



# ADVANCING CLINICAL EVALUATION AND DIAGNOSTICS WITH ARTIFICIAL INTELLIGENCE TECHNOLOGIES

**Marco Recenti**

March 2023

Department of Engineering, School of Technology

Reykjavík University

**Ph.D. Dissertation**

ISBN: 978-9935-539-07-6 Electronic version

ISBN: 978-9935-539-06-9 Print version

ORCID Marco Recenti 0000-0001-9440-8434





# Advancing Clinical Evaluation and Diagnostics with Artificial Intelligence Technologies

by

Marco Recenti

Dissertation submitted to the Department of Engineering, School of  
Technology  
at Reykjavík University in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**

March 2023

Thesis Committee:

Paolo Gargiulo, Supervisor  
Professor, Reykjavík University, Iceland

Magnús Kjartan Gíslason, Co-advisor  
Associate Professor, Reykjavík University, Iceland

Antonio Fratini, Co-advisor  
Associate Professor, Aston University - Birmingham, UK

Hannes Petersen, Co-advisor  
Professor, Medical School, University of Iceland, Iceland

Leandro Pecchia, Examiner  
Professor, Bio-Medical Campus of the University of Rome, Italy

Copyright  
Marco Recenti  
March 2023

The undersigned hereby certify that they recommend to the Department of Engineering, School of Technology, Reykjavík University, that this dissertation entitled **Advancing Clinical Evaluation and Diagnostics with Artificial Intelligence Technologies**, submitted by **Marco Recenti**, be accepted as partial fulfilment of the requirements for the degree of **Doctor of Philosophy (Ph.D.) in Engineering**

.....  
date

.....  
Paolo Gargiulo, Supervisor  
Professor, Reykjavík University, Iceland

.....  
Magnús Kjartan Gíslason, Co-advisor  
Associate Professor, Reykjavík University, Iceland

.....  
Antonio Fratini, Co-advisor  
Associate Professor, Aston University - Birmingham, UK

.....  
Hannes Petersen, Co-advisor  
Professor, Medical School, University of Iceland, Iceland

.....  
Leandro Pecchia, Examiner  
Professor, Bio-Medical Campus of the University of Rome, Italy

The undersigned hereby grants permission to the Reykjavík University Library to reproduce single copies of this Dissertation entitled **Advancing Clinical Evaluation and Diagnostics with Artificial Intelligence Technologies** and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the Dissertation, and except as herein before provided, neither the Dissertation nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

.....  
date

.....  
Marco Recenti  
Doctor of Philosophy





# Advancing Clinical Evaluation and Diagnostics with Artificial Intelligence Technologies

Marco Recenti

March 2023

## Abstract

Machine Learning (ML) is extensively used in diverse healthcare applications to aid physicians in diagnosing and identifying associations, sometimes hidden, between different biomedical parameters. This PhD thesis investigates the interplay of medical images and biosignals to study the mechanisms of aging, knee cartilage degeneration, and Motion Sickness (MS).

The first study shows the predictive power of soft tissue radiodensitometric parameters from mid-thigh CT scans. We used data from the AGES-Reykjavik study, correlating soft tissue numerical profiles from 3,000 subjects with cardiac pathophysiologicals, hypertension, and diabetes. The results show the role of fat, muscle, and connective tissue in the evaluation of healthy aging. Moreover, we classify patients experiencing gait symptoms, neurological deficits, and a history of stroke in a Korean population, revealing the significant impact of cognitive dual-gait analysis when coupled with single-gait. The second study establishes new paradigms for knee cartilage assessment, correlating 2D and 3D medical image features obtained from CT and MRI scans. In the frame of the EU-project RESTORE we were able to classify degenerative, traumatic, and healthy cartilages based on their bone and cartilage features, as well as we determine the basis for the development of a patient-specific cartilage profile.

Finally, in the MS study, based on a virtual reality simulation synchronized with a moving platform and EEG, heart rate, and EMG, we extracted over 3,000 features and analyzed their importance in predicting MS symptoms, concussion in female athletes, and lifestyle influence. The MS features are extracted from the brain, muscle, heart, and from the movement of the center of pressure during the experiment and demonstrate their potential value to advance quantitative evaluation of postural control response.

This work demonstrates, through various studies, the importance of ML technologies in improving clinical evaluation and diagnosis contributing to advance our understanding of the mechanisms associated with pathological conditions.

**Keywords:** Machine Learning, Biomedical Data-set, Aging, Motion Sickness, Knee Cartilage, Medical Imaging, Biosignals

# Að Efla Klínískt Mat og Greiningu með Gervigreindartækni

Marco Recenti

mars 2023

## Útdráttur

Tölvulærdómur (Machine Learning eða ML) er algjörlega viðurkennt og nýtt í ýmsum heilbrigðisþjónustuviðskiptum til að hjálpa læknum við að greina og finna tengsl milli mismunandi líffærafræðilegra gilda, stundum dulinna. Þessi doktorsritgerð fjallar um samspil læknisfræðilegra mynda og lífsmerkja til að skoða eðli aldrunar, niðurbrot hnéhringjar og hreyfikerfissjúkdóms (Motion Sickness eða MS).

Fyrsta rannsóknin sýnir spáarkraft midjubeins-CT-skanna í því að fullyrða staðfestar meðalþyngdarlíkön, þar sem gögn úr AGES-Reykjavík-rannsókninni eru tengd við hjarta- og æðafræðilega sjúkdóma, blóðþrýstingsveikindi og sykursýki hjá 3.000 þátttakendum. Niðurstöðurnar sýna hlutverk fitu, vöðva og tengikjarna í mati á heilbrigðum öldrun. Þar að auki flokkum við sjúklinga sem upplifa gangvandamál, taugaeinkenni og sögu af heilablóðfalli í kóreanskri þjóð, þar sem einstök gangtaksskoðun er tengd saman við tvískoðun.

Önnur rannsóknin setur upp ný tölfræðisfræðileg umhverfisviðmið til matar á hnéhringju með samhengi 2D og 3D mynda sem aflað er úr CT og MRI-skömmtum. Í rauninni höfum við getuð flokkað niðurbrots-, slys- og heilbrigðar hnéhringjur á grundvelli bein- og brjóskmerkja með raun að sækja niðurstöður í umfjöllun um sjúklingar eftir réttu einkasniði.

Að lokum, í MS-rannsókninni, notum við myndræn tilraun samþættaða með hreyfanlegan grundvöll og EEG, hjartslátt, EMG þar sem yfir 3.000 aðgerðir eru útfránn og greindir til að átta sig á áhrifum MS, höfuðárás hjá konum sem eru íþróttamenn, lífsstíl og fleira. Einkenni MS eru aflöguð úr heilanum, vöðvum, hjarta og frá hreyfingum þyngdupunktsins á meðan tilraunin stendur og sýna mög.

**Efnisorð:** Tölvulærdómur, Læknisfræðilegt gagnasett, Aldrandi, Ferðaveiki, Hnébrjósk, Læknisfræðileg myndgreining, Lífsmerki

*To those who have loved me, love me now, and will love me in the future.*



# Acknowledgements

Although I am not adept at finding the most fitting words to express my gratitude towards all those who have contributed to my PhD journey over the past three years, I assure you that I have given it my all.

The first person who deserves my heartfelt gratitude is undoubtedly my mentor and supervisor, Prof. Paolo Gargiulo. When I joined in 2018 for my master thesis, I had no prior experience in the research field but a lot of aspirations. With your kindness, patience, passion, professionalism, friendship, and curiosity, I feel that I have not only become a better researcher but also a better person. I am immensely thankful for the opportunity that you gave me and for allowing me to work here as a teacher now. I am excited for this new phase of my life, where I get to continue working with you and the amazing team that you have built in the lab. It's a privilege to be a part of this great and inspiring work environment.

I would like to thank all the researchers and professors who supported and guided me throughout these years: a big thank to Magnus and Antonio for the help as co-advisors. Thank you to Hannes, for the knowledge about the clinical aspect of motion sickness and for the unforgettable summer weekends in the cottage and your gracious hospitality in Akureyri. I am eagerly looking forward to returning to the north soon. Thanks to professor Leandro Pecchia, for taking the time to read and evaluate my thesis.

I would like to express my gratitude to Dr. Milan Chang for her precious assistance in organizing my exchange program in Korea and her contributions to the related publications, both submitted and future. Additionally, I cannot thank Prof. Seung-Uk Ko enough for his support during my six-month stay at Chonnam National University in Yeosu, where he supervised my research on gait. Despite being on the other side of the planet, I felt at home thanks to his warmth and guidance. I would also like to extend my appreciation to the members of the Global Education Center, directed by Prof. Ko, for their assistance and for providing me with the opportunity to share my life experiences with local school children and teenagers.

Thanks to all the amazing people at the Institute of Biomedical and Neural Engineering, including those who visited our lab for various reasons and those I can't wait to see again in the future: Romain, Deborah, Kyle, Fabio, Riccardo, Federica, Lorena, Zakia, Gabriele, Federica, Andrea, Solveig, Gunnar, Halldor, Valdis, and Thordur and Oli. Words cannot describe how much I enjoyed the time, the discussions, the food, and the laughter in the office with each and every one of you. I would also like to express my gratitude to my "Icelandic" friends, both those who have passed through here, those who stayed and those who returned to Reykjavik: Giulio, Michelangelo, Camilla, Valentina, Sara, Giordano, Maria, Roberta, Elena, Sara, the people from the Erasmus Groups and those who are still here: Vittorio, Piergiorgio, Giulia, Goda, Binni, Dario, Viktor, Matteo, Brian, Solveig, Idunn, Urdur, Jillian, Renata, and ev-

erybody who shared with me nice moments in this wonderful city.

A big thanks to Carlo, who, as a friend and colleague, taught me a lot during my first months as a researcher. I co-authored most of my relevant publications with him, and it's no coincidence. Also, thanks to Alfonso and all the wonderful people from Naples. I'm sure there's still a lot to share in the future.

Thanks to all my friends in Alba: Andrea and Riccardo, it is incredible how much we have achieved even though we are on different sides of the globe, and thank you for always being straightforward with me. Alberto, Serena, Matteo, Alberto, Eugenio, Rizio, Febo, Momo, Simo, Steve, Cesco, Yass, Jack, Al, Anna, Mara, Madda, Chiara, Andre, and all the lovely people who will always welcome me back to my hometown or wherever in the world with true friendship and without any kind of fake attitudes. Thanks to Luca, Michele, Tommaso and Riccardo for the always inspiring and deep daily current affairs and old memories conversations.

Thanks to my referee friends, the ones who always really cared about me.

Thanks to Saeah: I shared with you the last incredible months of my life in Korea, Italy, Iceland and not only. I can't wait to explore together other sides of the world and share more and more and even more happy moments. Thanks for every single moment, past, present and future.

Thanks to my grandmothers, and all my relatives: it may sound cliché to say that family is important, but it truly is when everyone cares and respects each other. Thank you for your love and support throughout my journey.

I want to conclude by mentioning my family: my brother Filippo, and my parents, Antonella and Gianpiero. Thank you for your unwavering support from the very beginning, despite the distance, for being available whenever I needed you, and for your warm smiles and happiness every time I see you in Italy, Iceland, or even on a screen. I would never have been able to become who I am now and achieve these incredible accomplishments without you.

# Preface

This dissertation is an original work by the author, me, Marco Recenti.

I graduated as a information and telecommunication engineer at Polytechnic University of Torino in April 2019. During my master studies, I did my final thesis at the Institute of Biomedical and Neural Engineering (IBNE) of Reykjavik University, under the supervision of Paolo Gargiulo starting September 2018. The master thesis relative to AGES-Reykjavik Dataset perfectly fit my desire of combining my ICT expertise, my passion for the biomedical studies, and my big curiosity of living in such a wonderful country like Iceland.

Thanks to the opportunity that Paolo gave me, from June 2019 I started to work in IBNE as a researcher and I officially became a PhD candidate in January 2020. I enjoyed to contribute in the development and success of the Motion Sickness Lab and of the relative BioVRSea system as well as to contribute in numerous other researches like the European Restore Project. I had the great opportunity to do an exchange period of six months between May and October 2022 in Yeosu, South Korea, at the Chonnam National University where I worked with a elderly population gait dataset from the Seoul National University Hospital.

The main results achieved in these years are described below in the different chapters of this thesis.





# Publications Presented in the Thesis

Most of the text of the present PhD thesis has been adapted from some of the publications, of which I, Marco Recenti, am an author. I am affiliated with the Institute of Biomedical and Neural Engineering in the Engineering Department of Reykjavik University, Iceland.

All publications have been submitted for peer-review in their respective journals or conferences during the duration of my doctoral studies. Most of them are published and are available online, some of them, as indicated, are currently under the peer-review process.

I fully acknowledge that these papers are written in collaboration with my co-authors and that this thesis therefore contains some text and concepts that were conceived in collaboration. I hereby confirm that the inclusion of all previously published text was written either by myself and/or included here with the permission of my co-authors. The publications list follows a chronological order starting from the oldest:

- Recenti, M., Ricciardi, C., Gislason, M., Edmunds, K., Carraro, U., & Gargiulo, P. (2019, September). Machine learning algorithms predict body mass index using nonlinear trimodal regression analysis from computed tomography scans. In *Mediterranean Conference on Medical and Biological Engineering and Computing* (pp. 839-846). Springer, Cham; DOI: 10.1007/978-3-030-31635-8\_100.
- Ricciardi, C., Edmunds, K. J., Recenti, M., Sigurdsson, S., Gudnason, V., Carraro, U., & Gargiulo, P. (2020). Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions. *Scientific reports*, 10(1), 1-13; DOI: 10.1038/s41598-020-59873-9.
- Recenti, M., Ricciardi, C., Edmunds, K. J., Gislason, M. K., Sigurdsson, S., Carraro, U., & Gargiulo, P. (2020). Healthy aging within an image: Using muscle radiodensitometry and lifestyle factors to predict diabetes and hypertension. *IEEE Journal of Biomedical and Health Informatics*, 25(6), 2103-2112; DOI: 10.1109/JBHI.2020.3044158.
- Recenti, M., Ricciardi, C., Monet, A., Jacob, D., Ramos, J., Gislason, M., ... & Gargiulo, P. (2021). Predicting body mass index and isometric leg strength using soft tissue distributions from computed tomography scans. *Health and Technology*, 11(1), 239-249; DOI: 10.1007/s12553-020-00498-3.
- Recenti, M., Ricciardi, C., Aubonnet, R., Picone, I., Jacob, D., Svansson, H. Á., ... & Gargiulo, P. (2021). Toward predicting motion sickness using virtual reality and a moving platform assessing brain, muscles, and heart signals. *Frontiers in Bioengineering and Biotechnology*, 9, 635661; DOI: 10.3389/fbioe.2021.635661.

- Jacob, D., Unnsteinsdóttir Kristensen, I. S., Aubonnet, R., Recenti, M., Donisi, L., Ricciardi, C., ... & Gargiulo, P. (2022). Towards defining biomarkers to evaluate concussions using virtual reality and a moving platform (BioVRSea). *Scientific Reports*, 12(1), 1-22; DOI: 10.1038/s41598-022-12822-0.
- Aubonnet, R., Ramos, J., Recenti, M., Jacob, D., Ciliberti, F., Guerrini, L., ... & Gargiulo, P. (2022). Toward New Assessment of Knee Cartilage Degeneration. *Cartilage*, 19476035221144746; DOI: 10.1177/19476035221144746.
- Recenti, M., Jacob, D., Aubonnet, R., Burgunder, B., i Escalona, I. M., Gunnarsson, A. E., ... & Gargiulo, P. (2022, October). Predicting lifestyle using BioVRSea multi-biometric paradigms. In *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)* (pp. 329-334). IEEE; DOI: 10.1109/MetroXRINE54828.2022.9967685..
- Recenti, M., Gargiulo, P., Chang, M., Ko, S.B., Kim, T.J., & Ko, S.U. (2023). Predicting stroke, neurological and movement disorders using single and dual-task gait in Korean older population. *Gait&Posture*. (**Under Review**)

# Contents

<b>Acknowledgements</b>	<b>xiii</b>
<b>Preface</b>	<b>xv</b>
<b>Publications Presented in the Thesis</b>	<b>xvii</b>
<b>Contents</b>	<b>xix</b>
<b>List of Figures</b>	<b>xxii</b>
<b>List of Tables</b>	<b>xxv</b>
<b>List of Abbreviations</b>	<b>xxix</b>
<b>1 Prologue</b>	<b>1</b>
<b>2 Introduction</b>	<b>3</b>
2.1 Artificial Intelligence and Machine Learning . . . . .	3
2.1.1 Supervised and Unsupervised Approach . . . . .	3
2.1.2 Train-Test sets . . . . .	4
2.1.2.1 Cross Validation Techniques . . . . .	4
2.1.3 ML Algorithms . . . . .	5
2.1.3.1 Tree-Based Algorithms . . . . .	5
2.1.3.2 K-Nearest Neighbors . . . . .	7
2.1.3.3 Support Vector Machine . . . . .	7
2.1.3.4 Neural Networks and Multilayer Perceptron . . . . .	7
2.1.4 Evaluation Metrics . . . . .	8
2.1.5 Tuning of Hyperparameters . . . . .	10
2.1.6 Data Augmentation . . . . .	10
2.1.7 Feature Selection . . . . .	10
2.1.8 Deep Learning . . . . .	12
2.1.9 Applications in Healthcare . . . . .	12
2.1.9.1 Large Datasets . . . . .	12
2.1.9.2 Medical Imaging . . . . .	13
2.1.9.3 Biosignals . . . . .	14
2.2 Clinical and Diagnostics Applications . . . . .	17
2.2.1 Aging Study . . . . .	18
2.2.1.1 Sarcopenia and Cardiovascular Pathophysiology . . . . .	18
2.2.1.2 Healthy Lifestyle in Aging . . . . .	20
2.2.1.3 Single and Dual Gait Analysis . . . . .	20

2.2.2	Knee Osteoarthritis Study . . . . .	21
2.2.3	Motion Sickness and Postural Control Study . . . . .	22
2.2.3.1	Motion Sickness: pathogenesis theories and ML applications . . . . .	22
2.2.3.2	Sport Concussion and its Relation with Postural Control	23
2.2.3.3	Lifestyle Influence on Motion Sickness . . . . .	24
<b>3</b>	<b>Aging Study</b>	<b>27</b>
3.1	Soft Tissue Radiodensitometric parameters from mid-thigh CT Scan . .	27
3.1.1	AGES-Reykjavik data-set . . . . .	27
3.1.1.1	AGES Features . . . . .	27
3.1.1.2	Non Linear Trimodal Regression Analysis - NTRA . .	28
3.1.1.3	Body Mass Index and Isometric Leg Strength . . . . .	30
3.1.2	Cardiovascular Risk Assessment . . . . .	31
3.1.2.1	Results . . . . .	31
3.1.2.2	Discussion of the main findings . . . . .	38
3.1.3	NTRA & Lifestyle factors: Diabetes and Hypertension prediction	40
3.1.3.1	Materials and Methods . . . . .	40
3.1.3.2	Results . . . . .	42
3.1.3.3	Discussion of the main findings . . . . .	50
3.2	Single and Dual Gait Analysis in Older Korean Population . . . . .	53
3.2.1	Seoul National University Hospital Gait Dataset . . . . .	53
3.2.1.1	Materials and Methods . . . . .	53
3.2.1.2	Results . . . . .	55
3.2.1.3	Discussion . . . . .	59
<b>4</b>	<b>Knee Osteoarthritis Study</b>	<b>63</b>
4.1	European Union Horizon 2020 RESTORE Project . . . . .	63
4.1.1	Materials and Methods . . . . .	64
4.1.1.1	Participants . . . . .	64
4.1.1.2	2D features . . . . .	65
4.1.1.3	3D features . . . . .	69
4.1.1.4	Machine Learning . . . . .	69
4.1.2	Results . . . . .	70
4.1.2.1	2D and 3D Features . . . . .	70
4.1.2.2	Machine Learning . . . . .	71
4.1.3	Discussion . . . . .	74
<b>5</b>	<b>Motions Sickness and Postural Control Study</b>	<b>77</b>
5.1	Introduction . . . . .	77
5.1.1	BioVRSea definition . . . . .	77
5.1.2	BioVRSea Acquisition's Protocol . . . . .	78
5.1.3	Biomedical Data Acquisition . . . . .	80
5.1.4	BioVRSea cohorts . . . . .	81
5.1.5	Motion Sickness Questionnaire . . . . .	82
5.1.6	Motion Sickness and Lifestyle Indexes . . . . .	83
5.1.7	Sport Concussion Assessment Questionnaire . . . . .	86
5.2	Motion Sickness prediction with first cohort . . . . .	88
5.2.1	BioVRSea Setup for first cohort study . . . . .	88

5.2.1.1	Biomedical Features . . . . .	88
5.2.1.2	Statistical Analysis and ML tools . . . . .	89
5.2.2	Results . . . . .	91
5.2.2.1	Statistical Analysis . . . . .	91
5.2.2.2	Machine Learning . . . . .	94
5.2.3	Discussion . . . . .	95
5.3	Female Athletes Concussion Cohort . . . . .	97
5.3.1	BioVRSea Concussion Experiment Setup . . . . .	98
5.3.2	Results . . . . .	100
5.3.3	Discussion . . . . .	101
5.4	Lifestyle prediction with EEG and CoP signals . . . . .	103
5.4.1	BioVRSea Features for MS Lifestyle study . . . . .	103
5.4.2	Results . . . . .	105
5.4.3	Discussion . . . . .	107
<b>6</b>	<b>Conclusion and Future Development</b>	<b>109</b>
<b>7</b>	<b>List of Publications from the Author</b>	<b>113</b>
	<b>Bibliography</b>	<b>117</b>

# List of Figures

2.1	Random Forest Classification Diagram (N=Number of Decision Trees in the Forest) . . . . .	6
2.2	Artificial Neural Network Representation with the respective layers, nodes and links . . . . .	8
2.3	Filter and Wrapper Feature Selection Diagrams . . . . .	12
2.4	Short description of the major diagnostic medical images with brain images examples [91] . . . . .	15
2.5	Basic EEG setup: EEG cap, amplifier, software and computer, from ANT-neuro, eegoTM mylab [115] . . . . .	16
2.6	Workflow for the Machine Learning approach in Digital Health Engineering. . . . .	19
3.1	Mid-tigh CT Scan from the AGES data-set: on the right fat is orange, muscle is red and connective tissue is blue . . . . .	29
3.2	Non Linear Trimodal Regression Analysis - NTRA: visual representation of the 11 parameters . . . . .	29
3.3	Workflow of the NTRA study on Cardiovascular risk assessment . . . . .	31
3.4	ROC curves for coronary heart disease (CHD), cardiovascular disease (CVD) and chronic heart failure (CHF) classification with K-fold cross-validation and nonlinear trimodal regression analysis by $k = 12$ . . . . .	34
3.5	Results from tissue-based machine learning feature importance. (A) Example of a segmented false-color CT cross-section to illustrate the morphology of fat (orange), loose connective (blue), and lean muscle (red) tissue. (B) Total model accuracy (%) for each algorithm and cardiac pathophysiology, visually illustrating (with analogous colors) the compositional accuracy of each model with respect to tissue type. (C) Compositional accuracy (%) for each model with respect to tissue type. . . . .	36
3.6	Workflow of the IEEE Journal of Biomedical and Health Informatics Study . . . . .	41
3.7	Three-Level binary tree of the IEEE Journal of Biomedical and Health Informatics Study . . . . .	42
3.8	Comparative representation of tissue importance in predicting DM and HTN from NTRA parameters. . . . .	49
3.9	Scatterplots with regression lines for four of the most significant features in the classification of subjects with and without an history of strokes in relation to age. . . . .	59
3.10	Scatterplots with regression lines for four of the most significant features in the classification of subjects with and without gait symptoms in relation to age. . . . .	60
3.11	Scatterplots with regression lines for four of the most significant features in the classification of subjects with and without focal neurological deficit in relation to age. . . . .	61

