comparison of diffusion tensor imaging metrics in normal-appearing white matter to cerebrovascular lesions and correlation with cerebrovascular disease risk factors and severity,” *Journal of Biomedical Imaging*, vol. 2022, jan 2022. doi:10.1155/2022/5860364
- [428] K.-B. Zaidi *et al.*, “Methods for Improving Screening for Vascular Cognitive Impairment Using the Montreal Cognitive Assessment,” *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques*, vol. 47, no. 6, pp. 756–763, Nov. 2020. doi:10.1017/cjn.2020.121
- [429] Ontario Brain Institute, “Brain-CODE Portal,” Jul. 2022. [Online]. Available: <https://www.braincode.ca/>
- [430] A. Rush *et al.*, “The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression,” *Biological Psychiatry*, vol. 54, no. 5, pp. 573–583, Sep. 2003. doi:10.1016/S0006-3223(02)01866-8
- [431] P. M. McLaughlin *et al.*, “The Quality Assurance and Quality Control Protocol for Neuropsychological Data Collection and Curation in the Ontario Neurodegenerative Disease Research Initiative (ONDRI) Study,” *Assessment*, vol. 28, no. 5, pp. 1267–1286, Jul. 2021. doi:10.1177/1073191120913933
- [432] A. A. Dillio *et al.*, “Association of apolipoprotein E variation with cognitive impairment across multiple neurodegenerative diagnoses,” *Neurobiology of Aging*, vol. 105, pp. 378.e1–378.e9, Sep. 2021. doi:10.1016/j.neurobiolaging.2021.04.011
- [433] A. J. Jak *et al.*, “Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment,” *The American Journal of Geriatric Psychiatry*, vol. 17, no. 5, pp. 368–375, May 2009. doi:10.1097/JGP.0b013e31819431d5
- [434] H. Nugroho, N. P. Utama, and K. Surendro, “Normalization and outlier removal in class center-based firefly algorithm for missing value imputation,” *Journal of Big Data*, vol. 8, no. 1, p. 129, Dec. 2021. doi:10.1186/s40537-021-00518-7

- [435] ONDRI Investigators *et al.*, “Motor Phenotype in Neurodegenerative Disorders: Gait and Balance Platform Study Design Protocol for the Ontario Neurodegenerative Research Initiative (ONDRI),” *Journal of Alzheimer’s Disease*, vol. 59, no. 2, pp. 707–721, Jul. 2017. doi:10.3233/JAD-170149
- [436] G. C. Cawley and N. L. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” *J. Mach. Learn. Res.*, vol. 11, p. 2079–2107, aug 2010.
- [437] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi:10.1023/A:1010933404324
- [438] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [439] A. Narasimhamurthy, “Theoretical bounds of majority voting performance for a binary classification problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1988–1995, Dec. 2005. doi:10.1109/TPAMI.2005.249
- [440] R. A. Armstrong, “When to use the Bonferroni correction,” *Ophthalmic and Physiological Optics*, vol. 34, no. 5, pp. 502–508, Sep. 2014. doi:10.1111/opo.12131
- [441] S. Seabold and J. Perktold, “statsmodels: Econometric and statistical modeling with python,” in *9th Python in Science Conference*, 2010.
- [442] V. Guimarães, I. Sousa, and M. V. Correia, “A Deep Learning Approach for Foot Trajectory Estimation in Gait Analysis Using Inertial Sensors,” *Sensors*, vol. 21, no. 22, p. 7517, Nov. 2021. doi:10.3390/s21227517
- [443] F. Petraglia *et al.*, “Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis,” *European Journal of Physical and Rehabilitation Medicine*, vol. 55, no. 2, pp. 265–280, Apr. 2019. doi:10.23736/S1973-9087.18.05306-6
- [444] I. L. Kyrдалen, P. Thingstad, L. Sandvik, and H. Ormstad, “Associations between gait speed and well-known fall risk factors among community-dwelling older adults,” *Physiotherapy Research International*, vol. 24, no. 1, p. e1743, Jan. 2019. doi:10.1002/pri.1743
- [445] H.-W. Jung *et al.*, “Usual gait speed is associated with frailty status, institutionalization, and mortality in community-dwelling rural older adults: a longitudinal analysis of the Aging Study of Pyeongchang Rural Area,” *Clinical Interventions in Aging*, vol. Volume 13, pp. 1079–1089, Jun. 2018. doi:10.2147/CIA.S166863