4.1	Graphical Abstract of the Cartilage Journal manuscript . . . . .	64
4.2	Bone observations. . . . .	66
4.3	Cartilage and joint space observations. . . . .	67
4.4	Femoral cartilage thickness measurements, medial and lateral compartments. Fitting cylinder method to obtain 3 regions of interest, anterior (-30o-0o), medial (0o-30o) and posterior (30o-60o) in the lateral compartment (A) and the medial compartment (B). . . . .	68
4.5	tibial cartilage thickness measurements. anterior, middle and posterior points are measured along the tibial cartilage in medial (A) and lateral (B) compartments. . . . .	68
4.6	Cartilage thickness measurements, femoropatellar compartment. Measurement of the articular cartilage thickness at 3 points on the patella (A) and articular cartilage thickness of the femoral trochlea on 2 different points (B). . . . .	69
4.7	Segmentation workflow for 3D analysis . . . . .	70
4.8	3D model from the registration for the 3 groups of patients (A) Degenerative (B) traumatic (C) Control. . . . .	70
5.1	A subject on the BioVRSea platform wearing the VR goggles during the sea virtual simulation in the Motion Sickness Lab at Reykjavik University . . . . .	78
5.2	BioVrSea Acquisition's Protocol . . . . .	79
5.3	BioVRSea Biosgnals setup: EEG wet cap and EMG sensors, gastrocnemius lateral, and soleus muscles. . . . .	80
5.4	Force Plates and VR goggles on the moving platform of BioVRSea . . . . .	81
5.5	Age and Gender piecharts relative to MSSQ answers from 355 subjects . . . . .	82
5.6	Three classes LSI piechart from 355 subjects . . . . .	84
5.7	MSPI piechart from 355 subjects . . . . .	85
5.8	BVSEI piechart from 355 subjects . . . . .	88
5.9	BioVRSea structure for the present study: the moving platform, is combined with a rough sea VR scenario and with EEG, EMG, and HR bio-signal acquisition to predict MS using ML. . . . .	89
5.10	Description of the 19 biometric parameters that compose the database. . . . .	90
5.11	Difference of the objective brain, muscle, and health bio measurements between the first static protocol and the light (1 Hz – green) and the hard one (3 Hz – red) for all the patients (the one that did not perform the 3-Hz protocol is not included). . . . .	91
5.12	Percentage of the MSSQ answers for each symptom, and percentage of zeros and ones for the eight computed indexes. . . . .	91
5.13	Difference of the subjective MS symptoms between the first static protocol and the light (1 Hz – green) and the hard (3 Hz – red) for all the patients (the one that did not perform the 3-Hz protocol is not included). . . . .	92
5.14	Significance of the 19 biometric parameters calculated with the univariate statistical analysis (Mann–Whitney test) for all the eight indexes. . . . .	92
5.15	Feature importance (%) for IPV, INM, and IMS using Random Forest algorithm. . . . .	94
5.16	BioVRSea experimental Setup for the Present Concussion Study . . . . .	98

5.17	Feature Selection and Machine Learning Workflow. 51 features are selected from Brain, Muscles and Heart signal and on them, the PCA is performed to obtain 9 PCA features used to the binary classification of the concussion assessment using different ML algorithms; 22 SCAT5 features are then used as well as features for the concussion classification; the 22 SCAT5 and 9 PCA are then combined and used together to classify concussed and non-concussed subjects. . . . .	99
5.18	Symptoms (SCAT5) reported by all participants before and after VR acquisition. Overview of concussion symptoms reported by all participants before VR acquisition. . . . .	100
5.19	The results of 5 different algorithms (RF- Random Forest, SVM—Support Vector Machine, ADA-B— Adaptive Boosting , MLP- Multilayer Perceptron, GB-Gradient Boosted) and combinations of features. Highest accuracy in all algorithms is achieved using both the PCA and SCAT5 features, but highest accuracy overall (95.5%) when computed with an SVM algorithm.	102
5.20	BioVRSea: workflow for the study presented in Rome at IEEE MetroX-RAINE Conference, October 2022. . . . .	104
5.21	Boxplot for the 3 most important features for the lifestyle index classification. The red indicates the unhealthy people while green is referred to healthy participants. . . . .	107



# List of Tables

2.1	Classification metrics formulas based on TP, FP, TN, FN definitions . . . . .	9
3.1	Main features of the AGES Reykjavik data-set . . . . .	28
3.2	The evaluation metrics for the BMI regression prediction . . . . .	30
3.3	AGES-I: Summary statistics and nonlinear trimodal regression analysis parameters with relative standard deviation (SD) by cardiac pathophysiology (coronary heart disease (CHD), cardiovascular disease (CVD), chronic heart failure (CHF), and no condition). . . . .	32
3.4	AGES-II: Summary statistics and nonlinear trimodal regression analysis parameters with relative standard deviation (SD) by cardiac pathophysiology (coronary heart disease (CHD), cardiovascular disease (CVD), chronic heart failure (CHF), and no condition). Note: *From the total sample size of $n = 3,157$ subjects that participated in both the AGES-I and AGES-II studies, 585 individuals presented with more than one cardiac pathophysiology. . . . .	33
3.5	The 11 nonlinear trimodal regression analysis parameters were used to assess cardiovascular risks through machine learning algorithms. The evaluation metrics [%] by cardiac pathophysiology were computed. . . . .	33
3.6	The 11 nonlinear trimodal regression analysis parameters grouped by tissue type (fat, connective and muscle) were used to assess cardiovascular risks through machine learning algorithms and evaluation metrics [%] were computed. . . . .	35
3.7	The 11 nonlinear trimodal regression analysis parameters were used to assess cardiovascular risks on subjects grouped by age [years] through ML algorithms and evaluation metrics [%] were computed. . . . .	37
3.8	The 11 nonlinear trimodal regression analysis parameters from AGES-I were used to predict the presence of chronic heart failure in AGES-II through machine learning algorithms and evaluation metrics [%] were computed. . . . .	38
3.9	The Influence Of Lifestyle Health On Ntra Parameters (* $p < 0.05$ ; ** $p < 0.01$ ; *** $p < 0.001$ ) . . . . .	42
3.10	The Influence of DM And HTN on NTRA Parameters Between LSH And LSNH Individuals (* $p < 0.05$ ; ** $p < 0.01$ ; *** $p < 0.001$ ) . . . . .	44
3.11	The Influence of Cardiac Pathophysiology on NTRA Parameters Between LSH and LSNH Subjects affected by either DM or HTN (* $p < 0.05$ ; ** $p < 0.01$ ; *** $p < 0.001$ ) . . . . .	45
3.12	Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms Applied to the Present Tree with SMOTE . . . . .	45
3.13	Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms Applied to the Present Tree without SMOTE . . . . .	46

3.14	Classification metrics [%] of DM and HTN, Independent from Lifestyle Index with SMOTE . . . . .	46
3.15	Classification metrics [%] of DM and HTN, Independent from Lifestyle Index without SMOTE . . . . .	47
3.16	Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms when grouping NTRA Parameters by tissue type with SMOTE . . . . .	47
3.17	Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms when grouping NTRA Parameters by tissue type without SMOTE . . . . .	48
3.18	Evaluation Metrics [%] in the Longitudinal Prediction of DM Incidence from AGES-I to AGES-II with SMOTE . . . . .	49
3.19	Evaluation Metrics [%] in the Longitudinal Prediction of DM Incidence from AGES-I to AGES-II without SMOTE . . . . .	49
3.20	Sex differences and mean age in respect to ST, GS, and NS binary features	56
3.21	The influence of ST, GS, and NS on GAIT features . . . . .	56
3.22	ML learning models evaluation metrics for the ST, GS, NS, binary classification . . . . .	57
3.23	Features groups relevance in % for all the RF classification models . . . . .	57
3.24	Most important features and relative importance in % for all the RF classification models using Gait + Clinical Features Selection . . . . .	58
3.25	Most important features and relative importance in % for all the RF classification models using Gait Features Selection . . . . .	58
4.1	Description of the patients demographics (age, sex) by group . . . . .	64
4.2	Feature Selection Sets used as inputs for the ML analysis . . . . .	71
4.3	2D measurements: Summary of bone pathologies, meniscal pathology and synovitis-effusion as percentages for each group of patients. . . . .	72
4.4	2D measurements: Average values of the ICRS grading, CI and ACT for each group (and standard deviation between parentheses) and their location (compartment/bone). . . . .	73
4.5	3D measurements: The results show the average variable for each group (with standard deviation between parentheses) . . . . .	73
4.6	classification metrics (Recall (Re) Precision (Pr) and F1 [%]) for the 2 different tree-based ML algorithms and the seven different features selections (Degenerative (D) - Traumatic (T) - Control (C) . . . . .	74
4.7	12 most important features [%] for the RF classification model with 96 Tot Features . . . . .	74
4.8	Importance of the groups of features [%] for the RF classification model with 96 Tot Features . . . . .	75
5.1	VR/platform synchronization protocol . . . . .	79
5.2	Percentages of subjects related to Lifestyle Index and Motion Sickness Proneness Index . . . . .	87
5.3	Percentages of subjects related to Lifestyle Index and BioVRSea Effect Index	87
5.4	Percentages of subjects related to Motion Sickness Proneness Index and BioVRSea Effect Index . . . . .	87
5.5	Evaluation metrics [%] after the classification ML analysis for IPV, INM, and IMS. (The bold values are the most significant results) . . . . .	93

5.6	Feature importance (%) for IPV, INM, and IMS using Random Forest algorithm. (The marked values are the three most significant features for each index classification) . . . . .	93
5.7	Age and athletic status of participants. *One athlete without a history of concussion did not clarify if she was retired or still active (3.6%). . . . .	98
5.8	The 51 selected features from the BioVRSea measurements. . . . .	99
5.9	Concussion classification evaluation metrics for the three different feature selections proposed: The 9 PCA features; the 22 SCAT5 features; the 31 combined features of PCA and SCAT5. . . . .	101
5.10	ML Results on Different Feature Groups with Tuning Based on Recall . . .	105
5.11	ML Results on Different Feature Groups with Tuning Based on Accuracy .	105
5.12	Most Significant CoP Features for the highest accuracy and recall with KNN algorithm . . . . .	106



# List of Abbreviations

ADA-B - Ada Boosting  
AF - Atrial Fibrillation  
AG - Ahlbäck grading  
AGES - Age Gene/Environment Susceptibility Study  
AI - Artificial Intelligence  
ANN - Artificial Neural Network  
ANOVA - Analysis of Variance  
AUCROC - Area Under the Curve Receiver Operating Characteristics  
BMI - Body Mass Index  
BVSEI - BioVRSea Effect Index  
C - Control (Chapter 4.1)  
CAD - Coronary Artery Disease  
CHD - Coronary Heart Disease  
CHF - Chronic Heart Failure  
CI - Cumulative Index  
cLDB - chondral lesion Database  
CNN - Convolutional Neural Network  
CNS - Central Nervous System  
CoP - Center of Pressure  
CT-Scan - Computed Tomography Scan  
CV - Cross Validation  
CVD - Cardiovascular Disease  
D - Degenerative (Chapter 4.1)  
DL - Deep Learning  
DL - Dyslipidemia (Chapter 3.2)  
DM - Diabetes Mellitus  
DT - Decision Tree  
EEG - Electroencephalogram  
EGG - Electrogastrogram  
EMG - Electromyogram  
EOG - Electro-Oculography  
EX-T - Extra Tree  
F1 - F Score  
fMRI - Functional Magnetic Resonance Imaging  
FN - False Negative  
FP - False Positive  
GB or GRAD-B - Gradient Boosting  
GL - Gastrocnemius Lateralis  
GS - Gait Symptoms

HTN or HT - Hypertension  
HR - Heart Rate  
HU - Hounsfield Unit  
ICRS - International Cartilage Repair Society  
ICT - Information and Communication Technologies  
IDizz - Dizziness and Vertigo Symptom Index  
Ifatig - Fatigue Related Symptoms Index  
IGenDis - General Discomfort Symptom Index  
Ihead -Head Related Symptoms Index  
INM - Neurological Muscle Strain Related Symptoms Index  
IMS - Motion Sickness Index  
IPV - Physiological/Vegetative Related Symptoms Index  
ISO - Isometric Leg Strength  
Isto - Stomach Related Symptoms Index  
JI - Jaccard Index (Accuracy)  
KNN - K Nearest Neighbors  
LG - Low Gamma  
LSI - Lifestyle Index  
LSH - Lifestyle Healthy  
LSNH - Lifestyle Not Healthy  
MAE - Mean Absolute Error  
ML - Machine Learning  
MLP - Multilayer Perceptron  
MRI - Magnetic Resonance Imaging  
MS - Motion Sickness  
MSE - Mean Squared Error  
MSI - Motion Sickness Index  
MSPI - Motion Sickness Proneness Index  
MSS - Motion Sickness Susceptibility  
MSSQ - Motion Sickness Susceptibility Questionnaire  
MVELO - Mean Velocity of Point on Support Plane (TOTEX/Total Time)  
mTBI - mild Traumatic Brain Injury  
N - Amplitude of NTRA  
NN - Neural Network  
NS - Neurological Sequelae  
NTRA - Nonlinear Trimodal Regression Analysis  
OA - Osteoarthritis  
PC - Postural Control  
PCA - Principal Component Analysis  
PO - Dual Task features (Chapter 3.2)  
PONV - Post-Operative Nausea and Vomiting  
PSD - Power Spectral Density  
PVE - Partial Volume Effect  
PY - Python  
 $R^2$  - Coefficient of Determination  
RF - Random Forest  
RMSE - Root Mean Square Error  
ROC - Receiver Operating Characteristics  
ROI - Region of interest

S - Soleus (Chapter 5)  
SCAT5 - Sport Concussion Assessment Tool version 5  
SCT - Sensory Conflict Theory  
SD or STD - Standard Deviation  
SL - Scikit-Learn  
SMOTE - Synthetic Minority Oversampling Technique  
ST - Stroke  
StepL - Step Length  
StepW - Step Width  
StrideL - Stride Length  
SVM - Support Vector Machine  
T - Traumatic (Chapter 4.1)  
TA - Tibialis Anterioris  
THA - Total Hip Arthroplasty  
TKA - Total Knee Arthroplasty  
TN - True Negative  
TOTEX - Summation over the elementary movement between consecutive samples on the support plan (CoP feature - Chapter 5)  
TP - True Positive  
VR - Virtual Reality  
 $\alpha$  - skewness of NTRA  
 $\mu$  - Location of NTRA  
 $\sigma$  - Width of NTRA





# Chapter 1

## Prologue

This thesis provides Artificial Intelligence (AI) and Machine Learning (ML) technology tools and illustrate different applications in the digital health and biomedical engineering fields. The researches are all related to the different study areas of the Institute of Biomedical and Neural Engineering in the Engineering Department at Reykjavik University. This dissertation will focus mainly on the use of features extracted from medical images and biosignals and their application on three main study subjects: aging, knee osteoarthritis and cartilage degeneration, and motion sickness and postural control.

The remainder of this thesis is structured in six chapters:

- **Chapter 2:** firstly introduces briefly the main ML technologies, then shows the methods, state of the art and the possible various applications in the biomedical engineering field with a focus on the importance of organizing the research based on a ML analysis having in mind a clear and specific clinical purpose. Finally an introduction relative to the specific biomedical areas of study covered in this thesis is presented.
- **Chapter 3:** presents the researches relative to the first area of study: healthy aging. The chapter firstly investigates the AGES-Rekjavik Dataset and the use of soft tissue radiodensitometric parameters extracted from a mid-thigh CT-Scan to classify elderly subjects affected by different comorbidities. After proving that NTRA parameters have a predictive value regarding BMI and ISO, the first main section relative to AGES dataset presents and discusses the results obtained studying radiodensitometry and cardiovascular diseases [1]. The second section about AGES dataset provides the impactful results of NTRA related to healthy lifestyle of older individuals and their diabetes and hypertension incidence [2]. The chapter finally describes the results of the research on Single and Dual Gait data analysis based on older Korean Population. The study presented, actually under review [3], was done at Chonnam National University in Yeosu, South Korea, under the co-supervision of Prof. Seung-Uk Ko in the frame of the international exchange program supported by Reykjavik University.
- **Chapter 4:** presents the research focused on the knee osteoarthritis and cartilage degeneration study: it is developed in the frame work of the RESTORE

European Project, and focus on knee cartilage assessment with the goal of extracting new metrics to quantify cartilage degeneration. The study presented points out the benefit of using features from 2D and 3D image analysis combined with ML models to improve the assessment of cartilage degeneration [4].

- **Chapter 5:** presents the researches performed using biosignals features focusing on motion sickness, concussion and postural control studies. Firstly the BioVRSea system, its relative protocols, cohorts, biosignals and indexes are described in details; successively the results obtained with the first cohort is presented [5]; then the results obtained from a group of professional female athletes with an history of concussion are shown [6]; finally the research focuses on BioVRSea latest cohort and motion sickness relation to the healthy and unhealthy lifestyle [7].
- **Chapter 6:** this final section is dedicated to the general conclusions of the work with its relative discussion and future developments.

# Chapter 2

## Introduction

### 2.1 Artificial Intelligence and Machine Learning

*"Can machines do what we (as thinking entities) can do?"*

In 1950, Alan Turing, the father of theoretical computer science and artificial intelligence, posed the question that initiated the artificial intelligence revolution of the 21st century, along with its consequential scientific and philosophical debates that persist to this day [8].

*"Machine learning is the study of computer algorithms that allow computer programs to automatically improve performances through experience."*

Tom Mitchell gave a concise definition of Machine Learning as an application of artificial intelligence that provides the ability to automatically learn and improve from experience to make prediction or decisions without being explicitly programmed to perform the task [9].

Nowadays, Machine Learning (ML) technologies are commonly employed in various fields, including healthcare applications. They can assist physicians in diagnosis and the identification of possible hidden links between biomedical parameters from images or biosignals and the diseases or health status of different groups of individuals.

This PhD thesis aims to explore various applications of Artificial Intelligence (AI) technologies, particularly ML, in the digital health and biomedical engineering fields. This chapter provides a brief introduction to the principal ML techniques and technologies used in this study. Subsequently, the importance of the clinical objective in the digital health application of ML is discussed, alongside the areas of interest covered in this thesis, such as aging, cartilage degeneration, and motion sickness studies.

#### 2.1.1 Supervised and Unsupervised Approach

Within the field of AI and ML, two essential approaches exist: supervised and unsupervised learning [10]. The main difference between them is that the former employs pre-existing results to predict outcomes, while the latter does not.

The goal of supervised learning is to predict an output or a target that is already known: it uses labeled datasets designed with the specific scope of training an algorithm to classify two or more classes or predict a continuous value in the most possible accurate way. The model can in this way quantify its efficiency and learn, improving

over time.

Supervised learning can be split into two problems categories: classification and regression. The first has the scope of assigning the test data to a specific group, defined as class. Classes can be two (binary classification) or more (multiclass classification). The regression algorithms predict a continuous value.

Unsupervised learning uses diverse ML algorithms to cluster unlabeled data. They can explore data hidden patterns in autonomy without the human mediation, that is why they are called "unsupervised". They are performed for three main tasks defined as clustering, association and dimensionality reduction.

The design and employment of unsupervised learning models is out of the scope of the present thesis. The main focus is the design of optimal datasets and ML models for a supervised (mainly classification) learning analysis: this will be discussed presenting different biomedical and clinic applications that operate with features extracted from medical images and biosignals.

### 2.1.2 Train-Test sets

The train-test split is a commonly used technique for reliably evaluating the performance of ML models. This technique involves dividing a dataset into two distinct subsets. The first subset, known as the train set, is utilized for fitting or training the ML model. The second subset, known as the test set, is used to assess the model's fit. In supervised learning, the input element of the dataset is provided, and the model's predictions are compared to the known values.

The primary objective of an optimal train/test division is to evaluate the performance of the model on new data that the model has not yet encountered during training. While an 80/20 or 70/30 train/test split is a common approach, the most reliable technique for estimating model performance is the Cross Validation (CV). Various CV techniques exist, but they share the same fundamental concept: executing a loop that initially partitions the data into multiple subsets. Then, a set is held out and used as the test set while the model is trained on the remaining sets [11].

#### 2.1.2.1 Cross Validation Techniques

K-Fold CV is the most popular and efficient technique used for CV. This procedure uses a parameter  $k$ , which refers to the number of splits applied to the entire dataset, hence it is called k-fold cross-validation. The dataset is split into  $k$  consecutive folds, where  $(k - 1)$  folds are used for training the model, and the remaining 1 fold is used for testing. The training is done for each combination of folds. The value of  $k$  is chosen such that each train/test group of data samples is large enough to be statistically representative [11]. It has been demonstrated that an unbalanced dimension of folds does not allow obtaining reliable results. Kohavi in 1995 [12] has demonstrated using different datasets that k-fold CV is most efficient when  $k = 10$ . However, it is possible to achieve equally efficient results with different values of  $k$  or different fold combinations. For example, the Leave-One-Out procedure uses  $k = n$ , where  $n$  is the number of samples considered. In this procedure only one sample is removed from the training set, and it is often performed for small dataset analysis [13].

Another application of k-fold CV is the stratified k-fold technique. This method guarantees a more balanced distribution of classes in each test set, and it can be optimal for unbalanced ML problems. This technique works by returning stratified folds, where

each fold contains approximately the same percentage of samples of each class as the full dataset [14].

K-Fold CV is applied in all the research presented in the following chapters. In particular, stratified k-fold CV is performed for the models in Chapter 5.4.

### 2.1.3 ML Algorithms

At present, there are numerous supervised ML algorithms available in the literature, and their effectiveness can vary depending on the input features and classification or regression objectives. Below, you will find a brief description of the algorithms proposed in this thesis.

#### 2.1.3.1 Tree-Based Algorithms

This group of algorithms is considered very efficient in terms of quality of results and impact on clinical problems [15]. They are denominated Tree Based because the basic units on which they are built are regression and classification trees - Decision Trees (DT) [16]. DT is a greedy algorithm that performs a recursive binary partitioning of the feature space. The tree predicts the same label for each leaf partition. Each partition is chosen greedily by selecting the best split from a set of possible splits, in order to maximize the information gain at a tree node. DT are easy and efficient as they do not need any normalization of the input data. They offer a regular system of binning patients, and they are far more efficient if the single tree is expanded to forests and ensemble trees with the main goal of achieving better classification accuracy and addressing instability that affects the single trees. Furthermore, clinicians find them appealing because they can offer a "very human data representation" [17]. The ensemble learning techniques of randomization, bagging, and boosting can efficiently be applied to decision trees. DT can be a weak and unstable learner, and ensemble techniques are useful to improve the performances and reduce the noise in the dataset [18].

**Random Forest** The first tree-based ML method employed in this thesis is the Random Forests (RF) ensemble learning method, which was firstly introduced by Tin Kam Ho in 1995 [19] and was subsequently discussed in detail in a book by Leo Breiman [20]. RF features Decision Trees (DT) that share identical basic properties and the ability to avoid overfitting. The set of DT used in RF are trained separately so that the training can be executed in parallel with the others. However, some randomness is injected into the training process to reduce the variance of the predictions. Randomness is introduced by subsampling the original dataset on each iteration to obtain a different training set, or by considering different random subsets of features to split on at each tree node. To make a prediction on a new instance, an RF must aggregate the predictions from its set of DT. In the case of classification, the aggregation is done by majority vote, where each prediction is counted as a vote for one class, and the label is predicted to be the class which receives the most votes. In the case of regression, the mean of the obtained predictions of the individual trees is evaluated as output. A simplified diagram of the RF classification algorithm is shown in Fig. 2.1.

**Extremely Randomized Tree** Extremely Randomized Trees (EX-T) take randomness one step further in the way splits are computed [21]. Like in RF, a random subset

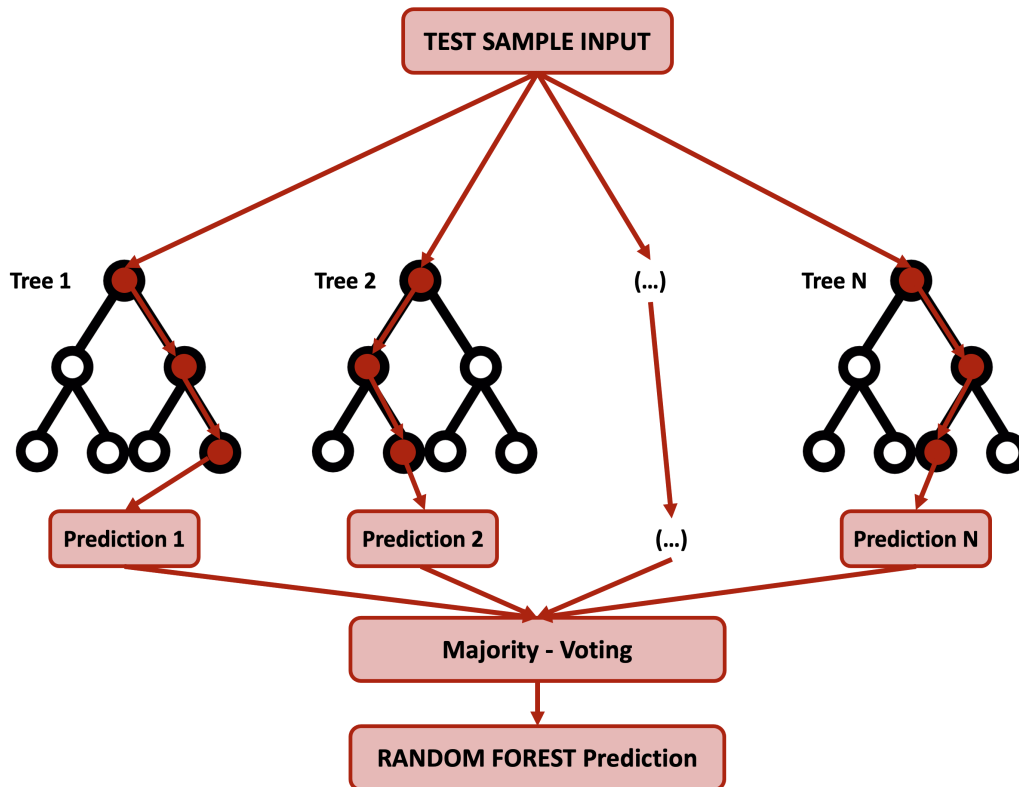


Figure 2.1: Random Forest Classification Diagram ( $N$ =Number of Decision Trees in the Forest)

of candidate features is used, but, instead of looking for the most discriminant thresholds, these are drawn randomly for each possible feature, and the best of these random thresholds is picked as the splitting rule. The random selection of the threshold allows further reduction of the variance of the model.

**Ada-Boost** Ada-Boost (ADA-B) is an ensemble method belonging to the boosting family, whose core principle is to strengthen weak learners [22]. During ADA-B training, only the parameters that improve the predictive power of the model are selected, reducing model complexity in terms of dimension and thereby improving execution time. During the boosting iteration, data modifications are applied by assigning weights to every training sample, starting with the original training data. For all other successive iterations, sample weights are modified, and the learning algorithm is applied again to the data with its new weight. At a given step, samples used for training that were wrongly predicted by the boosted model in the previous step have their weights increased, whereas these weights are decreased for examples that were predicted correctly. As the iterations proceed, samples that are difficult to predict/-diagnose receive ever-increasing influence. Each sequential weak learner is then forced to concentrate on samples that were previously missed [23].

**Gradient Boosting** Gradient Boosting (GB) produces highly competitive and robust models for both classification and regression tasks. It is especially appropriate for mining sub-optimally clean data, and the implementation proposed in this thesis follows the algorithm proposed by Friedman [24]. This method not only employs ran-

domization and bagging principles but also includes a special form of boosting to build an ensemble of weak models, specifically decision tree models.

### 2.1.3.2 K-Nearest Neighbors

K-nearest neighbors (KNN) is a non-parametric classification algorithm that assigns a class to a new instance based on the class of its k-nearest neighbors in the training set, where "nearest" is defined by a distance metric [25]. KNN is capable of solving problems related to overfitting, small datasets, non-linear and/or high-dimensional data, and can be used for both classification and regression. The algorithm calculates the distances between the new instance and all the instances in the training set and selects the k-nearest neighbors. The class of the new instance is then assigned as the majority class among its k-nearest neighbors. If  $k=1$ , the algorithm is called the nearest neighbor algorithm. A distance metric is used to measure the similarity between instances. The most common distance metric used is Euclidean distance, but other metrics, such as Manhattan distance, can be used as well [26].

### 2.1.3.3 Support Vector Machine

A Support Vector Machine (SVM) is a powerful machine learning algorithm that can be used for both linear and non-linear classification tasks. The SVM algorithm finds the best hyperplane that separates the data into different classes based on their distance from similar training data points [27]. SVM maps the training data as points in space to maximize the width of the gap between the classes. New data points are then mapped into this space and classified based on which side of the gap they belong to [28]. To perform non-linear classification, SVM can use kernel functions to map the data into higher dimensional feature spaces. The radial basis function (RBF) and the sigmoid kernels are commonly used kernel functions in practice due to their effectiveness. In this thesis, both linear and non-linear approaches are proposed, including the use of the sigmoid kernel for non-linear classification [29].

### 2.1.3.4 Neural Networks and Multilayer Perceptron

The concept of mimicking the functionality of a biological brain creating Neural Network (NN) was firstly introduced in 1943 by McCullough and Pitts [30]. NN, also known as Artificial Neural Networks (ANN), are based on a collection of nodes linked together, the so-called artificial neurons, inspired by the model of neurons in a biological brain. Each connection, like the synapses, can transmit signals to other connected neurons, which then receive the signals and process them. The signal is a real number, and inside a single neuron, it is computed by some non-linear function of the sum of its inputs, creating a specific output. Neurons and links have weights that are adjusted during the learning process. The strength of the signal at each connection increases or decreases the node weight. Neurons may be artificially generated with thresholds, such that the signal is only delivered if the aggregate signal crosses the specified threshold. Typically, neurons are grouped in different layers, which perform different processing of their inputs (Fig. 2.2). Signals move from the first input layer to the output layer, traversing the layers in between multiple times [31]. ANN are the basic unit for Deep Learning (DL) (see paragraph 2.1.8), a ML technique applied in various healthcare fields like bioinformatics or medical image analysis [32]. While a deep understanding

of ANN is beyond the scope of this thesis, their mention is necessary given their extreme popularity in various biomedical applications, as the reader will see in the next chapters.

One of the most representative examples of simplified NN is the Multilayer Perceptron (MLP), which consists of a multiple layers NN (at least three): an input layer, one or more hidden layers, and an output layer [33]. All the nodes, excluding the inputs, are neurons that use a nonlinear activation function, and the training is achieved using the backpropagation of errors technique or some of its variants [34].

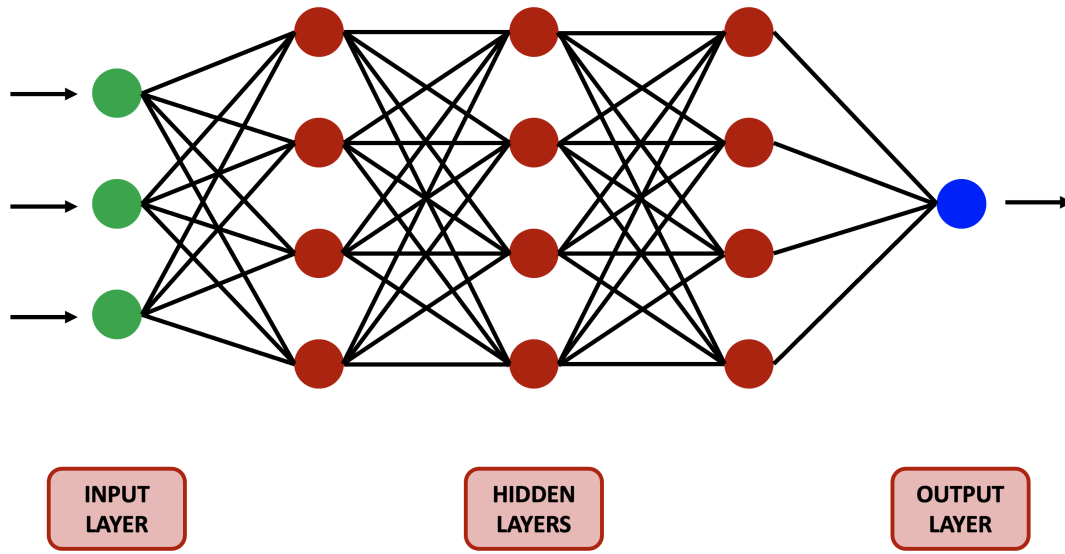


Figure 2.2: Artificial Neural Network Representation with the respective layers, nodes and links

### 2.1.4 Evaluation Metrics

Evaluation metrics are used to measure and quantify the quality of performance of an ML model. These metrics differ between regression and classification problems.

**Regression** To evaluate the performance of a regression model, several reliable metrics can be considered, including the Coefficient of Determination ( $R^2$ ), mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE).

$R^2$  provides a measure of how well future samples are likely to be predicted by the regression model. The best possible score is 1, while the final result can also be negative (because the regression model can be arbitrarily worse). Therefore, if  $R^2$  is unity, all variation has been explained, and there is a perfect fit. If the coefficient is zero, the regression does not explain anything, and the prediction is poor [35].

MAE refers to the magnitude of difference between the prediction of an observation and its true value [36]. MSE is the most common loss function; it is calculated as the difference between the model's predictions and the ground truth, squared, and averaged across the whole dataset. RMSE is the square root of the value obtained from the MSE function. It helps to plot the difference between the estimated and actual value of a parameter of the ML regression model [37].



**Classification** In a typical biomedical classification problem, where a disease has to be identified in a group of healthy and unhealthy subjects, the following terms are defined:

- True Positive (TP): sick people correctly classified as sick
- False Positive (FP): healthy people wrongly classified as sick
- True Negative (TN): healthy people correctly classified as healthy
- False Negative (FN): sick people wrongly classified as healthy

Several classification evaluation metrics are considered in this dissertation. Accuracy is defined as the number of correct predictions divided by the total number of predictions. Precision is the number of patients that we correctly identify as having the disease out of all the patients classified as sick. Recall (also known as sensitivity) tells us how many subjects the algorithm correctly identified as having the disease, out of all the subjects that actually have the disease. F-measure (F1) is defined as the harmonic mean between precision and recall. Specificity is the probability of identifying a healthy subject, conditioned on the individual truly being healthy [38] (Table 2.1).

Metric	Description
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$
Precision	$\frac{TP}{TP + FP} \quad (2.2)$
Recall or Sensitivity	$\frac{TP}{TP + FN} \quad (2.3)$
Specificity	$\frac{TN}{TN + FP} \quad (2.4)$
F1	$\frac{TP}{TP + \frac{FP+FN}{2}} \quad (2.5)$

Table 2.1: Classification metrics formulas based on TP, FP, TN, FN definitions

Finally, the Area Under the Curve of the Receiver Operating Characteristics (AU-CROC) can only be evaluated for binary classifications and measures the probabilistic performance of the classification. ROC is a probability curve, and AUC represents the

degree of separability. This metric can indicate how well the ML classification model is able to distinguish between the two classes. The higher the AUC, the better the model is at predicting class 0 as 0 and class 1 as 1, meaning that a higher AUC indicates better ability to distinguish between patients with and without the disease [39].

### 2.1.5 Tuning of Hyperparameters

In ML, a hyperparameter is a parameter assigned to the algorithmic model itself, whose value is used to control the learning process [40]. Hyperparameters are related to the model selection task, rather than the training process, and although they do not directly affect the model's performance, they impact the speed and quality of the learning process. Examples of model hyperparameters include the topology and size of a neural network, the number or maximum allowable depth of decision trees in a random forest model, or the kernel used in support vector machines.

The tuning of hyperparameters is the algorithmic process used to select the best hyperparameters to obtain the most significant evaluation metrics [41]. The tuning can be done, for example, to maximize accuracy or recall, as shown in Chapter 5.4.

### 2.1.6 Data Augmentation

Some supervised learning algorithms, such as decision trees, require a balanced class distribution to achieve more realistic and accurate classification performance. This issue is common in biomedical datasets where many classification studies are focused on rare diseases. If the minority class is significantly smaller than the majority class, some classification metrics, especially accuracy, can be misleading and not reflect the real prediction quality, which may be poor. In such cases, data augmentation or reduction techniques are employed on the training set to balance the class distribution and achieve more significant results [42].

**SMOTE** SMOTE (Synthetic Minority Over-sampling Technique) is one of the most popular data augmentation algorithms introduced by Chawla et al. in 2002 [43]. It generates artificial data by extrapolating between a real object of a given class and one of its nearest neighbors of the same class. Then, it chooses a point along the line between these two objects and determines new object attributes based on this randomly chosen point. The minority class is balanced by creating new artificial data. The implementation and efficiency of this technique are widely discussed in the next chapters.

### 2.1.7 Feature Selection

In case of datasets which contain a vast number of features, a reduction of this number to proceed with the ML model is often needed. This is performed to avoid overfitting, reduce training and evaluating time, and to improve the reliability of the evaluation metrics when the most useful features are selected avoiding the redundant ones. Several manual or algorithmic approaches can be computed to select the most significant features [44]. A brief description of the most common algorithmic methods is here presented.

**Filter Methods** In filter methods (see Fig. 2.3), features are selected based on their performances in various statistical tests [45]. Some of the most popular statistical filter methods are:

- **Pearson's Correlation:** This method measures and quantifies the linear dependence between two continuous variables, with its value ranging from -1 to +1.
- **Linear Discriminant Analysis:** This method finds a linear combination of features that defines or separates the classes of a categorical variable.
- **Analysis of Variance (ANOVA):** Similar to Linear Discriminant Analysis, ANOVA operates using one or more categorical independent features and one continuous dependent feature. It provides a statistical test of whether the means of several groups are equal or not. An example of the application of ANOVA feature selection can be found in Chapter 5.4.
- **Chi-Square:** This method is applied to categorical features to estimate their correlation tendency by evaluating their frequency distribution.

**Wrapper Methods** Wrapper methods (see Fig. 2.3) are computationally more expensive than filter methods, as they test a subset of features in a loop and train them with a selected algorithm, adding or removing variables from the tested subset [46]. Some common examples of wrapper methods include:

- **Forward Selection:** This method begins with no features in the model, and at each iteration, the most significant features are added to the model until a further variable addition does not improve the performance anymore.
- **Backward Elimination:** In contrast to forward selection, this method starts with all the features present, and the least significant feature is removed at each iteration until a variable elimination does not improve the performance anymore.
- **Recursive Feature Elimination:** This method creates models in a loop and keeps track of the most and least performing feature at each iteration. It builds the next model with the selected features, ranking them based on elimination order.

**Embedded Methods** Finally embedded methods merge the qualities of filter and wrapper methods. LASSO and RIDGE regression are the two most popular algorithmic embedded approaches [47].

**Principal Component Analysis** Principal Component Analysis (PCA) is a widely used technique for analyzing datasets with a high number of features. It was first theorized by Pearson at the beginning of the 20th century [48]. PCA preserves the maximum amount of information from the data by transforming it into a new coordinate system using an orthogonal linear transformation. This creates a new set of  $N$  PCA features. Typically, the number of PCA features  $N$  is selected to explain 95-99% of the variance and its value depends on the number of initial features and their correlation [49]. An application of PCA can be found in Chapter 5.3.

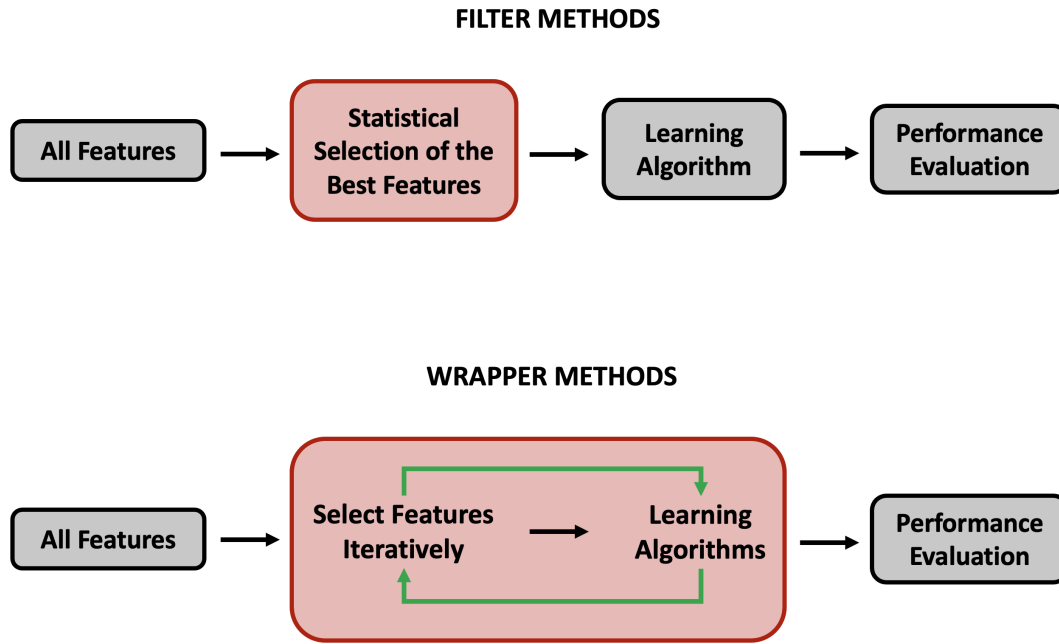


Figure 2.3: Filter and Wrapper Feature Selection Diagrams

### 2.1.8 Deep Learning

Deep Learning (DL) is a subclass of Machine Learning (ML) that uses multiple layers to extract progressively higher-level features from an input [50]. Based on the Artificial Neural Networks (ANN) shown in Fig. 2.2, DL can be performed using supervised and unsupervised learning techniques. DL has shown great potential in several domains, including computer vision, speech recognition, natural language processing, drug design, bioinformatics, genomic dataset management, and medical image analysis [51]–[55].

While DL has become a powerful tool in various fields, this thesis mainly focuses on the use of ML for medical images and large biomedical datasets. Advanced techniques of ANN technologies in DL are beyond the scope of this thesis. However, the different applications of ML in healthcare are discussed in various contexts.

### 2.1.9 Applications in Healthcare

#### 2.1.9.1 Large Datasets

The investigation of large datasets, commonly referred to as "big data analysis," covers a vast variety of phenomena, including biomedical big data, which has recently gained more attention due to its potential for diagnosis, prevention, and rehabilitation [56]. Biomedical big data of academic and scientific relevance can be found in many different structures, such as aggregated clinical trials [57], bio-genetic and microbiomic sequencing data [58][59], or digital health records in clinics and hospitals [60]. Big data are also generated from wearable devices and smartphone health applications or via social media with the use of increasingly popular "personal health monitoring" technologies [61]. Such data are mainly stored in virtual research repositories or biobanks, which can be proprietary or open access [62].

The size of bio-genetic datasets can be in the order of millions or even billions of features for thousands of patients [63]. For example, the Human Genome Diversity

Panel, which includes 940 individuals, has a total of 604 million observed genotypes [64]. The 1000 Genomes Project consists of 1,718 individuals for a total of 3.2 billion observed genotypes [65], and the Human Origin dataset has 1,941 individuals and 748 million observed genotypes [66]. Another influential example of a large genetic dataset existing in the literature is the US Million Veteran Program, which includes genomic data from around 300,000 multi-ethnic participants [67].

The analysis of DNA sequence genomic datasets is generally performed using sophisticated ANN with DL techniques [68] or with the use of specific advanced and optimized ML algorithm techniques capable of managing millions of inputs in a reasonable time-frame [69]. For example, Alipanahi et al. [70] developed convolutional neural networks (CNN) (a class of ANN defined as advanced regularized versions of MLP) that can identify and visualize damaging genetic and deleterious genomic variants by predicting the sequence specificities of DNA- and RNA-binding proteins from thousands of sequences of various lengths obtained from experimental data. Listgarten et al. [71] used ML algorithms to successfully identify a subset of three single nucleotide polymorphisms distributed over 45 genes of potential relevance to breast cancer etiology as key discriminators between breast cancer and control subjects, achieving a 69% accuracy with a quadratic kernel SVM model.

The study of large datasets is not one of the goals of this thesis. However, we investigated the AGES-Reykjavik dataset [72], introduced in Chapter 3.1.1, which consists of a fairly high number of elderly individuals (>3000), but the number of features considered is considerably less compared to the aforementioned big data repositories.

### 2.1.9.2 Medical Imaging

Medical imaging is undoubtedly one of the most efficient diagnostic techniques available today [73]. A summary of the main medical imaging techniques with brief descriptions of their key characteristics is provided in the table in Fig. 2.4.

In recent decades, the development of ML and DL has consistently improved the potential of medical imaging [74][75]. As a result, there are now several hundred publications on this subject [76]. Advanced DL applications can efficiently segment human anatomy and predict diseases using various imaging techniques, such as MRI [77], CT-scans [78], PET [79], and ultrasound [80], and can also generate 3D images [81], useful for identifying patterns in rare conditions [82]. DL has recently been recognized as an efficient and popular tool for identifying COVID-19 infections using chest imaging of the lungs and bronchi [83][84].

One example of DL imaging application is presented by Roth et al. [85], who proposed a method based on CNN for organ- or body part-specific anatomical classification. Specifically, they trained their deep neural network using 4,298 axial 2D CT images from 1,675 patients to classify five human body parts: the neck, lungs, liver, pelvis, and legs. Their experiments achieved an anatomy-specific classification error of 5.9% and an average AUCROC value of 0.998. Zhang et al. [86] designed different CNN architectures to segment infant brain tissues based on multimodal MRI images, demonstrating that a DL system can outperform manual segmentation performed by professionals.

DL is primarily applied to images, specifically to the matrices of pixels within those images. On the other hand, ML techniques, more so than DL techniques, have a greater impact on the study of radiologic features extracted from medical images themselves [87]. ML techniques have shown high performance in oncology studies, as seen in Parmar et al. [88]. Macyszyn et al. [89] used SVM to predict survival and molecular

subtype in glioblastoma, achieving a 75% accuracy for the four tumor subtypes classification and an 88.57% accuracy for the 18-month survival SVM binary classification model. Similarly, Zhang et al. [90] used multiple feature selection methods and classification algorithms to successfully predict local and distant failure in nasopharyngeal carcinoma. Their study utilized 970 radiomics features from 110 patients, with RF and ADA-B algorithms producing the best classification results with respective AUCROC values of 0.84 and 0.82.

In the present study, ML analysis is utilized to examine the effect of sarcopenia on elderly individuals and to assess knee cartilage degeneration using features extracted from different medical images of the mid-thigh and knee, as described in Chapters 3.1 and 4, respectively.

### 2.1.9.3 Biosignals

Biosignals are defined as any signal produced by a living organism that can be measured and quantified scientifically. In human healthcare studies, biosignals refer to electrical signals, which are changes in electric potential across specific human tissues, organs, or systems, such as muscles or brain activity [92]. Among the most popular signals used in biomedical engineering are Electromyography (EMG), Electrocardiography (ECG), and Electroencephalography (EEG). Advanced feature extraction techniques enable numerous scientifically significant applications of ML technologies with these signals.

**EMG** Surface EMG involves the placement of wired or wireless sensors on the skin to measure the electric potential generated by muscle cells when they are electrically or neurologically activated after an induced stimulus [93]. Features extracted from EMG signals are mainly used to investigate possible abnormalities and disorders in postural and motor control, or to diagnose neuromuscular diseases [94]. Abnormal patterns in EMG signals have been identified, for example, in patients affected by Parkinson's disease [95] or amyotrophic lateral sclerosis [96]. Both ML and DL techniques are widely used with EMG signals, and they find applications, for example, in improving the performance of prosthetic limbs [97] or in studying sport performances and quantifying muscle fatigue [98][99]. Another example of its application is presented by Yu et al. [100], who used both DL and ML techniques (in particular, RF) to distinguish between 287 subjects with and without a history of strokes using EMG features extracted from thighs and calves, achieving an accuracy of over 90%.

**ECG** An electrocardiogram (ECG) is a simple and affordable test that is primarily used to check the heart's rhythm and electrical activity. A few sensors are attached to the skin to detect the electrical signals produced by the heart at every beat. ECG is used to investigate symptoms of potential heart problems and can detect arrhythmia, coronary heart disease, heart attacks, and cardiomyopathy [101]. Nowadays, AI and ML technologies facilitate the diagnosis and prediction of different heart diseases or detection of abnormalities in the heartbeat with ECG signals [102]. Arrhythmia can be detected with high accuracy using ECG features with both single and ensemble SVM [103], and also with an ANN with 7 hidden layers [104]. Khosla et al. [105] demonstrated that "any ECG abnormality" is a highly ranked feature in the longitudinal prediction of strokes, indicating that all ECG abnormalities could be more indicative of stroke than just the common atrial fibrillation diagnosis. Similar results

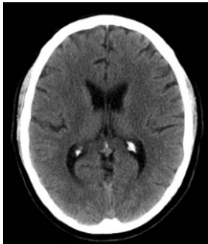
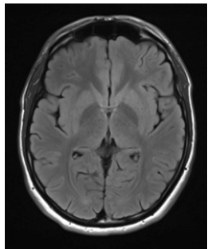
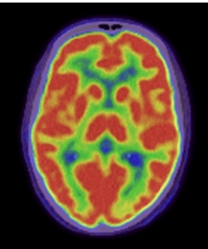

NAME	DESCRIPTION
<p data-bbox="359 327 483 353">CT-SCAN</p> 	<p data-bbox="592 405 1390 562">It combines a series of X-ray images from different angles around the patient's body and with the help of advanced computer processing creates cross-sectional images (slices) showing blood vessels, soft tissues like muscles and fat, and bones. CT-scans can provide more-detailed information than basic X-rays images.</p>
<p data-bbox="392 701 450 728">MRI</p> 	<p data-bbox="592 763 1390 954">It is a noninvasive medical imaging test that uses radio waves and a large magnet to produce detailed images of the internal structure of the human body, including organs, bones, muscles, and blood vessels. No radiation is produced during an MRI exam, unlike X-rays. MRI is better than CT at telling the difference between types of soft tissues and between normal and abnormal soft tissues.</p>
<p data-bbox="392 1081 450 1108">PET</p> 	<p data-bbox="592 1178 1374 1301">It is a functional imaging technique that uses radioactive substances to measure and show changes in physiological activities. Different kind of tracers can be used for the different purposes of the imaging analysis, depending on the target process within the body.</p>
<p data-bbox="355 1462 486 1489">Ultrasound</p> 	<p data-bbox="592 1529 1374 1720">It is a non-invasive diagnostic technique used to image inside the body. The ultrasound probes, usually placed on the skin, produce sound waves with a precise frequency. Anatomical ultrasound produces images of internal organs or other structures. Functional ultrasound combines information relative to the movement or speed of tissue or blood or the density of tissue with anatomical images.</p>

Figure 2.4: Short description of the major diagnostic medical images with brain images examples [91]

were demonstrated by Rathakrishnan et al. [106], showing that ECG features of elderly stroke and healthy subjects are of high importance in stroke identification.

**EEG** Electroencephalography (EEG) is a technique used to measure brain electrical activity by placing electrodes on the scalp of the head [107]. It provides precise temporal resolution, allowing the detection of changes in neural activity and rigorous tracking of brain dynamics and activation [108]. EEG is essential for tracking and monitoring cognitive events that occur within tens to hundreds of milliseconds [109]. Thousands of different temporal, spectral, and connectivity domains can be extracted from an EEG signal. EEG systems are relatively cheap, adaptable to mobile and various acquisition processes. However, they require appropriate electrode positioning, signal amplification, and intensive training [110]. An example of a basic EEG setup is shown in Fig. 2.5.

ML models with EEG features find multiple applications in recent scientific literature [111]. For instance, Mumtaz et al. [112] used EEG-derived synchronization likelihood features as input data for automatic diagnosis of major depressive disorder and found that the SVM algorithm was the most effective. Wang et al. [113] used RF to identify normal and abnormal amplitude integrated EEG in neonatal infants, achieving an accuracy of 91% by raising the minority class weight to handle the imbalanced dataset. Tuyisenge et al. [114] did not use ML and EEG features to classify a disease but worked on the signal itself by implementing an ensemble bagging classifier that can automatically detect bad channels in intracranial EEG signals.



Figure 2.5: Basic EEG setup: EEG cap, amplifier, software and computer, from ANT-neuro, eegoTM mylab [115]

EMG, EEG and other biosignals from postural control assessment are combined in a biosignal-dataset and examined in this thesis in the context of the Motion Sickness studies using the novel BioVRSea acquisition system (Chapter 5).



**Gait** Biosignals extracted from gait analysis acquisitions have a significant impact on biomedical engineering studies. Typically, gait analysis is performed in a specialized laboratory with several cameras, either video or infrared, placed around a walkway. Patients are fitted with multiple markers on their body and are asked to walk at a specified speed on a force platform with floor-mounted load transducers that measure the Center of Pressure (CoP) [116]. Gait analysis mainly involves temporal/spatial features, including the computation of speed, step length, pitch, and other features. ML models have demonstrated high performance in using gait parameters to distinguish neurodegenerative disorders, identifying, for example, Parkinsonian symptoms [117] or Friedreich's ataxia characteristics [118], as well as identifying various types of prostheses with SVM [119]. They have also been shown to be efficient for rehabilitation purposes [120], such as in patients affected by knee osteoarthritis [121]. ML has found numerous applications in the study of Parkinson's disease using gait features, one of which is proposed by Rehman et al. [122]. They performed a recursive feature elimination technique with SVM to classify patients with early-stage Parkinson's and healthy subjects, obtaining an accuracy between 73-97% and identifying mean step velocity, mean step length, step length variability, mean step width, and step width variability as the most significant temporal/spatial features for the model. In this thesis, we compute temporal/spatial gait features from single and cognitive dual tasks to classify individuals of Asian descent with motor or neural problems or a history of heart attacks in Chapter 3.2.

## 2.2 Clinical and Diagnostics Applications

In the field of digital health engineering, a precise and optimized system is required for the implementation of ML approaches, which can be substantially different from other types of database analyses. For instance, in the network management sector, ML can be leveraged to aid in the management of live data-traffic or long-range capacity planning. In finance, ML is widely employed for asset management, risk evaluation, credit scoring, and loan approval. A significant difference between these domains and the field of healthcare is that, in most cases, the data used in other fields refer to fictitious values that do not have a direct correlation to a tangible value, such as the measurement of muscle tissue density or the frequency of heartbeats. On the other hand, most of the biomedical features used in AI applications are derived from medical images or biosignals, which were initially developed to support physicians and medical professionals in making accurate diagnoses or planning rehabilitation or prevention strategies. In this context, AI and ML can significantly augment the impact of biomedical features and provide clinicians with insights that would not have been possible with a purely "human" approach.

As depicted in Figure 2.6, an optimal implementation of an ML prediction process in digital health engineering should begin with a precise and accurate data acquisition phase that aims to create the most complete and reliable dataset possible. Following proper data acquisition and rigorous engineering processing of feature extraction, the resulting database should be subject to deep analysis, with initial attention given to missing data (NaN) and outliers. Proper NaN management is strongly necessary to prevent loss of valuable data, and an outlier analysis can be useful to differentiate between real outliers and inaccurate preliminary feature extraction. Subsequently, a univariate or multivariate statistical analysis enables an exhaustive statistical view of

the dataset. In classification processes, frequently utilized in healthcare for identifying specific diseases or comorbidities, this analysis helps in understanding which features are statistically significant with respect to the prediction class.

Subsequently, feature selection may be conducted using one of the methodologies described earlier, depending on the requirements. In healthcare, Principal Component Analysis (PCA) may be a viable option for feature selection if feature importance analysis is not necessary post the ML model. The feature importance analysis provides valuable insights such as the most influential part of the body or brain region, or the most significant biosignal in the prediction process, which is carried out immediately after the regression or classification model. However, if PCA is used, the model's input features would be artificial and would not reflect the actual biomedical input data. Thus, the feature importance analysis will not provide any useful clinical information. Nevertheless, PCA may be useful for dimension reduction.

Successively, the optimal train-test division method is selected based on the number of subjects and features. For larger datasets, the 10-fold cross-validation method is widely considered the most effective. ML algorithms are then applied, utilizing the best strategies for achieving the desired goal. In this PhD study, tree-based approaches have been demonstrated to be the most reliable in most cases. Additionally, linear and simplified ANN approaches have also demonstrated a significant impact. Evaluation metrics are then calculated and a feature importance analysis is performed. This analysis can provide useful insights as some of the features that were considered insignificant in previous research can turn out to be highly relevant in the new prediction process. The success of the research significantly relies on the fruitful interaction between medical professionals and engineers. As emphasized at the beginning of this paragraph, the ultimate clinical evaluation goal is crucial for making a strong impact in the healthcare system and improving the daily lives of individuals with health conditions.

In the next sections, the different state of the arts relative to the clinical topics addressed in this thesis are presented, starting from the healthy aging problem and continuing with cartilage degeneration and motion sickness study.

## 2.2.1 Aging Study

### 2.2.1.1 Sarcopenia and Cardiovascular Pathophysiology

The progressive degeneration of skeletal muscle is consistently identified as an independent risk factor for significant morbidity, disability, and mortality in aging individuals [123][124]. Defined as sarcopenia, recent literature has interrogated its mediating and moderating roles in a wide range of adverse health outcomes, including its role in the etiology of cardiovascular pathophysiology [125].

Catabolic inflammatory cytokine production and characteristic adiposity from the progression of sarcopenia have been linked with the onset of diabetes [126], hypertension [127], and dyslipidemia [128], all of which are well-established risk factors for CHD [129] and all-type CVD [130]. CHF patients frequently develop cardiac cachexia [131], a similar muscle wasting condition whose advanced stage has been implicated as an accelerated analogue of sarcopenic muscle degeneration [132]. Indeed, the progression of sarcopenia in older CHF patients may be considerably entangled with embedded cachexic effects [133]. While literature cites the associations and potential causal mechanisms between cardiovascular pathophysiology and downstream changes in skeletal

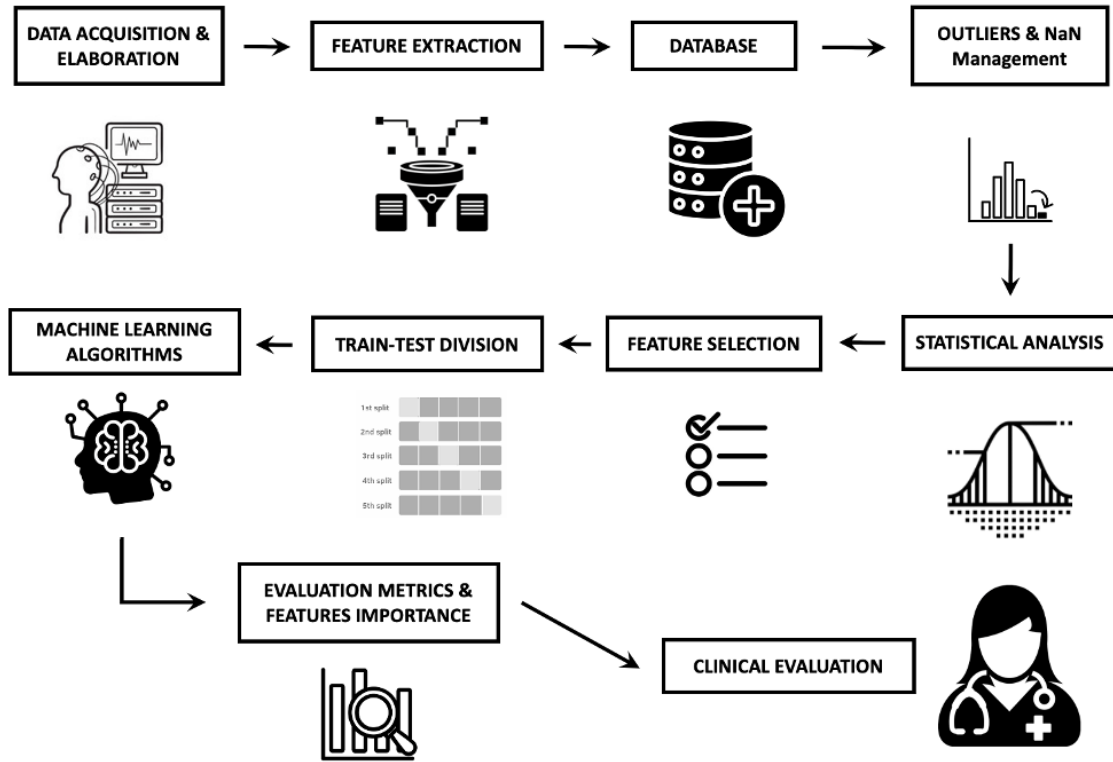


Figure 2.6: Workflow for the Machine Learning approach in Digital Health Engineering.

muscle form and function [134], validating standardized predictive models for these conditions remains debated. Furthermore, incorporating more nuanced quantitative methods for the non-invasive prediction of these events remains a priority in literature. Identifying such a methodology would further establish the generalizability of skeletal muscle research to the early detection of cardiovascular pathophysiology and facilitate the identification of compensatory targets for clinical intervention.

The concomitant loss of muscle mass and increase in adipose tissue in aging individuals suggest the use of quantitative imaging techniques, such as X-ray CT or MRI to characterize overall changes in skeletal muscle [135]. These changes altogether present a reduction in muscle ‘quality’, which has been cited as a significant causal mechanism in the loss of muscle function - particularly when in conjunction with reduced muscle mass [136][137]. CT imaging has shown particular utility in quantifying these changes [135]. This is often performed via the use of radiodensitometric absorption values, measured in Hounsfield units (HU). Here, changes in segmented cross-sectional areas have been used to illustrate changes in volume [138][139], and changes in average HU values have been used to illustrate changes in muscle quality [140]. The utility of modelling entire radiodensitometric distributions from CT cross-sections of the mid-thigh, highlighting the nonlinear trimodal regression analysis (NTRA) method has been previously shown [141][142].

The definition and the relative formula of NTRA is presented in section 3.1.1.2. The potential and efficiency of the features extracted from a mid-thigh CT-Scat using NTRA are discussed in details in the following chapter 3.1 with a focus on cardiovascular pathophysiology in chapter 3.1.2.

### 2.2.1.2 Healthy Lifestyle in Aging

Leading a healthy lifestyle not only contributes to general psychophysical well-being, but may also protect against a host of deleterious health outcomes, such as type 2 diabetes mellitus [143], hypertension [144], cancer [145], as well as cardiovascular [146], liver [147], lung [148], or kidney diseases [149]. While literature associates many lifestyle factors with these health outcomes, low physical activity and cigarette smoking have shown particularly strong causal evidence for adverse health [150] and increased mortality risk [151]. Indeed, regular exercise and abstinence from smoking during young adulthood has been consistently associated with reduced blood pressure and cardiovascular disease risk in aging [152], as further evidenced by increased life expectancy and fewer years of disability before death [153]. Low physical activity has also been identified as an independent risk factor for incident diabetes [154], for which the concomitant effects of cigarette smoking have likewise been well-established [155]. For example, in a large cohort of middle age Finnish men and women, cigarette smoking was indicated as the main risk factor for diabetes, even when controlling for physical activity, coffee consumption, and alcohol use [156]. Furthermore, the recent work by Reis et al. [143] illustrated that elderly women with a low-risk lifestyle profile, defined as non-smokers who engage in regular physical activity and moderate alcohol consumption, had a dramatically lower diabetes incidence risk than subjects with unhealthy lifestyle profiles.

While the many deleterious consequences of smoking have been well-documented [150] understanding the protective role of physical activity against aging pathophysiology is comparatively limited by its complex causal mechanisms and divergent measurement constructs. For example, while physical activity may be objectively measured using established methods such as accelerometry or oxygen consumption, these tools may be unsuitable for individuals of advanced age. For example, oxygen consumption criteria for measuring fitness has been shown to significantly underestimate cardiorespiratory fitness in aging individuals [157]. As such, self-reported assessment of physical activity is often performed, despite risking bias from faulty recall or social desirability [158][159]. For these reasons, systematic review literature consistently cites substantial heterogeneity in physical activity reporting units, which ultimately limits the generalizability of available evidence. Understanding the protective role of exercise against aging pathophysiology may therefore necessitate a deeper look into the direct and indirect effects on health from changes in skeletal muscle.

The effect of the NTRA features on the lifestyle of elderly subjects and on the diabetes and hypertension is discussed in chapter 3.1.3.

### 2.2.1.3 Single and Dual Gait Analysis

Walking is one of the most common human physical activities, and its investigation, through gait analysis, can provide multiple and significant information about physical and neurological status. Gait analysis involves measurements of temporal and spatial features of the subject's walking on a pressure mat without any other action requested: This is known as single-task gait [160]. The dual-task gait consists of adding a motor or cognitive task while walking, such as bouncing a basketball or counting a 3-digit number backward [161]. The dual-task gait has been demonstrated to have a high impact in the performance evaluation of the healthy and unhealthy population [162][163]. Walking has received growing attention in studies on stroke and neurological and mo-

tor disorders among older adults [164]. Significant abnormalities in gait are found in patients of Parkinson's disease, Diplegia, Hemiplegia and Huntington Chorea [165]. Abnormal gait patterns in temporal and spatial features are also evident in patients with any kind of stroke [166].

Often, gait acquisition on multiple patients is performed to study specific diseases or degenerations in population of interest and one of the most common approaches for analyzing these medical datasets is the use of artificial intelligence technologies. These technologies nowadays offer multiple techniques to interpret biomedical datasets using ML algorithmic approaches [167]. The tree-based algorithms were demonstrated as really effective in classifying Parkinson disease using gait features as input [168]. Other ML techniques like SVM, PCA or KNN were also demonstrated to be efficient starting from single gait features because they can predict mild cognitive impairments [169], hemiplegia and back and leg pain [170] or even distinguish between age groups [171].

The application of ML analysis to study elderly Korean subjects and their neurological and motor status, using single and dual gait analysis, is discussed in chapter 3.2.

### 2.2.2 Knee Osteoarthritis Study

Osteoarthritis is a prevalent form of arthritis, happening when protective hyaline cartilage between bones breaks down through injury or disease [172]. Osteoarthritis of the knee is one of the main reason for impairment, being a significant burden on healthcare systems [173] with a greater risk to develop with obesity and aging [174][175].

The study of cartilage thickness is essential to both identify and control the evolution of osteoarthritis [176]. Diagnosis relies on a clinical assessment and a radiographic exam of the joint [177]. Magnetic resonance imaging (MRI) is the most advanced imaging technique for the evaluation of hyaline cartilage, and presented many improvements in acquisition and image modality in the past years [178][179][180]. MRI gives a visual evaluation of the cartilage. Extensive examination of MRI sequences for assessing morphological and structural aspects of knee cartilage are reported in previous studies [177][178][181]. MRI is able of precisely measuring the thickness of articular cartilage. MRI also visualize, other tissues involved in osteoarthritis, such as subchondral bone, meniscus, and soft tissue. It is crucial to understand that osteoarthritis is a disease of the whole organ, involving multiple joint tissues. Computed tomography (CT) imaging also presents a great 3D representation of cortical bone, osteophytes, and soft tissue calcification. It has been used to study changes in the joint, such as trabecular bone changes, subchondral cysts, and bone sclerosis, that can be osteoarthritis-related alterations in the joint [182].

In the scientific literature, there are several applications of ML and DL for MRI of the knee [183][184]. Liu et al.[185] used DL to detect cartilage degeneration and acute cartilage injuries within the knee joint while Bien et al.[186] developed an efficient DL method to detect general abnormalities and specific diagnosis on MRI exams. DL was also applied for osteoarthritis diagnosis [187] and to predict patients at high risk of total knee replacement to prevent the surgery with an early-stage diagnosis using both MRI and non-image features [188]. Kwon et al. [121] used gait data and radiographic images to multi-classify the severity of osteoarthritis based on the Kellgren–Lawrence grade system using DL. Different ML algorithms using MRI as an input are also applied to predict the progression of osteoarthritis using a principal component analysis (PCA) approach on the extracted features [189] or using plain radiographs and clinical

data [190].

A detailed description of novel features extracted from 2D and 3D analysis and used to assess knee cartilage status, is shown in chapter 4.1

### 2.2.3 Motion Sickness and Postural Control Study

This section will introduce the basics of MS and its relative relation with concussion and lifestyle, giving the reader a comprehensive overview of the subject. The results relative to the MS and PC studies are presented in chapter 5.

#### 2.2.3.1 Motion Sickness: pathogenesis theories and ML applications

Postural control (PC) is a central nervous system (CNS) feedback control system that governs human upright stance and gives a platform for locomotion and task-driven behavior, as well as several autonomic responses. The PC system works on a subconscious level and is based on continuous CNS input from visual, vestibular, proprioceptive, and somatosensory receptors [191]. The CNS then processes this information to direct (efferent signals) both somatic (muscular) and autonomic (blood pressure etc.) responses. The PC system can be disturbed in two ways: the first one is a disease disruption (lost function) at all levels, and the second is a physiological “overstimulation” (increased function), which gives rise to motion sickness (MS). Motion sickness is experienced by those who passively travel and is more common in women and at a young age. Although there are great individual differences, sex and age are both predictors of MS and motion sickness susceptibility (MSS) in general populations, probably due to gene–environment interaction [192]. In addition, MS and MSS also fluctuate across age, i.e., in general, humans from 2 years of age begin to feel motion sick during traveling, peaking at 13 years of age and declining postpubertal [193][194].

One of the best-known manifestations of MS is seasickness [195]. Due to modern technology, humans have faced new MS situations such as spaceflights (space sickness [196]) or when playing computer games, including the phenomenon of “cybersickness” in virtual reality (VR) environments [197]. MS is a polysymptomatic disorder, where the primary symptoms are nausea and vomiting, but sweating, facial pallor, increased salivation, drowsiness, and dizziness are also frequent [192]. There is varying susceptibility among the general population, but all those with a fully or partially functional vestibular system can experience MS. Females report higher incidence in MS history (higher frequency and severity of symptoms) and are more susceptible to seasickness, simulator sickness, and visually induced MS than males of the same age [198][199][200]. Two main theories regarding the pathogenesis of MS exist. The “sensory conflict theory” (SCT) [201] states that MS is caused by conflict between visual, vestibular, and/or somatosensory inputs. In the case of passive travel, such as being a passenger in a car or on a ship, the physical motion perceived by the vestibular system does not match the expected signals from the visual system. Sensory conflict can also occur due to a purely visual stimulus, as can be experienced by people during VR simulations who may perceive a visual movement, but vestibular signals do not match this. Recent studies report possible “sensory conflict neurons” in the brainstem and cerebellum [202][203] and also brain networks that mediate nausea and vomiting [204], which appear to further support the sensory conflict theory.

The second theory of pathogenesis in MS is the “postural control theory” [205]. It states that prolonged postural instability precedes the subjective symptoms of MS, i.e., that

MS is directly brought on by an inability to control the posture during motion rather than a detection of any sensory conflict. The ability to remain bipedal/upright is crucial to human survival and MS appears to be closely linked to postural instability; studies have shown that greater postural instability or increased body sway correlates with greater MSS [206][207].

Regardless of the underlying pathophysiology, CNS adaptive signals as well as efferent signals involved in the corrective processes, preceding and during MS, are measurable via various methods. Koohestani et al. [208] give an overview of objective biosignal measures in MS research. Some studies have looked into possible correlations between MS levels and physiological biosignals such as electroencephalography (EEG), electrogastrography (EGG), electro-oculography (EOG), skin conductivity, heart rate (HR), blood pressure, body temperature, and cerebral blood oxygen demand visualized in functional magnetic resonance imaging. Relationships between the levels of MS experienced by subjects have been demonstrated in various EEG, EGG, and eye movement studies. Objective kinematic measures such as center of pressure (COP) are also documented as having relationships to MS levels in the literature [209][210] as well as spectral characteristics of spontaneous sway, which have been measured as a possible objective measurement for a predictive MS parameter [211].

Objective measures can be useful for tracking the onset of MS, as it may be possible to use biosignals to predict the likelihood of the subject experiencing MS. Therefore, a subject's MSS can be linked to measurable physiological or kinematic parameters in some cases by correlating the objective measurements with a standardized subjective MSS test/experienced MS level test. This is crucial for all further genetic studies on MS.

Motion sickness susceptibility is generally assessed by means of a questionnaire: subjective reporting of experienced levels of typical MS symptoms during biosignal measurement is a method used extensively in recent experimental studies [212][213][214]. Correlation of various biosignals and subjective reporting of MS levels is a task to which ML is contributing. EEG has been used as a technique to correlate biosignal measurements with MS levels in multiple subjects using ML for VR-related MS [215]. Ko et al. [216] used neural network ML algorithms to estimate patient's MS level based on the EEG power spectra from possible stimulated brain areas. Li et al. [217] also studied EEG, COP, and head and waist motion markers correlated to a subjective MS questionnaire using ML following visually induced MS. Wang et al. [218] used postural difference measures pre- and post-visually induced MS calculated with a deep long short term memory model. These studies used visually induced MS exclusively for estimating physiological response in virtual environments. Hell and Argyriou [219] also used ML to predict MS using a VR rollercoaster simulation tool and a neural network architecture predicting MS and the intensity of roller coasters in order to improve the gaming experience.

### 2.2.3.2 Sport Concussion and its Relation with Postural Control

A concussion, or mild traumatic brain injury (mTBI), is a short-lived functional neurological impairment caused by a blow to the head or by a force transmitted to the head [220]. Participation in sports is a risk factor for sustaining multiple concussions [221], with some sports presenting a greater risk than others [222]. The possible acute symptoms of concussion can include headaches, emotionality, loss of consciousness, am-

nesia, problems with balance, and sleep/wake disturbance [223]. Although most cases of concussions resolve spontaneously, they can have persistent psychological, physical, and cognitive complications and protracted recovery times [223][224].

There is currently no objective way to diagnose a concussion, nor is there a precise, universal concussion definition [225]. When diagnosing concussions, medical professionals rely on clinical assessment, which can be problematic as symptoms are non-specific and could relate to other illnesses, mental or physical. In research, concussion assessments are usually carried out through questionnaires built around the most recent consensus on the definition of concussion. The Concussion Assessment Tool, fifth edition (SCAT5), has been useful when assessing symptoms after an incident and tracking recovery [220]. Although it is easily administered, it should not be used on its own for diagnosis, not unlike other similar concussion assessment scales [226][227]. Medical records and clinical interviews are sometimes thought of as the gold standard in concussion research [228], inevitably resulting in a loss of accuracy as many who suffer from a concussion do not seek medical assistance [229].

Due to the complex etiology of concussive symptoms, a multi-faceted approach to concussion assessment and treatment is essential [230], including multiple concussion measures and techniques. As a part of a multi-faceted approach, virtual reality (VR) offers a novel way to evaluate and manipulate postural control and cortical activity [231]. VR gives an option to assess responses in a secure setting while exposing participants to a visually and physically demanding task.

ML is widely used in concussion research, and its efficacy has been positively demonstrated in recent scientific literature. Both supervised and unsupervised ML models can be used to study concussions [232][233]. An unsupervised clustering approach has been used to identify and characterize distinct concussion subtypes [234], while Visscher et al. [235] used a self-organizing map to divide subjects with and without vestibular disorders after a concussive event. The supervised approach ranges from deep learning for concussion prediction on professional athletes [236] to fuzzy trees algorithms to predict post-concussion symptom recovery [237]. Of relevance is the Castellanos et al. study from 2021 [238] with the use of a large but relatively unbalanced dataset with the linear ML approach of the SVM which allowed the prediction of sport-related concussion risk with good accuracy.

### 2.2.3.3 Lifestyle Influence on Motion Sickness

The effects of lifestyle factors on MS have been widely investigated in literature: the effects of smoke and nicotine in relation to MS symptoms in particular are studied to understand why smokers do not suffer of strong postoperative nausea and vomiting (PONV) [239][240], but the reason is currently still unknown. Golding et al. [241] were the only ones who studied the effect of smoking nicotine tobacco and its deprivation on MS, demonstrating that nicotine withdrawal promotes MS whereas deprivation protects from it: the effect of tobacco in smokers is approximately half of the effect of an anti-motion drug. This may have a link with PONV suggesting that nicotine withdrawal can help smokers to have reduced risk of these post surgical consequences. Multiple MS studies are present in literature about physical activity, focusing mainly on spinning athletic movements and concussion in different sports [242][243]. Elbin et al [244] reported that young athletes with high MS sensitivity have more vestibular symptoms and impairments while Caillet et al. [245] were the only ones who studied



the effect of physical activity on MS sensitivity. Their study showed that people who have regular physical activity are less dependent on MS visual input and use vestibular afferences.



# Chapter 3

## Aging Study

This chapter is dedicated to the aging studies. It focuses on two main researches:

- The first research is dedicated to the study of soft tissue radiodensitometric parameters extracted from a mid-thigh CT-Scan to classify elderly subjects with different comorbidities including cardiac pathophysiologicals, hypertension and diabetes. The AGES-Reykjavik dataset developed by the Icelandic Heart Association is considered for these researches described below.
- The second research focus on spatio-temporal gait features used to classify elderly patient affected by gait symptoms, neurological deficits or with an history of ischemic or hemorrhagic stroke in a Korean population.

### 3.1 Soft Tissue Radiodensitometric parameters from mid-thigh CT Scan

#### 3.1.1 AGES-Reykjavik data-set

The Age Gene/Environment Susceptibility Study (AGES) [72] recruited in Iceland approximately 3,400 elderly subjects from 66–98 years of age (mean: 77.46) to participate in a series of two sets of multimetric assessments separated by approximately five years, collectively defined as the AGES-I and AGES-II database. Same features (table 3.1) are present in both AGES-1 and AGES-II. Data are provided by Hjartavernd -Icelandic Heart Association: they cannot be made publicly available, since the informed consent signed by the participants prohibits data sharing on an individual level, as outlined by the study approval by the Icelandic National Bioethics Committee.

##### 3.1.1.1 AGES Features

AGES dataset includes in total thousands of features: a limited number of them are considered in this thesis. The main features of AGES Reykjavik include general demographic information, neurological, and lifestyle parameters measured or surveyed, and incidence of multiple different comorbidities typical of the aging population. They are described in table 3.1.

All participants in the AGES-Reykjavik database were scanned twice with a CT-Scan, the second time approximately after 5 years from the first time, with a 4-row CT

Feature	Description
Age	Age of the subject
Sex	Sex of the subject
Smoking Status	‘never smoked’ - ‘smoke regularly’ (100 cigarettes or 20 cigars in lifetime) - ‘current smoker’
Physical Activity	Both for past (youth or midlife) and present (within 12 months of the survey): ‘never’ - ‘rarely’ - ‘occasionally’ - ‘moderate’ - ‘high’
BMI	Body Mass Index - $Weight/Height^2[kg/m^2]$
ISO	Isometric Leg Strength [N] - Max strength in dominant leg
CVD	Coronary vascular event before entering AGES (including stroke)
CHD	Coronary heart disease event before entering AGES
CHF	Coronary heart failure before entering AGES
DM	Diabetes Mellitus
HTN	Hypertension

Table 3.1: Main features of the AGES Reykjavik data-set

detector system at 120-kV (Sensation; Siemens Medical Systems, Erlangen, Germany). The localized scanning region extended from the iliac crest to the knee joints; prior to transaxial imaging, correct positions were determined by measuring the maximum femoral length on an anterior-posterior localizer image, followed by the localization of the center of the femoral long axis (Fig. 3.1). After image acquisition, for each subject, a single 10 mm section was taken from mid-thigh, midway between the acetabulum of the hip joint and the knee joint. Pixels from this slice were then processed to obtain subject-specific distributions of radiodensitometric values across the range of -200 to 200 Hunsfield Unit (HU) to extract the eleven Non Linear Trimodal Regression Analysis features.

For each subject, HU distributions were derived from summing and transforming each pixel’s CT number value according to the following linear transformation expression:  $HU = CT \times 2.26625 - 190$ . Following transformation, HU values were binned into 128 bins, as typical for CT assessment protocols [246]. Resultant histograms were smoothed by a non-parametric fitting algorithm to obtain underlying empirical probability density functions (PDF) for each histogram. Each PDF was then exported for NTRA regression analyses.

### 3.1.1.2 Non Linear Trimodal Regression Analysis - NTRA

The Non Linear Trimodal Regression Analysis (NTRA) method, firstly introduced by Edmunds et al. [141] [142] considers each Hunsfield Unit (HU) distribution as a quasi-probability density function defined by three Gaussian distributions. This definition resulted from the hypothesis that cross-sectional soft tissue HU distributions are trimodal or consisting of three separate tissue types whose linear attenuation coefficients primarily occupy their own distinct HU domain. These three soft tissue types and their HU ranges are as follows: adipose or fat tissue [-200 to -10 HU], loose connective

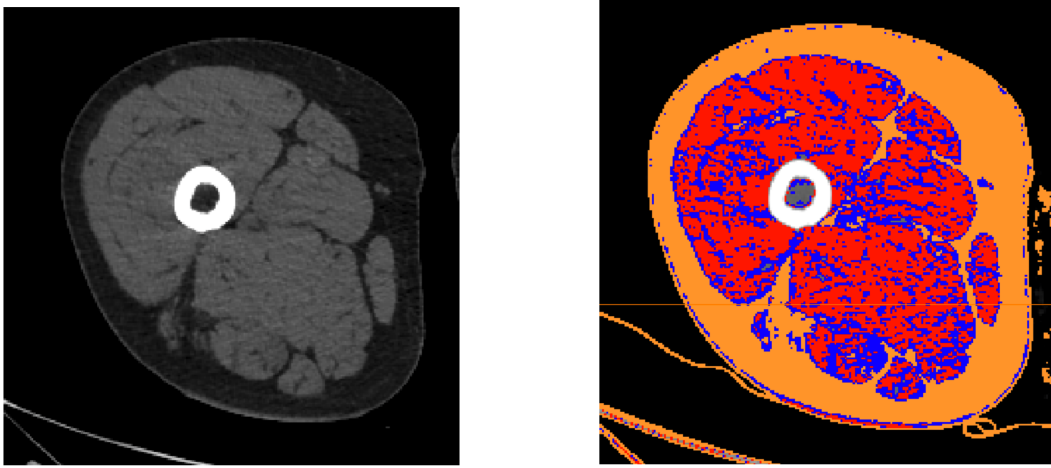


Figure 3.1: Mid-thigh CT Scan from the AGES data-set: on the right fat is orange, muscle is red and connective tissue is blue

tissue and atrophic muscle [-9 to 40 HU], and lean muscle [41 to 200 HU]. The general form of this trimodal quasi-probability density function is defined as:

$$\sum_{i=1}^3 \varphi(x, N_i, \mu_i, \sigma_i, \alpha_i) = \sum_{i=1}^3 \frac{N_i}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \operatorname{erfc}\left(\frac{\alpha_i(x-\mu_i)}{\sigma_i \sqrt{2}}\right) \quad (3.1)$$

where  $N$  is the amplitude,  $\mu$  is the location, and  $\sigma$  is the distribution width. The skewness term  $\alpha$ , captures the inwardly sloping asymmetries of fat and muscle distributions, while  $\alpha$  is considered zero (non-skewed) for the central connective tissue distribution. This definition allows for the generation of theoretical HU distribution curves via standard error minimization at each radio-absorption bin using a generalized reduced gradient algorithm, resulting in the extraction of 11 subject-specific NTRA parameters. Distributions and relative parameters of fat, muscle and connective tissues can be visualized in Fig. 3.2.

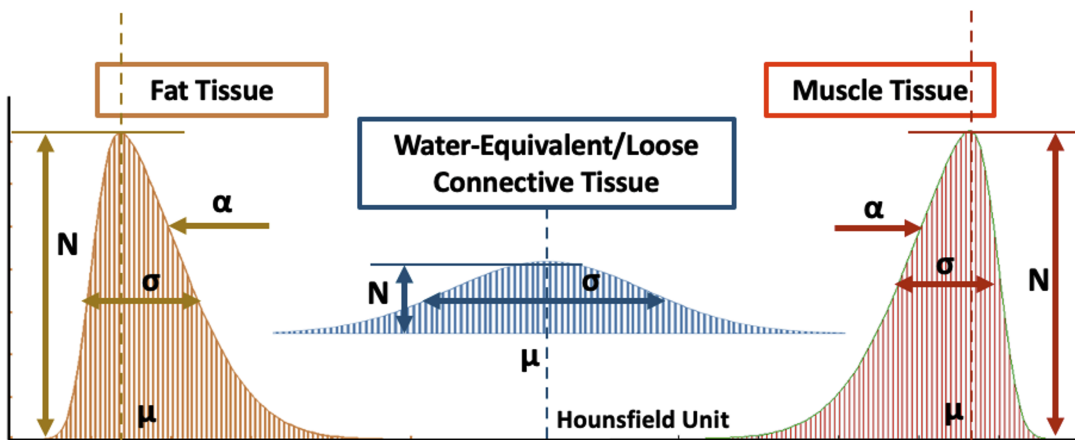


Figure 3.2: Non Linear Trimodal Regression Analysis - NTRA: visual representation of the 11 parameters

AGES I						
ALG.	Avg $R^2$	Min $R^2$	Max $R^2$	MSE	RSME	MAE
RF	70.1	62.5	82.1	4.78	2.16	1.67
EX-T	71.1	62.7	82.7	4.64	2.13	1.65
ADA-B	68.5	57.0	81.9	5.01	2.21	1.72
GB	71.4	61.8	81.5	4.55	2.12	1.65
AGES II						
ALG.	Avg $R^2$	Min $R^2$	Max $R^2$	MSE	RSME	MAE
RF	74.4	63.9	86.6	4.47	2.10	1.66
EX-T	76.0	69.9	84.8	4.22	2.04	1.62
ADA-B	74.2	67.3	86.2	4.51	2.11	1.67
GB	<b>76.6</b>	69.6	<b>88.9</b>	4.13	2.01	1.59

Table 3.2: The evaluation metrics for the BMI regression prediction

### 3.1.1.3 Body Mass Index and Isometric Leg Strength

The first methodological studies with AGES dataset aimed at proving that an ML approach can show the predictive value of the 11 NTRA parameters predicting Body Mass Index (BMI) and Isometric Leg Strength (ISO). These works were presented in Coimbra, Portugal at the *IFMBE-MEDICON* Conference in September 2019 and successively published on the dedicated conference special issue in the *Health&Technology* journal [247][248].

Regression analysis was performed with four tree-based algorithmic models: RF, EX-T, ADA-B and GB. Table 3.2 shows the evaluation metrics obtained for the regression on the BMI for each of the 4 tree-based algorithms performed. The highest  $R^2$  is 88.9 and it is obtained with GRAD-B, with an average  $R^2$  of 76.6. The most important features are the NTRA amplitudes: Nconn and Nfat in each possible combination always cover more than 50% of the total importance.

For what concerns ISO prediction the maximum  $R^2$  mean value is obtained with EX-T but the maximum value (66.5) comes from the RF algorithm. Nmusc covers almost 50% of the total feature importance while the connective tissue features, especially the  $\mu$  values, have again a significant importance in the prediction. This strengthens the results obtained with the BMI: connective tissue is highly relevant, and it can be considered as one of the main prediction factors.

The main limitation that arises in these preliminary studies is the apparent uselessness of predicting parameters such as BMI and ISO, that are easily obtainable at low cost, without an expensive CT scan approach. The power of this methodological approach has been further demonstrated (see following chapters 3.1.2 and 3.1.3). Therefore, the subsequent phase of NTRA parameters with ML is their combination to predict cardiac risks; indeed, we showed the possibility to employ them to predict cardiovascular diseases, cardiac heart failure and coronary heart disease as described in detail in the following section of this thesis (Chapter 3.1.2).

### 3.1.2 Cardiovascular Risk Assessment

This section is adapted from the first relevant publication relative to the ML studies on AGES data-set published on *Scientific Report* [1]. The main aim of this study is to highlight the substantial capacity of NTRA-based ML modelling to predict three of the main cardiovascular health outcomes (Cardiac Vascular Disease (CVD), Chronic Heart Failure (CHF), Coronary Heart Disease (CHD)). These findings are most evidenced by the high classification scores of RF models with CHF, findings which are further validated by the robust predictive performance of CHF incidence from longitudinal data. The present study altogether serves as a substantial step forward in the construction of reproducible tools for predicting cardiovascular health in elderly individuals.

In the present study, we compare the integration of the 11 NTRA parameters to classify elderly at risk for CHD, CVD, and CHF using multivariate logistic regression modelling and three different tree-based ML algorithms: random forests (RF), ADA-Boost (ADA-B), and gradient boosting (GB). Accuracy, precision, sensitivity, specificity, and AUCROC are used as evaluation metrics. Figure 3.3 depicts the study workflow. Results from each ML model were assembled over a typology of four predictive comparisons: total classification score, classification by tissue type, tissue-based feature importance, classification by age. Further model validation was compared for each ML model using longitudinal CHF data.

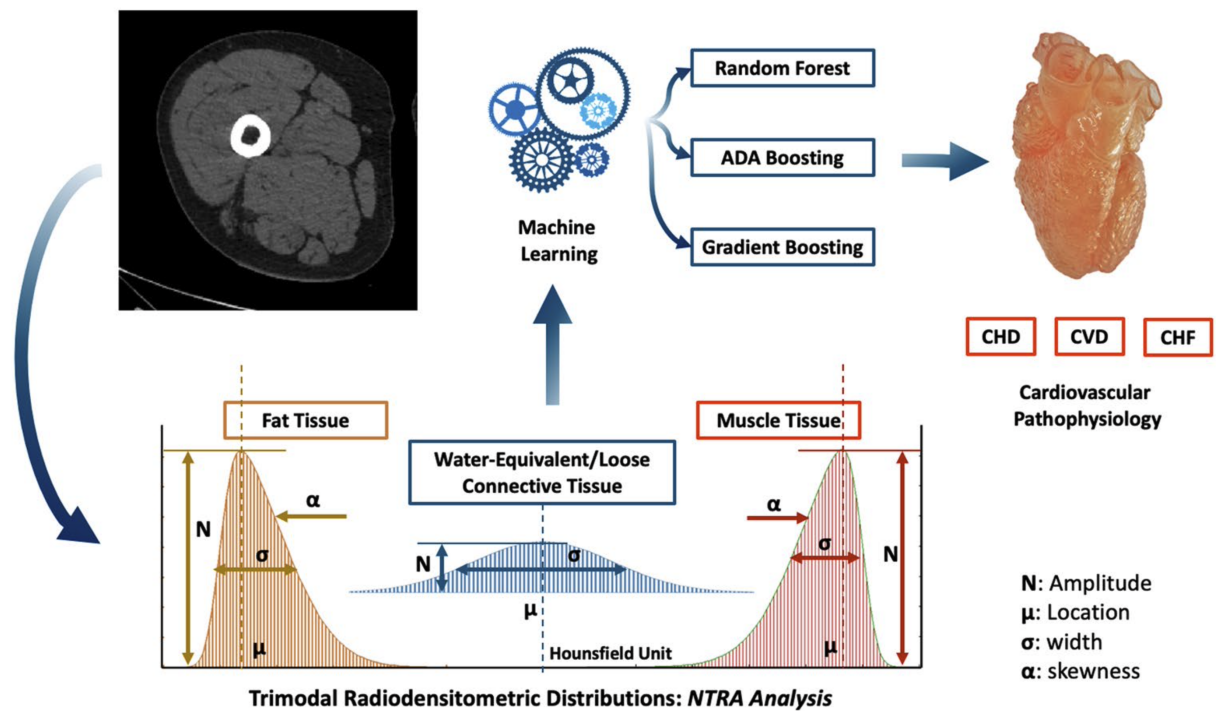


Figure 3.3: Workflow of the NTRA study on Cardiovascular risk assessment

#### 3.1.2.1 Results

**Descriptive AGES-Reykjavik statistics and NTRA parameters** Prior to the construction of the ML models, descriptive statistics and mean NTRA parameters were assembled from the AGES-I and AGES-II databases. Tables 3.3 and 3.4 contains a summary of these values. From the total sample size of  $n = 3,157$  subjects who were

AGES-I Dataset					
		<i>Cardiac Pathophysiology:</i>			
<i>Summary measure:</i>		<i>CHD</i>	<i>CVD</i>	<i>CHF</i>	<i>No Condition</i>
	Sample size (n)*	628	753	59	2394
	Age: Mean (SD)	75.5 (4.7)	75.6 (4.8)	76.6 (5.3)	74.6 (4.8)
	Sex (Male)	419	464	34	859
	Sex (Female)	209	289	25	1535
Mean NTRA (SD)					
<i>Tissue:</i>	<i>Parameter:</i>	<i>CHD</i>	<i>CVD</i>	<i>CHF</i>	<i>No Condition</i>
Fat	Amplitude: N	51.5 (28.5)	53.5 (29.1)	54.8 (25.8)	64.6 (33.9)
	Location: $\mu$	-118.0 (3.8)	-117.9 (3.7)	-116.6 (4.2)	-117.8 (3.2)
	Width: $\sigma$	9.6 (6.8)	9.2 (6.5)	8.4 (4.8)	7.9 (5.7)
	Skewness: $\alpha$	-2.8 (2.2)	-2.7 (2.2)	-2.3 (1.4)	-2.4 (2.0)
Connective	Amplitude: N	43.6 (9.0)	43.2 (9.0)	43.2 (10.4)	41.2 (8.2)
	Location: $\mu$	-14.5 (28.2)	-16.7 (28.3)	-21.0 (28.2)	-26.3 (28.3)
	Width: $\sigma$	24.3 (6.0)	24.5 (5.9)	24.9 (6.1)	25.3 (5.7)
Muscle	Amplitude: N	82.4 (18.9)	81.2 (18.6)	78.3 (19.7)	77.1 (17.6)
	Location: $\mu$	61.5 (2.7)	61.4 (2.7)	61.0 (2.9)	61.5 (2.6)
	Width: $\sigma$	8.5 (2.3)	8.6 (2.2)	8.9 (2.1)	8.6 (2.1)
	Skewness: $\alpha$	2.9 (0.9)	2.9 (0.9)	3.0 (0.7)	2.8 (0.7)

Table 3.3: AGES-I: Summary statistics and nonlinear trimodal regression analysis parameters with relative standard deviation (SD) by cardiac pathophysiology (coronary heart disease (CHD), cardiovascular disease (CVD), chronic heart failure (CHF), and no condition).

present for both studies, there were no changes in subsamples for CHD or all-type CVD ( $n = 628$  and  $753$ , respectively). However, the number of subjects with CHF increased from  $n = 59$  to  $183$  in the five years between these datasets. Mean NTRA parameters were similar between subjects presenting with cardiovascular pathophysiology, but amplitude (N), location  $\mu$ , and width  $\sigma$  parameters differed somewhat for individuals with no condition.

**ML models: Total classification scores: K-fold cross-validation with  $k = 12$**  Prior to the generation of ML models, the SMOTE technique was applied for all cardiac conditions to obtain a balanced data-set with an equal distribution of sick and healthy people. In this phase, the 11 NTRA parameters were employed to make the predictions with GB, RF and ADA-B. K-fold cross-validation was employed three times ( $k = 8, 10$ , and  $12$ ) to compute the pathophysiology predictions; here, the 12-fold cross-validation was empirically found to be the best option for predicting all three conditions. The results from  $k = 12$  analyses are summarized in Table 3.5 and the respective ROC curves are shown in Fig. 3.4.

Regarding the ML analyses, CHF was classified with the highest overall scores; specifically, the RF method yielded the best results, evidenced by an accuracy of 95.9%, an exceptionally high AUCROC of 0.994, and all additional scores above 95.0%. Nevertheless, ADA-B likewise surpassed 90.0% accuracy and obtained a high AUCROC (0.987). Concerning the CHD condition, ADA-B again obtained the second highest accuracy among all pathophysiologies, and RF was again the best algorithm (85.0% in



AGES-II Dataset					
		<i>Cardiac Pathophysiology:</i>			
<i>Summary measure:</i>		<i>CHD</i>	<i>CVD</i>	<i>CHF</i>	<i>No Condition</i>
	Sample size (n)*	628	753	183	2322
	Age: Mean (SD)	80.7 (4.7)	80.8 (4.8)	82.3 (5.3)	79.7 (4.8)
	Sex (Male)	419	464	96	831
	Sex (Female)	209	289	87	1491
Mean NTRA (SD)					
<i>Tissue:</i>	<i>Parameter:</i>	<i>CHD</i>	<i>CVD</i>	<i>CHF</i>	<i>No Condition</i>
Fat	Amplitude: N	52.7 (28.9)	54.2 (29.3)	58.8 (29.6)	64.6 (33.9)
	Location: $\mu$	-117.0 (5.3)	-116.8 (5.9)	-115.9 (4.5)	-117.3 (4.2)
	Width: $\sigma$	9.1 (6.2)	8.9 (6.0)	8.5 (5.2)	7.9 (5.6)
Connective	Skewness: $\alpha$	-2.7 (2.0)	-2.6 (2.0)	-2.7 (1.8)	-2.4 (1.9)
	Amplitude: N	43.9 (9.7)	43.4 (9.6)	42.2 (10.2)	41.4 (9.0)
	Location: $\mu$	-17.7 (27.3)	-19.2 (27.6)	-27.7 (30.7)	-27.6 (28.0)
Muscle	Width: $\sigma$	23.9 (5.5)	24.0 (5.6)	22.2 (5.3)	24.9 (5.5)
	Amplitude: N	76.3 (18.4)	75.0 (18.2)	69.8 (17.7)	72.0 (17.1)
	Location: $\mu$	60.9 (2.9)	60.7 (3.0)	59.5 (3.2)	61.0 (2.8)
	Width: $\sigma$	9.1 (2.5)	9.1 (2.5)	10.1 (3.4)	9.1 (2.6)
	Skewness: $\alpha$	2.9 (0.8)	2.9 (0.8)	3.2 (0.9)	2.9 (0.8)

Table 3.4: AGES-II: Summary statistics and nonlinear trimodal regression analysis parameters with relative standard deviation (SD) by cardiac pathophysiology (coronary heart disease (CHD), cardiovascular disease (CVD), chronic heart failure (CHF), and no condition). Note: \*From the total sample size of  $n = 3,157$  subjects that participated in both the AGES-I and AGES-II studies, 585 individuals presented with more than one cardiac pathophysiology.

	<b>Alg</b>	<b>Acc Mean</b>	<b>Acc Max</b>	<b>Sens</b>	<b>Spec</b>	<b>Prec</b>	<b>AUCROC</b>
<b>CHD</b>	<b>GB</b>	75.9	77.7	70.0	81.7	79.3	0.864
	<b>RF</b>	85.0	87.4	81.7	88.4	87.6	0.936
	<b>ADA-B</b>	79.5	82.2	74.9	84.1	82.4	0.873
<b>CVD</b>	<b>GB</b>	73.1	75.7	67.1	79.1	76.2	0.834
	<b>RF</b>	82.1	83.9	78.8	85.5	84.5	0.914
	<b>ADA-B</b>	70.2	77.0	63.3	77.2	73.5	0.766
<b>CHF</b>	<b>GB</b>	88.6	90.3	85.0	92.1	91.5	0.962
	<b>RF</b>	95.9	96.5	95.0	96.9	96.8	0.994
	<b>ADA-B</b>	94.0	95.4	92.1	95.8	95.7	0.987

Table 3.5: The 11 nonlinear trimodal regression analysis parameters were used to assess cardiovascular risks through machine learning algorithms. The evaluation metrics [%] by cardiac pathophysiology were computed.

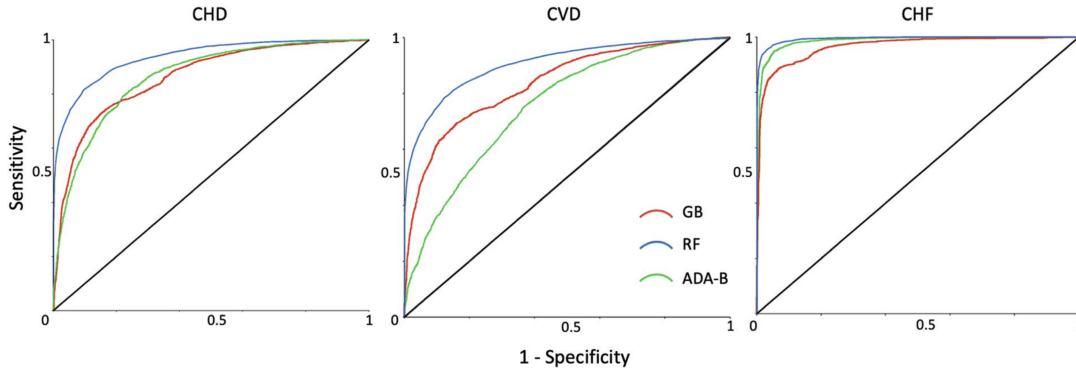


Figure 3.4: ROC curves for coronary heart disease (CHD), cardiovascular disease (CVD) and chronic heart failure (CHF) classification with K-fold cross-validation and nonlinear trimodal regression analysis by  $k = 12$

accuracy and AUCROC of 0.936). CVD was likewise accurately predicted, although the condition yielded the weakest overall results among the three, with a highest achieved predictive accuracy of 82.1% obtained from the RF method and AUCROC of 0.914.

**ML Models: NTRA-based classification by tissue type** ML analyses were further employed with features grouped by the three tissue types defined by their inherent NTRA parameters, as described:  $N$ ,  $\mu$ ,  $\sigma$ , and  $\alpha$  for fat and lean muscle, and  $N$ ,  $\mu$ , and  $\sigma$  for loose connective tissue. Table 3.6 details the evaluation metrics computed per ML algorithm in this regard, defined by each tissue type and cardiac pathophysiology.

When predicting cardiac pathophysiology from NTRA defined tissue type (Fig 3.6), the best results were again obtained from RF models; CHF was predicted with mean accuracies of 88.4%, 89.6% and 86.6% for fat, muscle, and connective tissue, respectively. Fat’s features, in general, yielded the best overall predictive value for CHF. In comparison, CHD was predicted with an accuracy of 79.6% by fat and muscle, and 78.4% by connective tissue; all tissues yielded nearly identical overall predictive results. In predicting CVD, the tissues, commensurate with the previous ML results, obtained the lowest overall scores (under 80.0%). The highest model performances, in accordance with AUCROC, were achieved with the prediction of CHF, wherein all models surpassed the value of 0.9.

**ML Models: Tissue-based feature importance** Next, feature importance was computed and grouped again by tissue type defined by NTRA parameters, allowing for the comparison of the respective contributions from fat, muscle, and connective tissue NTRA values towards the accuracy of pathophysiology prediction. These tissue contributions are detailed in Fig. 3.5, alongside an example of a segmented false-color CT cross-section that illustrates the morphology of each NTRA tissue type.

**ML Models: NTRA-based classification by age** We additionally sought to illustrate whether the excellent classification scores identified in initial ML analyses held with respect to age, indicating their relative dependencies. From the original database,

	Tissue	Alg	Acc Mean	Acc Max	Sens	Spec	Prec	AUCROC
<b>CHD</b>	<b>Fat</b>	<b>GB</b>	73.8	75.2	69.6	78.0	75.9	0.828
		<b>RF</b>	79.6	82.2	76.1	83.1	81.8	0.884
		<b>ADA-B</b>	63.9	65.0	52.0	75.8	68.3	0.674
	<b>Conn</b>	<b>GB</b>	74.3	77.5	70.0	78.6	76.6	0.824
		<b>RF</b>	78.4	80.2	74.4	82.4	80.9	0.876
		<b>ADA-B</b>	63.3	65.4	56.3	70.5	65.6	0.680
	<b>Muscle</b>	<b>GB</b>	74.0	76.4	69.0	78.9	76.6	0.824
		<b>RF</b>	79.6	82.2	76.6	82.6	81.4	0.885
		<b>ADA-B</b>	63.6	66	63.3	63.9	63.7	0.673
<b>CVD</b>	<b>Fat</b>	<b>GB</b>	71.0	73.3	66.1	75.8	73.2	0.794
		<b>RF</b>	76.8	78.1	73.8	79.8	78.5	0.855
		<b>ADA-B</b>	61.5	64.1	50.8	72.2	64.7	0.645
	<b>Conn</b>	<b>GB</b>	71.3	74.3	66.0	76.7	73.9	0.792
		<b>RF</b>	76.1	78.5	71.8	80.3	78.5	0.846
		<b>ADA-B</b>	61.6	63.4	58.3	64.8	62.4	0.654
	<b>Muscle</b>	<b>GB</b>	70.2	72.8	65.0	75.5	72.6	0.788
		<b>RF</b>	76.8	78.9	73.7	79.9	78.6	0.854
		<b>ADA-B</b>	60.7	63.8	56.8	64.6	61.6	0.644
<b>CHF</b>	<b>Fat</b>	<b>GB</b>	83.0	85.0	80.2	85.9	85.0	0.918
		<b>RF</b>	88.4	90.0	87.3	89.4	89.2	0.956
		<b>ADA-B</b>	85.6	88.7	83.3	88.0	87.4	0.927
	<b>Conn</b>	<b>GB</b>	82.4	83.8	80.2	84.6	83.9	0.907
		<b>RF</b>	86.6	87.6	86.5	86.7	86.7	0.939
		<b>ADA-B</b>	82.9	85.2	80.9	84.9	84.3	0.905
	<b>Muscle</b>	<b>GB</b>	84.0	86.5	81.4	86.6	85.8	0.922
		<b>RF</b>	89.6	91.2	89.4	89.8	89.8	0.96
		<b>ADA-B</b>	87.4	89.1	85.7	89.2	88.8	0.943

Table 3.6: The 11 nonlinear trimodal regression analysis parameters grouped by tissue type (fat, connective and muscle) were used to assess cardiovascular risks through machine learning algorithms and evaluation metrics [%] were computed.

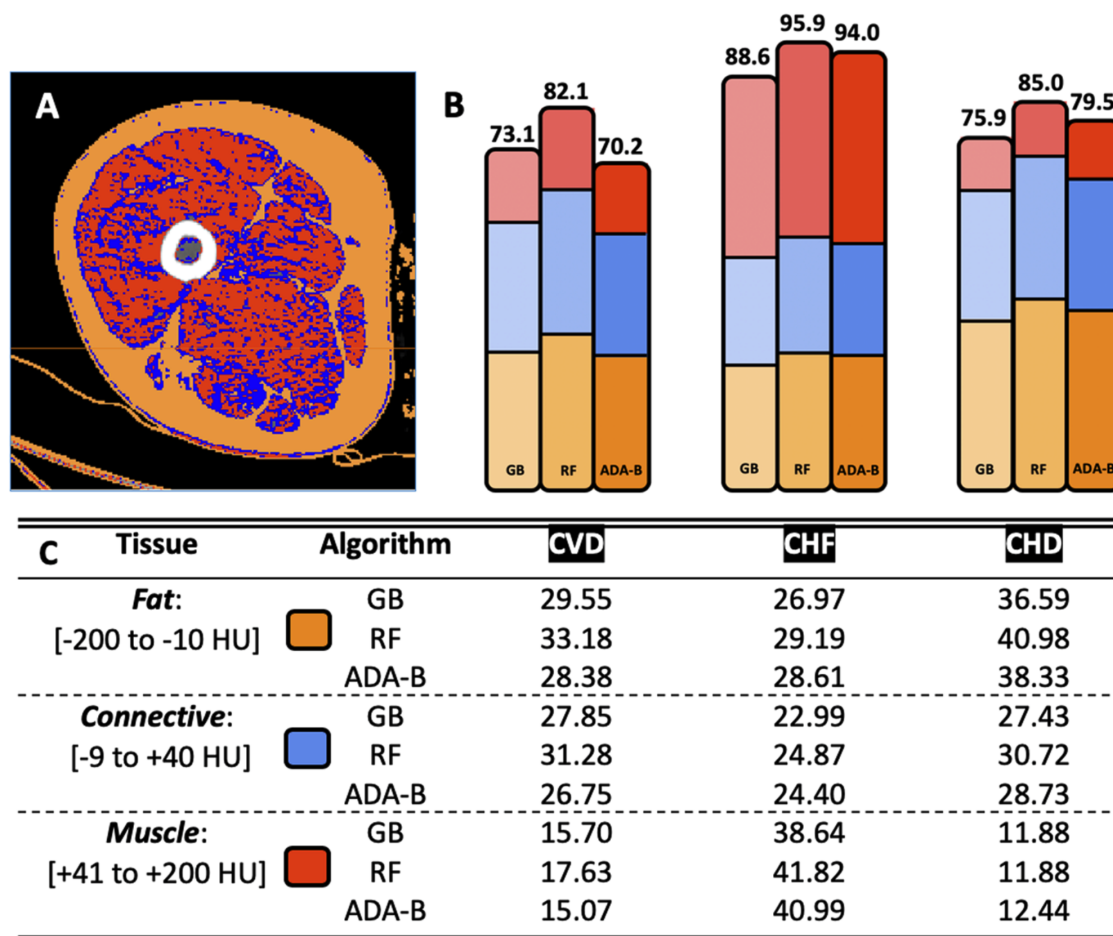


Figure 3.5: Results from tissue-based machine learning feature importance. (A) Example of a segmented false-color CT cross-section to illustrate the morphology of fat (orange), loose connective (blue), and lean muscle (red) tissue. (B) Total model accuracy (%) for each algorithm and cardiac pathophysiology, visually illustrating (with analogous colors) the compositional accuracy of each model with respect to tissue type. (C) Compositional accuracy (%) for each model with respect to tissue type.

individuals were classified into three subgroups according to their age: 66–75, 76–84, and 85–98 years old. Results from these analyses are shown in Table 3.7.

For CVD, the maximum classification accuracy and AUCROC with no age grouping were 82.1% and 0.914; splitting into three groups, RF kept on being the best algorithm and showed an accuracy between 78.0% and 85.4%, and an AUCROC between 0.875 and 0.937. Concerning CHD, the best accuracy and AUCROC were 85.0% and 0.937, respectively; subgrouping by age, RF obtained an accuracy above 82.0% for all subgroups and an AUCROC above 0.9 for each group. Finally, CHF showed again the best results with an accuracy range between 88.6% and 95.6% and AUCROC between 0.962 and 0.994 through RF. Despite subgrouping by age, results were still excellent, presenting an accuracy range of 92.6% to 97.9% and AUCROC between 0.981 and 0.998. These results confirm that ML classification is accurate, independent from age as a confounder, and considering the operation of these algorithms, it is further reasonable to assume an analogous classification independence from sex in prediction.

	Age	Alg	Acc Mean	Acc Max	Sens	Spec	Prec	AUCROC
<b>CHD</b>	<b>66-75</b>	<b>GB</b>	79.0	81.7	73.6	84.5	82.6	0.883
		<b>RF</b>	87.5	90.2	83.7	91.3	90.5	0.953
		<b>ADA-B</b>	84.3	89.3	79.5	89.2	88.0	0.920
	<b>76-83</b>	<b>GB</b>	75.9	79.7	69.5	82.3	79.7	0.862
		<b>RF</b>	85.3	87.4	82.9	87.7	87.1	0.930
		<b>ADA-B</b>	77.0	82.3	71.0	83.0	80.7	0.836
	<b>84-98</b>	<b>GB</b>	72.8	81.1	79.4	66.2	70.2	0.825
		<b>RF</b>	82.0	86.2	86.3	77.7	79.4	0.908
		<b>ADA-B</b>	71.3	81.0	81.0	61.7	61.7	0.769
<b>CVD</b>	<b>66-75</b>	<b>GB</b>	76.4	79.6	70.9	81.9	79.6	0.868
		<b>RF</b>	85.4	88.0	81.7	89.1	88.2	0.937
		<b>ADA-B</b>	78.3	82.1	73.6	83.1	81.3	0.858
	<b>76-83</b>	<b>GB</b>	72.0	76.2	66.4	77.6	74.8	0.817
		<b>RF</b>	80.6	82.4	78.5	81.9	82.7	0.890
		<b>ADA-B</b>	66.1	73.2	61.1	71.2	68.0	0.725
	<b>84-98</b>	<b>GB</b>	68.9	76.2	77.2	60.6	66.2	0.786
		<b>RF</b>	78.0	88.9	83.7	72.4	75.2	0.875
		<b>ADA-B</b>	62.9	67.6	64.3	61.5	62.5	0.659
<b>CHF</b>	<b>66-75</b>	<b>GB</b>	93.8	95.0	91.5	96.0	95.8	0.981
		<b>RF</b>	97.9	99.2	97.5	98.4	98.4	0.998
		<b>ADA-B</b>	97.5	98.7	96.2	98.8	98.8	0.997
	<b>76-83</b>	<b>GB</b>	88.1	90.5	84.6	91.7	91.0	0.963
		<b>RF</b>	96.0	97.0	94.7	97.4	97.3	0.995
		<b>ADA-B</b>	94.2	96.4	92.2	96.1	95.9	0.986
	<b>84-98</b>	<b>GB</b>	82.6	87.8	88.2	76.9	79.2	0.921
		<b>RF</b>	92.6	96.4	92.8	92.4	92.4	0.981
		<b>ADA-B</b>	89.9	93.5	92.5	87.2	87.9	0.964

Table 3.7: The 11 nonlinear trimodal regression analysis parameters were used to assess cardiovascular risks on subjects grouped by age [years] through ML algorithms and evaluation metrics [%] were computed.

	Alg	Acc Mean	Acc Max	Sens	Spec	Prec	AUCROC
CHF	GB	88.3	90.3	85.5	91.2	90.7	0.959
	RF	95.5	97.0	93.7	97.3	97.2	0.993
	ADA-B	94.3	95.8	96.2	92.3	96.0	0.986

Table 3.8: The 11 nonlinear trimodal regression analysis parameters from AGES-I were used to predict the presence of chronic heart failure in AGES-II through machine learning algorithms and evaluation metrics [%] were computed.

**ML Models: NTRA-based longitudinal assessment** In order to validate the ML prediction results, a cross sectional dataset obtained between AGES-I and AGES-II was used; here, only CHF was possible to assess due to no change in the number of individuals who received a CVD or CHD diagnosis between the two study timepoints. To test the predictive potential of our ML models against the diagnosis of CHF, an incidence index was defined; here, the null condition ‘0’ was assigned as a control to subjects without CHF in either AGES-I or AGES-II, whereas ‘1’ was assigned to those without CHF in AGES-I but with the condition in AGES-II. This method thereby removed all individuals presenting CHF at both timepoints. Table 3.8 illustrates the results from predicting CHF incidence using each of the aforementioned ML models. As shown in Table 3.8, the RF method again yielded the best predictive accuracy (95.2%) and AUCROC (0.993) for the prediction of CHF incidence. In contrast, ADA-B was analogously second-best in predictive accuracy (94.3%), and GB was the least accurate of the three (88.3%). Nonetheless, each ML algorithm surpassed an AUCROC value of 0.95, as well as specificity and precision values greater than 90.0%.

### 3.1.2.2 Discussion of the main findings

Deleterious changes in skeletal muscle in patients with poor cardiovascular health outcomes have been discussed in literature. Patients with CHF have been shown to develop significant ultrastructural abnormalities in their skeletal muscle, suggesting poor muscle oxidative capacity as reflected by decreased exercise capacity [249]. Indeed, abnormal skeletal muscle function, increased thigh intermuscular fat, and reduced exercise capacity have been cited as primary chronic symptoms in heart failure patients with preserved ejection fraction [250]. However, literature on the use of ML-modelling for the prediction of these conditions remains scarce, despite recent systematic review evidence that highlights its promising utility in datamining and classifying health outcomes [251].

At the time of this work, only one study could be found that reports using ML-modelling of CT images to classify individuals according to cardiovascular health outcomes. In this study, coronary CT angiography images were combined with ML-modelling to develop an artificial intelligence-based imaging biomarker to predict myocardial infarction in healthy subjects [252]. However, the use of CT images of skeletal muscle for classifying cardiovascular health outcomes remains unreported. Furthermore, the methodological heterogeneity between ML-based clinical studies is generally high, as predictive parameters or ML methods remain largely study-specific and unstandardized. As such, the present work aimed to explore ML-modelling techniques to classify individuals diagnosed with CHD, CVD, and CHF using CT-based NTRA parameters as a quantitative construct for skeletal muscle health.

It is likewise important to discuss the pathophysiological characteristics of the three cardiovascular outcomes utilized in this study to interrogate the particular predictive strength of CHF and relative similarity in prediction of CVD and CHD. Firstly, CVD is understood as an overarching typology of cardiovascular conditions that includes CHD alongside a host of other disease types, such as atherosclerosis or myocardial infarction [253]. As such, the comparative prediction of all-type CVD and CHD may be expected to be relatively similar. Contrastingly, CVD and CHD have been implicated as a primary etiology of CHF alongside other key comorbidities such as diabetes [254]. As such, while CHF may be a downstream consequence of CVD or CHD, its prediction likely relies on additional exogenous factors and may therefore be relatively independent.

From our ML models, there were again similarities between the classification accuracy of CVD and CHD, while CHF classification consistently outperformed the other two conditions. Nevertheless, all three conditions yielded high overall accuracies and excellent AUCROC values, suggesting the high general utility of NTRA-based modelling for all outcomes. Regarding tissue-based feature importance (Fig. 3.5), several key insights are shown, with differences apparent between cardiovascular conditions. Firstly, fat had a predominate role in classifying CHD (41.0%), while muscle had a comparatively minor contribution (11.9%). Contrastingly, lean muscle gave the highest contribution in classifying CHF (41.0%), while connective tissue yielded the lowest contribution (24.9%). Finally, fat and connective tissue gave almost the same contribution in classifying CVD (about 33.2% and 31.3%, respectively), while lean muscle was comparatively much lower (17.6%). These condition-based differences in classification indicate the potential specificity of tissue types to each condition, further suggesting the importance of segmenting classifying parameters by these three tissue types, which is one of the key features of NTRA computational modelling.

**The value of the present work** In general, this work features several key novelties for the use of skeletal muscle to classify cardiovascular health in advanced age. Firstly, we describe the NTRA computational modelling method, wherein radiodensitometric distributions from CT image cross-sections yield 11 subject-specific soft-tissue parameters that altogether present a robust and standardizable construct for quantifying muscle degeneration. This method has shown sensitivity and specificity to lower-extremity function and nutritional parameters in previous investigations [141][142], but the present use of these parameters to classify cardiovascular health outcomes is new. Finally, we validate the ML classification results using longitudinal CHF data to independently model the prediction of CHF incidence.

Altogether, a key advantage of this methodology is its derivation from CT images. As a non-invasive and standardized imaging modality that is widely utilized for diagnostic applications and pathophysiological monitoring, CT-derived HU distributions of soft-tissue radiodensity can be directly compared across clinical contexts. As such, the present use of NTRA-based classification is highly reproducible and can be readily built into existing CT analysis frameworks for patient evaluation. This tool can be further adapted into additional ML-based platforms for the detection and monitoring of adverse health outcomes, like the ones presented in the next chapter 3.1.3. Altogether, the present work serves as a substantial step forward in the construction of reproducible tools for associating skeletal muscle changes with cardiovascular health

outcomes in elderly individuals.

**Limitations** As the AGES-Reykjavik study consisted of otherwise-healthy volunteers (presenting with or without various pathologies), standard clinical measurements of key cardiac functions, such as coronary perfusion or ejection fraction measurement, were absent from the dataset. However, the validity of our results would be strengthened by the classification of these intermediate clinical measurements, as the outcomes of CVD, CHD, and CHF are largely heterogeneous in nature. The future use of our reported methods with clinical cardiovascular data would likewise allow for the interrogation of the causal relationship between cardiac health outcomes and changes in radiodensitometric NTRA values. Further testing of this relationship using independent patient cohorts may likewise be needed to further refine our ML models. Finally, while evidence for the classifying power of ML-modelling continues to grow, its literature base still lacks a standardized methodology, and the mechanisms governing some of these classifications may remain unclear. As such, exploring the contextual value of different ML-modelling algorithms remains essential.

### 3.1.3 NTRA & Lifestyle factors: Diabetes and Hypertension prediction

The present section reports the results from the second significant publication relative to the AGES studies with ML and NTRA parameters. It is adapted from the paper published on *IEEE Journal of Biomedical and Health Informatics* [2].

Smoking and low physical activity have been broadly associated with deleterious health outcomes in aging populations, but it remains unclear whether these relationships may be intercorrelated or modified by the degeneration of skeletal muscle. As such, the main objective of the present work was to interrogate the relationship between NTRA parameters and lifestyle health in the prediction of diabetes mellitus (DM) and hypertension (HTN) using multilevel Machine Learning (ML) classification. This work was performed using a large cohort of elderly subjects from the population-based AGES-Reykjavik Study database. Subjects were first classified by lifestyle health, followed by DM or HTN comorbidity diagnoses, followed by cardiovascular pathophysiology diagnoses. This classification methodology ultimately yielded a three-level binary tree. Fig.3.6 shows the workflow of this study: the 11 NTRA parameters extracted from the CT-scans and the pathologies data fill the database, which is used to create the binary tree, then ML tree-based algorithms are used to classify the subjects according to their pathology.

#### 3.1.3.1 Materials and Methods

The cohort for the present work contained 2,943 subjects, yielding a total ‘AGES-I+II’ database size of 5,886 records.

**Three-Level binary Tree: Lifestyle Health Index and Comorbidities Classification Methods** To construct the three-level binary tree for the present analyses,



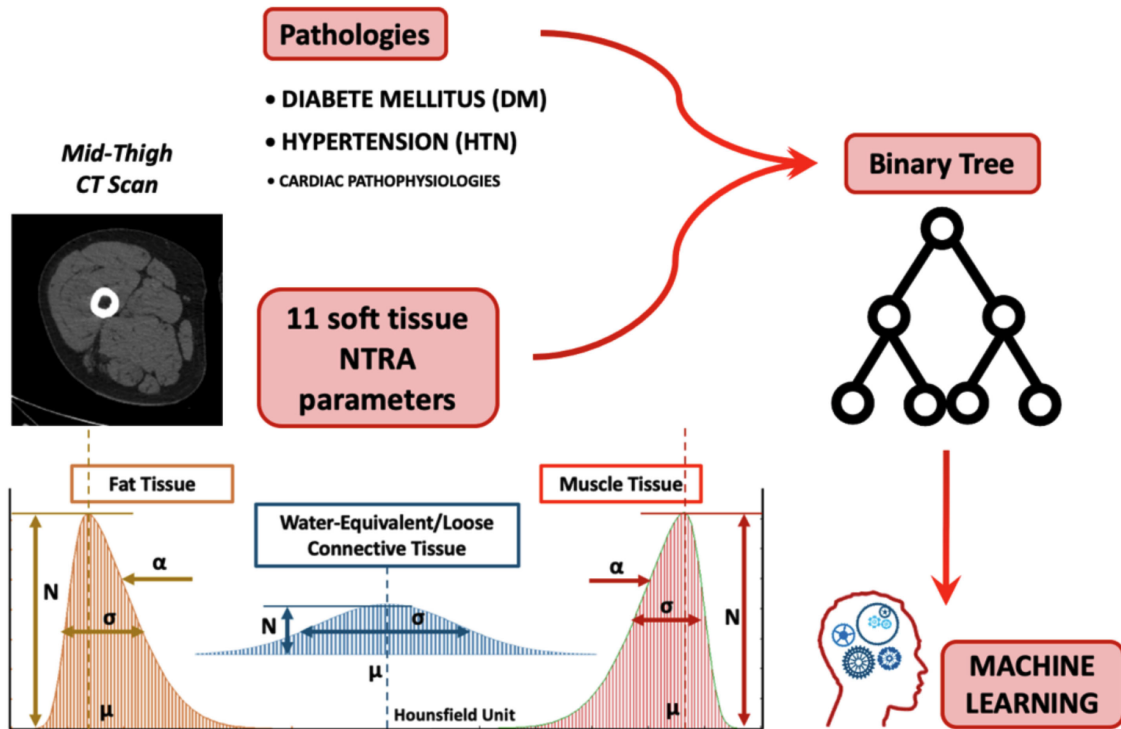


Figure 3.6: Workflow of the IEEE Journal of Biomedical and Health Informatics Study

a binary lifestyle health index ('Healthy': LSH versus 'Not Healthy': LSNH) was defined using smoking status and self-reported physical activity as key lifestyle parameters from AGES-I data.

Subjects were divided into three classes according to their smoking status: 'never smoked', 'smoke regularly' (100 cigarettes or 20 cigars in lifetime), and 'current smoker'. Next, five categories of physical activity frequency were defined for both past (youth or midlife) and present levels (within 12 months of the survey): 'never', 'rarely', 'occasionally', 'moderate' and 'high'. LSH individuals ( $n = 467$ ) were defined as those with no smoking history who reported, at a minimum, occasional past and present physical activity, while the LSNH group included all other individuals ( $n = 2,477$ ).

The second level, 'DM/HTN Comorbidities', divides patients with neither a DM or HTN diagnosis from those with at least one of the two comorbidities. The third level, 'Cardiac Pathophysiology', differentiates subjects who received a diagnosis for either CVD, CHD, or CHF from those without any cardiac pathophysiology.

Altogether, this classification methodology ultimately produced the three-level binary tree represented in Fig.3.7.

**Machine Learning and Evaluation Metrics** In the present ML classification analyses three tree-based algorithms (RF, ADA-B, and GB) were employed. The performances were evaluated after applying 10-fold cross-validation. Six evaluation metrics were considered: accuracy (Acc.), precision (Prec.), sensitivity (Sens.), specificity (Spec.), F-measure (F-Meas.), and Area Under the Curve Receiver Operating Characteristics (AUCROC). These metrics were computed both with and without SMOTE.

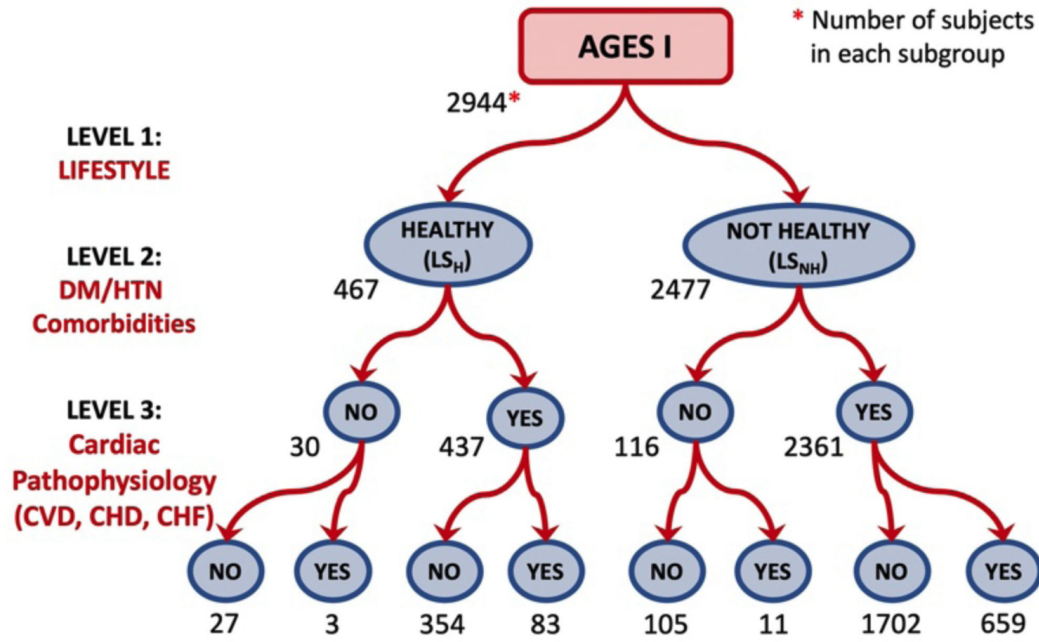


Figure 3.7: Three-Level binary tree of the IEEE Journal of Biomedical and Health Informatics Study

Tissue	Parameter	<i>LSH</i> [ $n=467$ ] Mean $\pm$ SD	<i>LSNH</i> [ $n=2477$ ] Mean $\pm$ SD	P-value
Fat	Amplitude: N	63.36 $\pm$ 33.85	61.70 $\pm$ 33.01	0.288
	Location: $\mu$	-117.88 $\pm$ 2.80	-117.82 $\pm$ 3.42	0.593
	Width: $\sigma$	8.23 $\pm$ 6.91	8.28 $\pm$ 5.82	0.000***
	Skewness: $\alpha$	-2.51 $\pm$ 2.23	-2.49 $\pm$ 2.05	0.328
Muscle	Amplitude: N	77.82 $\pm$ 16.69	78.39 $\pm$ 18.23	0.568
	Location: $\mu$	61.79 $\pm$ 2.49	61.42 $\pm$ 2.69	0.008**
	Width: $\sigma$	8.32 $\pm$ 1.96	8.69 $\pm$ 2.19	0.001***
	Skewness: $\alpha$	2.70 $\pm$ 0.74	2.86 $\pm$ 0.76	0.000***
Connective	Amplitude: N	40.47 $\pm$ 8.02	42.02 $\pm$ 8.50	0.000***
	Location: $\mu$	-26.60 $\pm$ 30.46	-23.45 $\pm$ 28.25	0.004**
	Width: $\sigma$	25.41 $\pm$ 5.99	25.04 $\pm$ 5.69	0.063

Table 3.9: The Influence Of Lifestyle Health On Ntra Parameters (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ )

### 3.1.3.2 Results

**Univariate Statistical Analyses Across Binary Tree Levels** For all three levels of the binary tree, univariate statistical analyses between node couples were performed in order to interrogate the independent roles of each NTRA parameter. All NTRA parameters underwent a Kolmogorov-Smirnov test and were not found to follow a normal distribution; as such, Mann-Whitney tests were performed for all comparisons. Pairwise statistics were first assessed at the topmost level of the binary tree (between the LSH and LSNH groups), and the results are shown in Table 3.9.

These results show that, in distinguishing between LSH and LSNH groups, differences in lean muscle NTRA parameters were preeminent; muscle  $\mu$ ,  $\sigma$ , and  $\alpha$  were all statistically significant ( $p = 0.008$ ,  $p = 0.001$ , and  $p = 0.000$ , respectively). In contrast, two connective tissue parameters were significantly different between groups, N ( $p = 0.000$ ) and  $\mu$  ( $p = 0.004$ ), along with only one fat tissue parameter,  $\sigma$  ( $p = 0.000$ ).

Table 3.10 contains the results from pairwise statistical assessment across the second level of the binary tree, which further separates LSH and LSNH individuals according to their diagnosis of DM or HTN. There were no significant differences between people affected by DM or HTN within the LSH group, but this may be primarily due to the large difference in sample sizes between the sub-groups. On the other hand, in the LSNH group, there were several significant NTRA parameter differences between individuals with either comorbidity and those without. In fat tissue, N ( $p = 0.047$ ) and  $\mu$  ( $p = 0.048$ ) parameters were significantly different, as well as lean muscle  $\sigma$  ( $p = 0.011$ ) and connective tissue N ( $p = 0.000$ ).

Finally, Table 3.11 contains the results obtained from pairwise comparison across the third level of the tree, in order to assess the potential influence of cardiac pathophysiology (CVD, CHD, or CHF). Pairwise comparisons of LSH individuals without an DM or HTN diagnosis were excluded at this level due to insufficient sample sizes for meaningful statistical comparison within these sub-groups. Results show that when distinguishing between individuals with or without a cardiac condition, the most significant NTRA differences occur in fat and connective tissues. This is particularly evidenced in LSNH individuals with diagnosed DM or HTN, where all fat and connective parameters were significantly different (with the exception of fat  $\mu$ ), compared to only one muscle parameter, N. In contrast, in LSH individuals, two fat parameters, N ( $p = 0.009$ ) and  $\sigma$  ( $p = 0.006$ ), were significantly different along with two connective tissue parameters,  $\mu$  ( $p = 0.020$ ) and  $\sigma$  ( $p = 0.000$ ).

**Machine Learning Classification Across Binary Tree Levels** Results from ML analyses using SMOTE are presented in Table 3.12. Since the present dataset was imbalanced across groups, all ML analyses were performed both with and without SMOTE. The results without SMOTE are shown in Table 3.13.

Regarding LSH and LSNH classification (first level), the RF algorithm achieved the best results across all evaluation metrics, with an accuracy of 85.4% and AUCROC of 0.935. In contrast, the ADA-B model obtained middling results (80.1% in accuracy and 0.881 in AUCROC), while the GB model yielded the lowest scores across all evaluation metrics.

Concerning the second level classification within the LSH group, the RF algorithm again resulted excellent classification scores, achieving an accuracy of 90.2% and an AUCROC of 0.978 for detecting DM or HTN. Similarly, second-level classification in the LSNH group with the RF algorithm yielded an accuracy of 94.6% and an AUCROC of 0.990. However, in both second-level classifications, the ADA-B model gave similarly high scores to the RF model in accuracy and AUCROC, even outperforming the RF model in accuracy, sensitivity, and F-measure in the LSH group. Across all comparisons, the GB algorithm again achieved the lowest scores, with the sole exception of F-measure for second-level classification within the LSNH group, where the GB model gave the highest value.

	Tissue	Parameter	DM or HTN		P-value
			No [n=30] Mean $\pm$ SD	Yes [n=437] Mean $\pm$ SD	
<i>LSHGroup</i>	Fat	Amplitude: N	55.71 $\pm$ 30.10	63.89 $\pm$ 34.06	0.192
		Location: $\mu$	-117.58 $\pm$ 2.74	-117.90 $\pm$ 2.80	0.325
		Width: $\sigma$	7.94 $\pm$ 5.87	8.25 $\pm$ 6.98	0.924
		Skewness: $\alpha$	-2.35 $\pm$ 1.09	-2.52 $\pm$ 2.29	0.406
	Muscle	Amplitude: N	74.59 $\pm$ 16.12	78.05 $\pm$ 16.73	0.148
		Location: $\mu$	62.08 $\pm$ 2.79	61.77 $\pm$ 2.47	0.802
		Width: $\sigma$	7.79 $\pm$ 1.57	8.35 $\pm$ 1.98	0.071
		Skewness: $\alpha$	2.68 $\pm$ 0.53	2.70 $\pm$ 0.75	0.940
	Connective	Amplitude: N	38.20 $\pm$ 7.93	4.63 $\pm$ 8.01	0.161
		Location: $\mu$	-24.96 $\pm$ 28.55	-26.71 $\pm$ 30.61	0.783
		Width: $\sigma$	26.74 $\pm$ 5.90	25.32 $\pm$ 5.99	0.268
	<i>LSNHGroup</i>			<b>No [n=116] Mean <math>\pm</math> SD</b>	<b>Yes [n=2361] Mean <math>\pm</math> SD</b>
Fat		Amplitude: N	57.69 $\pm$ 34.00	61.90 $\pm$ 32.96	0.047*
		Location: $\mu$	-118.23 $\pm$ 3.51	-117.81 $\pm$ 3.42	0.048*
		Width: $\sigma$	8.82 $\pm$ 6.56	8.25 $\pm$ 5.78	0.570
		Skewness: $\alpha$	-2.79 $\pm$ 2.78	-2.48 $\pm$ 2.01	0.422
Muscle		Amplitude: N	77.47 $\pm$ 17.99	78.43 $\pm$ 18.24	0.528
		Location: $\mu$	61.50 $\pm$ 2.08	61.42 $\pm$ 2.72	0.933
		Width: $\sigma$	8.17 $\pm$ 1.97	8.71 $\pm$ 2.19	0.011*
		Skewness: $\alpha$	2.72 $\pm$ 0.76	2.86 $\pm$ 0.76	0.053
Connective		Amplitude: N	39.09 $\pm$ 8.51	41.16 $\pm$ 8.47	0.000***
		Location: $\mu$	-20.74 $\pm$ 30.35	-23.58 $\pm$ 28.15	0.357
		Width: $\sigma$	25.37 $\pm$ 6.21	25.02 $\pm$ 5.66	0.228

Table 3.10: The Influence of DM And HTN on NTRA Parameters Between LSH And LSNH Individuals (\*p<0.05; \*\*p<0.01; \*\*\*p<0.001)

	Tissue	Parameter	CVD, CHD, or CHF		P-value
			No [n=354] Mean $\pm$ SD	Yes [n=83] Mean $\pm$ SD	
<i>LSH</i>	Fat	Amplitude: N	65.56 $\pm$ 34.14	56.77 $\pm$ 32.99	0.009**
		Location: $\mu$	-117.92 $\pm$ 2.72	-117.80 $\pm$ 3.16	0.776
		Width: $\sigma$	8.05 $\pm$ 7.17	9.10 $\pm$ 6.09	0.006***
		Skewness: $\alpha$	-2.49 $\pm$ 2.29	-2.67 $\pm$ 2.31	0.166
	Muscle	Amplitude: N	77.23 $\pm$ 16.38	81.55 $\pm$ 17.81	0.058
		Location: $\mu$	61.77 $\pm$ 2.44	61.80 $\pm$ 2.60	0.920
		Width: $\sigma$	8.31 $\pm$ 1.87	8.55 $\pm$ 2.38	0.446
		Skewness: $\alpha$	2.68 $\pm$ 0.71	2.81 $\pm$ 0.91	0.148
	Connective	Amplitude: N	40.39 $\pm$ 8.00	41.65 $\pm$ 8.01	0.082
		Location: $\mu$	-28.16 $\pm$ 29.17	-20.51 $\pm$ 35.66	0.020*
		Width: $\sigma$	25.83 $\pm$ 5.83	23.11 $\pm$ 6.20	0.000***
	<i>LSNH</i>			<b>No [n=1707] Mean <math>\pm</math> SD</b>	<b>Yes [n=654] Mean <math>\pm</math> SD</b>
Fat		Amplitude: N	64.95 $\pm$ 33.97	53.94 $\pm$ 28.70	0.000***
		Location: $\mu$	-117.77 $\pm$ 3.28	-117.90 $\pm$ 3.75	0.325
		Width: $\sigma$	7.94 $\pm$ 5.50	9.07 $\pm$ 6.37	0.000***
		Skewness: $\alpha$	-2.41 $\pm$ 1.97	-2.66 $\pm$ 2.08	0.001**
Muscle		Amplitude: N	77.47 $\pm$ 17.93	80.94 $\pm$ 18.82	0.000***
		Location: $\mu$	61.40 $\pm$ 2.71	61.46 $\pm$ 2.75	0.536
		Width: $\sigma$	8.75 $\pm$ 2.18	8.62 $\pm$ 2.21	0.192
		Skewness: $\alpha$	2.85 $\pm$ 0.68	2.91 $\pm$ 0.91	0.189
Connective		Amplitude: N	41.71 $\pm$ 8.21	43.34 $\pm$ 9.02	0.000***
		Location: $\mu$	-26.02 $\pm$ 28.01	-17.21 $\pm$ 27.53	0.000***
		Width: $\sigma$	25.17 $\pm$ 5.60	24.63 $\pm$ 5.80	0.036*

Table 3.11: The Influence of Cardiac Pathophysiology on NTRA Parameters Between LSH and LSNH Subjects affected by either DM or HTN (\*p<0.05; \*\*p<0.01; \*\*\*p<0.001)

<i>LSH and LSNH (first level of the tree)</i>						
Alg.	Acc.	Prec.	Sens.	Spec.	F-Meas.	AUCROC
<b>RF</b>	85.4	88.2	81.9	89.0	84.9	0.935
<b>GB</b>	73.8	75.9	69.8	77.8	72.7	0.845
<b>ADA-B</b>	80.1	82.3	76.6	83.6	79.4	0.881
<i>LSH and DM or HTN (second level of the tree)</i>						
<b>RF</b>	90.2	96.6	83.3	97.0	89.4	0.978
<b>GB</b>	85.4	89.7	79.9	90.8	84.5	0.929
<b>ADA-B</b>	90.5	94.7	85.8	95.2	90.0	0.973
<i>LSNH and DM or HTN (second level of the tree)</i>						
<b>RF</b>	94.6	96.3	92.7	96.4	94.5	0.990
<b>GB</b>	85.5	87.8	82.4	88.6	95.0	0.944
<b>ADA-B</b>	92.5	94.6	90.1	94.9	92.3	0.977

Table 3.12: Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms Applied to the Present Tree with SMOTE

<i>LSH and LSNH (first level of the tree)</i>						
<b>Alg.</b>	<b>Acc.</b>	<b>Prec.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>F-Meas.</b>	<b>AUCROC</b>
<b>RF</b>	83.9	84.1	99.6	0.04	91.2	0.593
<b>GB</b>	73.2	85.0	82.7	22.7	83.8	0.593
<b>ADA-B</b>	81.4	84.7	95.2	0.09	89.6	0.543
<i>LSH and DM or HTN (second level of the tree)</i>						
<b>RF</b>	93.6	93.6	100	0	0.967	0.448
<b>GB</b>	85.4	93.4	90.8	0.07	92.1	0.467
<b>ADA-B</b>	92.7	93.5	99.1	0	0.962	0.451
<i>LSNH and DM or HTN (second level of the tree)</i>						
<b>RF</b>	95.3	95.3	100	0	97.6	0.55
<b>GB</b>	89.9	95.5	93.8	9.5	94.6	0.537
<b>ADA-B</b>	94.7	95.3	99.3	0.1	97.3	0.535

Table 3.13: Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms Applied to the Present Tree without SMOTE

<i>Evaluation metrics for the classification of DM</i>						
<b>Alg.</b>	<b>Acc.</b>	<b>Prec.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>F-Meas.</b>	<b>AUCROC</b>
<b>RF</b>	90.0	87.7	93.0	86.9	90.3	0.967
<b>GB</b>	79.7	77.6	83.6	75.8	80.5	0.899
<b>ADA-B</b>	86.2	83.5	90.2	82.2	86.7	0.945
<i>Evaluation metrics for the classification of HTN</i>						
<b>RF</b>	86.7	85.2	78.7	93.1	81.8	N/A
<b>GB</b>	79.7	77.7	69.0	90.1	73.0	N/A
<b>ADA-B</b>	81.6	77.6	71.8	89.6	74.6	N/A

Table 3.14: Classification metrics [%] of DM and HTN, Independent from Lifestyle Index with SMOTE

**Independent Classification of DM and HTN** To compare the detection of DM and HTN independent from lifestyle index, analogous ML models were constructed and compared. Results from these analyses are shown in Table 3.14 (with SMOTE) and Table 3.15 (without SMOTE).

These scores again suggest that the RF algorithm is optimum for classifying the presence of either DM or HTN from NTRA parameters, as primarily evidenced by excellent evaluation metrics in the detection of DM (90.0% and 0.967 for accuracy and AUCROC, respectively). RF scores were likewise high for the classification of HTN; however, AUCROC scores were not available for comparison, as HTN was presented in three classes.

**Independent Classification of DM and HTN by Tissue Type** In order to understand the tissue-based importance of NTRA parameters in predicting DM or HTN, the 11 parameters were grouped by tissue type and used to classify DM and HTN similarly to what was done previously in the cardiac pathophysiology chapter 3.1.2. Results from these models were then assessed for comparative tissue feature importance in each classification. These results showing classification by tissue type are shown in Table 3.16 (with SMOTE) and Table 3.17 (without SMOTE), and tissue importance is represented in Fig. 3.8

<i>Evaluation metrics for the classification of DM</i>						
<b>Alg.</b>	<b>Acc.</b>	<b>Prec.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>F-Meas.</b>	<b>AUCROC</b>
<b>RF</b>	89.6	0	0	100	NaN	0.655
<b>GB</b>	82.5	17.2	18.5	89.8	17.8	0.597
<b>ADA-B</b>	87.7	18.3	5.6	97.1	8.6	0.589
<i>Evaluation metrics for the classification of HTN</i>						
<b>RF</b>	78.4	78.7	0.2	99.4	0.4	N/A
<b>GB</b>	93.3	79.4	90.8	12.3	84.7	N/A
<b>ADA-B</b>	73.3	78.9	91.4	9.5	84.7	N/A

Table 3.15: Classification metrics [%] of DM and HTN, Independent from Lifestyle Index without SMOTE

<i>The classification of DM by tissue type</i>							
<b>Tissue</b>	<b>Alg.</b>	<b>Acc.</b>	<b>Prec.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>F-Meas.</b>	<b>AUCROC</b>
<b>Fat</b>	<b>RF</b>	82.9	80.6	86.7	79.1	83.5	0.910
	<b>GB</b>	74.7	73.2	78.0	71.4	75.5	0.839
	<b>ADA-B</b>	72.0	69.6	78.0	66.0	73.6	0.782
<b>Muscle</b>	<b>RF</b>	82.9	81.2	85.6	80.1	83.3	0.907
	<b>GB</b>	74.8	73.2	78.4	71.2	75.7	0.841
	<b>ADA-B</b>	68.4	69.1	66.7	70.1	67.9	0.757
<b>Connective</b>	<b>RF</b>	79.7	78.1	82.6	76.8	80.3	0.883
	<b>GB</b>	75.4	74.2	78.1	72.8	76.1	0.841
	<b>ADA-B</b>	68.6	66.8	73.8	63.3	70.1	0.751
<i>The classification of HTN by tissue type</i>							
<b>Fat</b>	<b>RF</b>	76.4	72.1	63.7	87.7	67.7	N/A
	<b>GB</b>	72.5	66.6	60.6	84.8	63.5	N/A
	<b>ADA-B</b>	72.3	65.6	61.2	83.9	81.7	N/A
<b>Muscle</b>	<b>RF</b>	76.1	73.0	66.1	87.8	69.3	N/A
	<b>GB</b>	73.0	69.1	63.9	85.7	66.4	N/A
	<b>ADA-B</b>	74.1	67.9	64.9	84.7	66.3	N/A
<b>Connective</b>	<b>RF</b>	73.4	70.0	62.0	86.7	65.8	N/A
	<b>GB</b>	70.4	66.1	60.2	84.5	63.0	N/A
	<b>ADA-B</b>	71.3	65.1	61.3	83.5	63.1	N/A

Table 3.16: Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms when grouping NTRA Parameters by tissue type with SMOTE

<i>The classification of DM by tissue type</i>							
<b>Tissue</b>	<b>Alg.</b>	<b>Acc.</b>	<b>Prec.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>F-Meas.</b>	<b>AUCROC</b>
<b>Fat</b>	<b>RF</b>	89.6	30.8	1.3	99.7	2.5	61.8
	<b>GB</b>	81.5	15.1	17.5	88.8	16.3	0.603
	<b>ADA-B</b>	88.9	30.0	6.0	98.4	9.9	0.582
<b>Muscle</b>	<b>RF</b>	89.7	0	0	99.9	NaN	0.506
	<b>GB</b>	79.8	7.0	7.9	88.0	7.5	0.479
	<b>ADA-B</b>	89.7	0	0	99.9	NaN	0.529
<b>Connective</b>	<b>RF</b>	89.3	16.7	1.0	99.4	1.9	0.583
	<b>GB</b>	80.9	13.5	15.9	88.4	14.6	0.571
	<b>ADA-B</b>	89.1	11.5	1.0	99.1	1.8	0.605
<i>The classification of HTN by tissue type</i>							
<b>Fat</b>	<b>RF</b>	77.7	78.8	98.1	1.8	87.4	N/A
	<b>GB</b>	71.7	79.0	89.1	11.7	83.7	N/A
	<b>ADA-B</b>	74.1	78.8	92.7	7.2	85.2	N/A
<b>Muscle</b>	<b>RF</b>	78.1	78.9	98.7	1.8	87.7	N/A
	<b>GB</b>	73.1	79.0	90.8	10.4	85.5	N/A
	<b>ADA-B</b>	75.0	78.9	94.0	6.4	85.8	N/A
<b>Connective</b>	<b>RF</b>	77.6	79.0	98.1	3.0	87.5	N/A
	<b>GB</b>	72.6	79.5	89.7	14.3	84.3	N/A
	<b>ADA-B</b>	78.3	78.8	99.2	1.1	87.9	N/A

Table 3.17: Evaluation Metrics [%] for RF, GB, and ADA-B Algorithms when grouping NTRA Parameters by tissue type without SMOTE

Each of the three tissue types generally yielded excellent results for the independent detection of DM and HTN, albeit with lower evaluation scores than our previous ML models. Fat and muscle had comparatively higher scores than fat across all algorithms, and RF models again outperformed GB and ADA-B algorithms across nearly all comparisons.

Fig. 3.8 shows differences in feature importance by tissue type defined by NTRA parameters, which allows for the comparison of the respective contributions from fat, muscle, and connective tissue values towards the accuracy of DM and HTN prediction. Results from this assessment indicate that fat holds the strongest role in classifying the presence of both DM and HTN, respectively contributing 47.14% and 35.83% towards each model’s accuracy. Connective tissue contributed more towards model accuracy than muscle in both predictive models. This hierarchical importance of fat and connective tissue over muscle is in agreement with results from our pairwise statistical comparisons and ML models.

**Prediction of DM Incidence from Longitudinal Data** Finally, a longitudinal DM assessment was performed to simulate the prediction of DM incidence. Here, NTRA parameters computed from AGES-I were used to classify individuals who received a DM diagnosis between the two study timepoints. In order to perform this analysis, an incidence index is created similar to what was done for CHF (chapter 3.1.2.1). Table 3.18 and Table 3.19 contain the evaluation metrics from this assessment with and without SMOTE.

This analysis further supports the predictive utility of NTRA parameters in the detec-



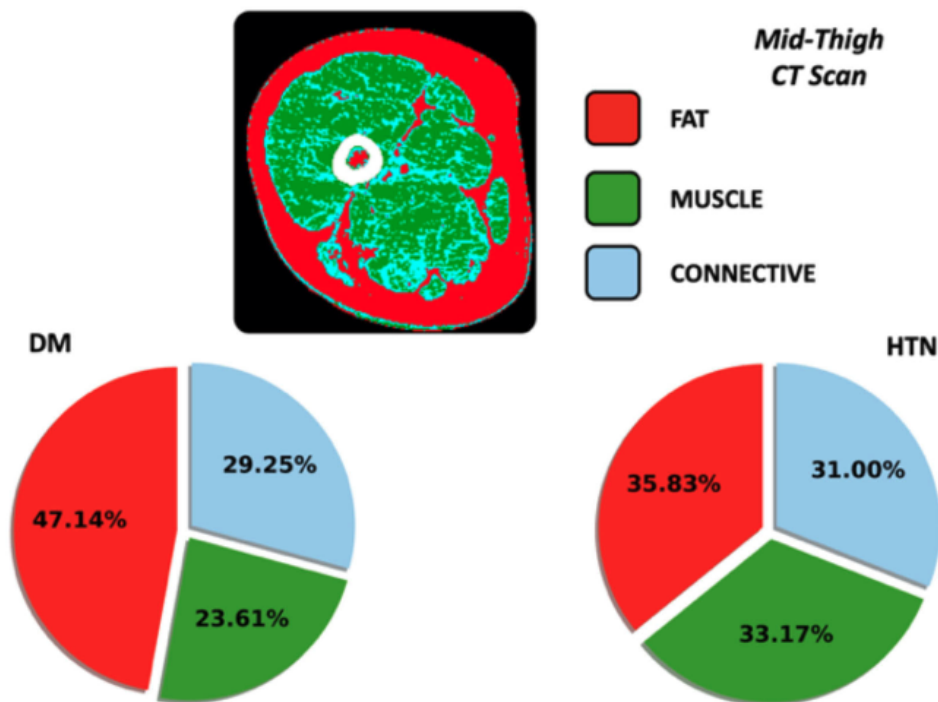


Figure 3.8: Comparative representation of tissue importance in predicting DM and HTN from NTRA parameters.

Alg.	Acc.	Prec.	Sens.	Spec.	F-Meas.	AUCROC
<b>RF</b>	94.9	97.5	92.3	97.6	94.8	0.991
<b>GB</b>	88.2	90.5	85.3	91.0	87.8	0.956
<b>ADA-B</b>	93.4	96.0	90.6	96.3	93.3	0.984

Table 3.18: Evaluation Metrics [%] in the Longitudinal Prediction of DM Incidence from AGES-I to AGES-II with SMOTE

Alg.	Acc.	Prec.	Sens.	Spec.	F-Meas.	AUCROC
<b>RF</b>	96.2	96.2	100	0	98.1	0.636
<b>GB</b>	92.4	96.4	95.6	9.3	96.0	0.631
<b>ADA-B</b>	95.7	96.2	99.4	0	97.8	0.579

Table 3.19: Evaluation Metrics [%] in the Longitudinal Prediction of DM Incidence from AGES-I to AGES-II without SMOTE

tion of DM incidence. In particular, a combined accuracy of 94.9% with an AUCROC of 0.991 from the RF model suggests that NTRA parameters can serve as robust quantitative indicators for the onset of DM. The ADA-B algorithm achieved similar results, combining an accuracy of 93.4% with an AUCROC of 0.984, and the GB model was again outperformed across every evaluation metric.

### 3.1.3.3 Discussion of the main findings

**Univariate Statistical Analyses** The strong age dependency of many chronic diseases and comorbidities likely reflects the cumulative effects from a variety of risk and protective factors that occur over one's life course. This notion has become increasingly explored in the etiology of diseases in aging populations. In the previous chapter we shown the robust classification of individuals at risk for cardiovascular pathophysiology using CT-based NTRA parameters. However, investigating the classification of DM and HTN from radiodensitometry in this regard remained unexplored, along with the commensurate modifying roles of smoking status and physical activity as key lifestyle factors.

For the present binary tree, univariate statistical analyses between node couples resulted in significant differences across all levels. In order to best interpret these differences, it is first important to outline the hypothesized physiological meaning behind each of the NTRA parameters and their respective tissue types. The amplitude parameter (N) largely reflects the relative abundance of CT image HU values for a given tissue, which may be understood as an indirect quantitative indicator for declining tissue mass [143]. In contrast, the location ( $\mu$ ), width ( $\sigma$ ), and skewness ( $\alpha$ ) terms generally serve as a multimetric quantitative construct for myosteatosis, a key characteristic of sarcopenic muscle degeneration defined by the reduction of dense contractile myofibers and commensurate infiltration of non-contractile adipose tissue [156][152]. In CT imaging, this characteristic adipose tissue infiltration drives what is known as the Partial Volume Effect (PVE), defined as the loss of fidelity in small regions or morphologies due to limitations in spatial resolution [255]. NTRA parameters uniquely capture this phenomenon, as the accumulation of pixels subject to PVE due to the presence of myosteatosis yields an increase in fat and lean muscle HU distribution width and skewness, as well as an inward shift in fat and muscle peak locations toward 0 HU. This may furthermore result in an increase in loose-connective (water-equivalent) pixels, resulting in an increased amplitude and decreased width in connective NTRA. In accordance with this physiological interpretation of NTRA, it is possible to infer whether myosteatosis indicative changes in these parameters are reflected by the different subgroups in the present binary tree. At the first level, where participants were sorted into LSH and LSNH groups, all of the observed significant differences in NTRA parameters indeed indicated that physical activity and abstinence from smoking may protect against sarcopenia. The LSH group had significantly narrower fat and muscle peaks, higher muscle  $\mu$ , reduced muscle  $\alpha$ , and a smaller connective tissue peak with its  $\mu$  further from 0 HU.

At the second level of univariate statistical analyses, which further divided subjects according to the presence of DM or HTN, significant differences in NTRA values were less apparent; indeed, there were no significant differences between any of the LSH subgroups. While this may have been due to the loss of statistical power from progressively smaller subgroup sample sizes, between-group differences in LSNH individuals again indicated healthier muscle in subjects without a DN or HTN diagnosis. LSNH subjects presenting with one of these comorbidities yielded significantly higher fat and connective tissue amplitudes, as well as significantly wider lean muscle distributions and a significant shift in fat  $\mu$  values toward 0 HU.

Finally, at the third level, further division according to cardiovascular pathophysiology diagnosis yielded significant NTRA differences independent of lifestyle classification, but dependent upon the presence of DM or HTN. For those diagnosed with either

comorbidity, individuals with CHD, CVD, or CHF again showed NTRA parameters that were largely indicative of myosteatosis, compared to those in good cardiovascular health. However, fat and muscle amplitudes showed an inverse relationship in this regard, where individuals with cardiovascular pathophysiology had significantly higher muscle amplitudes and lower fat amplitudes.

**Machine Learning Classification** The differing specificities of NTRA parameters across levels of the present binary tree motivates the use of more sophisticated classification methods, such as the three different ML algorithms explored in this work. In this regard, overall classification scores from ML analyses consistently indicated that the GB algorithm performed the poorest of the three, while the RF model scores were consistently the highest. This is in accordance with our previous ML classification work with NTRA and cardiovascular diseases, suggesting that the learning technique of randomization and bagging is optimum for NTRA classification.

ML classification generally yielded high evaluation scores across the first and second levels of the present binary tree, but with the best performance achieved at the second level for the classification of LSNH individuals diagnosed with DM or HTN. In particular, the RF model sensitivity, specificity, and accuracy exceeded 90.0%, indicating the model's high capacity for disseminating comorbidity presence in LSNH individuals. Furthermore, RF model AUCROC values were largely higher than 0.9, suggesting these models are highly capable of distinguishing between binary tree classes. Finally, independent classification performed worse across all evaluation metrics than second-level classification following LSH and LSNH group designation. This suggests that the classification of individuals according to lifestyle health served to improve the predictive capacity of NTRA-based assessment.

The predictive value of the NTRA parameters extracted from a CT scan has already been demonstrated, and the specific importance of connective tissue in classification has been underlined before but not with such significant results. Physiological assessment of loose connective tissue is often considered at a microscopic level rather than from a CT scan, where lean muscle and adipose tissues are largely considered primary targets for analysis. In this work, we have shown how connective NTRA parameters are not only statistically significant in distinguishing subject lifestyle health, but also in the identification of subjects with cardiovascular pathophysiology. Our ML models further demonstrate the impact of connective tissue in the classification of patients with DM or HTN. Indeed, DM classification models illustrate the comparatively high importance of connective tissue compared to lean muscle or adipose tissue, despite having one less NTRA parameter. Taken together, these results contribute to the notion that connective tissue may serve an important role in CT-based analyses.

Regarding tissue-based feature importance, several important differences are apparent when comparing cardiovascular pathophysiology classification. Firstly, fat was the predominate tissue when classifying individuals with CHD (41.0%), while muscle gave a comparatively minor contribution (11.9%). Contrastingly, lean muscle contributed the most in classifying CHF (41.0%), while connective tissue yielded the lowest contribution (24.9%). Finally, fat and connective tissue gave nearly the same contribution in classifying CVD (about 33.2% and 31.3%, respectively), while lean muscle was comparatively much lower (17.6%). These condition-based differences in classification indicate the potential specificity of tissue types to each condition, further suggesting

the importance of segmenting classifying parameters by these three tissue types, which is one of the key features of NTRA computational modelling.

Finally, the present longitudinal assessment from AGES-I to AGES-II showed that NTRA assessment can be considered not only for the predictive assessment of aging individuals according to their lifestyle and comorbidities, but also for the real prediction of DM. In particular, a combined accuracy of 94.9% with an AUCROC of 0.991 from the RF model suggests that NTRA parameters can serve as robust quantitative indicators for DM incidence.

**Strengths and Limitations** Altogether, the present work further extends our existing research on the assessment of health outcomes featuring the NTRA method and serves as an important milestone towards understanding the comparative roles of lifestyle factors and muscle health in promoting healthy aging. This represents, to our knowledge, the first time that NTRA-based radiodensity, or a similar approach, has been utilized in predictive models of DM and HTN. While this indeed represents a novel approach, the comparative value of our methods against traditional metrics for predicting risk has yet to be established, presenting an important direction for future investigation. Nevertheless, the primary strengths of this work stem from the use of ML classification within the large sample size offered by the population-based AGES-Reykjavik study. In this regard, this work is situated within the general context of the ‘big data’ movement in medical science research. A primary objective of this movement is to match the rapidly-accruing body of clinical or biological data with commensurate methods for pattern identification and computational modelling, such as the ML classification methods described here [256]. This data-oriented approach to clinical research represents an emergent yet growing field that continues to show promise in defining new tools for prognostic or diagnostic evaluation.

The main limitation of this research stems from the characterization of our lifestyle health index; while smoking status and physical activity are both robust indicators for health outcomes in aging, it may be important in future works to investigate additional lifestyle factors, such as Geriatric Depression Score, dietary intake, and alcohol use. Likewise, the consideration of key demographic data, such as income or education, may elucidate whether the observed relationship between NTRA parameters and comorbidities is partially modified by other factors. Some of these aforementioned lifestyle factors will be included in chapter 5.4 where they will be considered to study Motion Sickness using biosignals from brain, muscle and center of pressure acquisitions.

Finally, the main methodological limitation in this study is the use of SMOTE to balance the dataset. Nevertheless, the implementation of ML algorithms without SMOTE showed that the imbalance in sample size across tree nodes necessitated oversampling in order to perform most of the present analyses. Indeed, this imbalance between subgroups occurred primarily at the second level – especially for LSNH subjects, where subjects without diagnosed DM or HTN accounted for only 116 out of the total sample of 2477 individuals. At the third level, it is furthermore important to note that the highest number of subjects ( $n = 693$ ) presenting with at least one of the three cardiovascular pathologies was in the unhealthiest overall group, LSNH with DM or HTN. These differences were further reflected in the statistical analyses of the NTRA

parameters relating to the subgroups considered.

**Conclusion** finally we can say that this work serves as an important milestone toward the construction of predictive imaging tools for assessing the impact of lifestyle factors in promoting aging health. Our findings discussed here extend our existing work on NTRA-based soft tissue radiodensity assessment through ML-based classification, demonstrating the value of image features from CT scans to predict DM onset and, particularly, the comparative importance of lean muscle, adipose, and connective tissues toward promoting healthy aging.

## 3.2 Single and Dual Gait Analysis in Older Korean Population

### 3.2.1 Seoul National University Hospital Gait Dataset

In this chapter spatio-temporal gait features are taken in consideration to classify elderly patient affected by gait symptoms, neurological deficits or with an history of ischemic or hemorrhagic stroke. The data were obtained at the Department of Neurology and Critical Care of the National University Hospital in Seoul, South Korea. The work was done between May and October 2022 at Chonnam National University, in Yeosu, South Korea with the supervision of Prof. Seung-Uk Ko in the frame of the PhD exchange program supported by Reykjavik University. The following text is adapted from the manuscript actually under review in the journal *Gait&Posture*.

In the current study, we examined single and dual gait performance tests and used these gait characteristics to perform tree-based ML algorithms to classify patients with history of strokes, motor, and neurological disorders. Our study approach strengthens the importance of the ML analysis and the tree-based approach in medical datasets when using gait characteristics (with a considerable emphasis on dual-task features) to predict physical and neurological impairments among Korean older adults with specific disease history. Our findings may further help clinicians and therapists to diagnose motor and neurological disorders as well as to provide additional information to rehabilitation strategies planning for stroke patients.

#### 3.2.1.1 Materials and Methods

The dataset contains 122 non-hospitalized older adults of Korean origin with an average age of  $76.08 \pm 7.98$  years (age range 60-96). All the patients were diagnosed by the doctors during their first visit in the clinic with a neurological disease, including movement disorders, peripheral disease, headache, dizziness, and stroke.

**Gait Acquisition** Single and a dual-task gait tests were performed at the Department of Neurology and Critical Care of the Seoul National University Hospital in Seoul, South Korea. Gait tests (both single and dual-task) were conducted using the gait analyzer system Tango STEP Lite (Tango+ Life Design, Gwangju, South Korea), with a sampling frequency of 60 Hz on a 6 meters mat while participants were walking in a straight line at a normal speed. Spatio-temporal features including gait speed, stride length, and stride width were calculated in bundle by Tango STEP Lite, and all of

them were manually checked by a technician using custom made software written in R as previously applied [18]. Dual-task gait in the current study consisted of a cognitive assignment in which participants were asked to walk while saying out loud the days of the week in reverse order.

**Gait Features** Gait features include mean Step Length [meters] (*StepL*) and *StepL ratio*. *StepL ratio* is calculated and normalized between 0 and 1 with the quotient of the left and right leg measures. The same mean and ratio calculation is computed for Stride Length [meters] (*StrideL*) and *Speed* [meters/second]. In addition, *Step Width* [meters] and *Cadence* [steps/minute] are extracted. These 8 features are calculated for both single and dual gait, making a total of 16 input features for the ML analysis. “PO” prefix will distinguish the dual-task features from the single-task ones in the nomenclature of this manuscript.

**Demographic and Clinical Characteristics** Demographic and clinical characteristics (13 features) were collected by the clinicians during the visit before the gait test acquisition. Three clinical features were identified: *Stroke* (ST), *gait symptoms* (GS), and *Neurological Sequelae* (NS) based on the clinical history of diagnosis and symptoms. ST, GS, and NS are the features considered for the binary classification (yes=1, no=0) with ML algorithms.

*Stroke* (ST): Patients who reported history of ischemic or a hemorrhagic stroke (90 subjects) were defined as having ST.

*Gait Symptoms* (GS): Patients who reported suffering from gait and movement disorders including imbalance, shuffling, staggering, freezing, or ataxia were defined as having GS (72 subjects).

*Neurological Sequelae* (NS): Patients who reported suffering from focal neurological deficit including limb weakness (hemiparesis, monoparesis), sensory symptoms, dizziness, and imbalance related to neurological diseases were defined as having NS (49 subjects).

Not all the patients were able to successfully complete the dual-task saying out loud all the days of the week in the inverse order, so a binary *task error* feature is added: 56 subjects failed to correctly enunciate the names of the days backwards. Moreover, an additional feature is the binary *musculoskeletal pain*, which consists of 32 subjects suffering any kind of pain related to muscles or bones. *Symptoms Side* is the side of lower extremities associated with gait abnormality which can be left, right, bilateral, or absent.

Supplementary features include *age*, *sex*, and comorbidities like *hypertension* (HT), *diabetes* (DM), *atrial fibrillation* (AF), *coronary artery disease* (CAD) and *dyslipidemia* (DL). There are in total 29 features for each patient including gait, demographic, and clinical data (data not shown).

All the features listed in this paragraph are denominated as “*clinical data*” in the text.

**Statistics and Machine Learning** To study the independent roles of the gait features all of them underwent a Kolmogorov Smirnov test. The results show that none follow a normal distribution. Therefore, Mann Whitney tests was performed to understand which features can be considered statistically significant (p-value <0.05).

In this ML analysis three different tree-based ML algorithms were performed for the

classification analysis: Random Forest (RF), Gradient Boosting (GB) and ADA Boosting (ADA-B). These algorithms exploit different ensemble learning techniques and they were executed with the same random seed and with the default hyperparameters proposed by the Python SciKit-Learn library [257].

A feature importance investigation was performed on the ML model which was the most representative with regard to classification metrics.

**Cross-Validation and Classification Evaluation Metrics** For a complete classification performance evaluation k-fold cross validation is computed. A repeated stratified 10-fold cross validation is here implemented: the k-fold is repeated 3 times producing different splits in each repetition and the stratification produces sets of patients having approximately the same percentage of samples of each target class. A tuning on 10 different k-fold random seeds is computed to obtain the highest possible mean accuracy.

The following evaluation metrics are computed: accuracy mean and accuracy max (referred to the maximum accuracy achieved in a single cross validation fold), precision, recall, F-Score (F1), and Area Under the Curve Receiver Operating Characteristic (AUCROC).

**Features selections** Two sets of features are used to classify ST, GS, and NS in the Korean older population. The first set consist of all the 16 gait features (single and dual gait) plus the clinical data mentioned above. For the ST prediction all the features are considered, a total of 28 plus the ST class. For the GS and NS classification the *Symptoms Side* is excluded as it caused an algorithmic overfitting, so the final total of features is 27 plus the prediction class. The second set of features considers the 16 gait features solely.

### 3.2.1.2 Results

**Gait features by the history of ST, GS, and NS.** The summary of the descriptive characteristics according to the history of ST, GS, and NS is presented in Table 3.20. Table 3.21 shows statistically significant gait features in group of ST, GS, and NS (significant p-value <0.05).

For the ST, only three gait features had a p-value below 0.01 and all of them were from the dual-task set (Step Length, Stride Length and Speed). The same features for the single-task gait had p-value below 0.05.

For the GS the same 5 features were statistically significant both for the single and the dual-task gait (Step Length, Step Length\_ratio, StepWidth, Stride Length and Speed), making a total of 10 out of 16 significant features.

The statistically significant gait features for distinguishing NS were 5 from single gait and 3 from dual gait. Cadence was significant only for this NS prediction class, even if with a p-value of only 0.047, and just two features had a p-value <0.001 (Step Length and PO\_Step Width).

Ratio features were significant only for the GS, and between them only the Step Length Ratio, single and dual, had a p-values <0.05.

**Machine Learning – Classification** Table 3.22 shows the evaluation metrics obtained from the tree-based ML models for ST, GS, NS binary classification. For

	ST		GS		NS	
	No	Yes	No	Yes	No	Yes
Subject, N	32	90	50	72	73	49
Age, yo, (SD)	76.13 (8.03)	76.07 (7.98)	76.23 (8.00)	76.00 (8.04)	77.27 (7.98)	74.31 (8.05)
Male, N	13	61	27	47	40	34
Female, N	19	29	23	25	33	15

ST: Stroke, GS: Gait Symptom, NS: Neurological Sequelae

Table 3.20: Sex differences and mean age in respect to ST, GS, and NS binary features

	ST	GS	NS
Statistically Significant Features		p-value	
Step Length	0.039*	0.000***	0.001***
Step Length Ratio	-	0.001***	-
Step Width	-	0.002**	0.002**
Stride Length	0.038*	0.000***	-
Speed	0.030*	0.000***	0.001**
PO Step Length	0.009**	0.001***	0.004**
PO Step Length Ratio	-	0.005**	-
PO Step Width	0.013*	0.001***	0.001***
PO Stride Length	0.006**	0.000***	0.004**
PO Speed	0.004**	0.004**	-

"-": Not Significant, \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

Table 3.21: The influence of ST, GS, and NS on GAIT features

all the three binary classes the features selection which includes both gait and clinical data resulted the most efficient between the two considered, especially in terms of accuracy mean and AUCROC. RF can be considered the most effective algorithm while ADA-B was the worst. GB had slightly lower metric than RF and overcome it in accuracy only for NS classification with gait and clinical features. For the ST classification the highest recall of all Table 3.22 (91.9) was reached with RF and gait and clinical selection: the same model reached the maximum mean accuracy of 73.5. Same algorithm with only gait features as input had almost 70 of accuracy but a low AUCROC of 0.53. The best metrics were achieved for the GS classification, RF and gait and clinical features: the accuracy mean was almost 80 and relevant 85.1 of F1 and 0.8 of AUCROC were registered meaning that the model can distinguish correctly both the positive and negative GS subjects. Decent results were also gained for the NS classification but only when we used the gait and clinical features together: GB was here the best model with 78.2 of accuracy and 0.81 of AUCROC. At the opposite, the gait features can barely reach an accuracy of 60 classifying NS.

**Machine Learning – Features Analysis** The features importance was calculated for all the RF models, as this algorithm resulted the most reliable between the three considered, for all the classification performed. In only one occasion GB had a higher mean accuracy (NS prediction with Gait and Clinical features, Table 3.22). Table 3.23 shows the percentage of importance for each of the groups in the classification process while in Table 3.23 are printed the 8 most important features for each of the RF models.



Class	Feat Select	Alg.	Acc. Mean [%]	Acc. Max [%]	Rec [%]	Prec [%]	F1 [%]	AUCROC
ST	Gait and Cli	RF	73.5	83.3	91.9	77.2	83.9	0.691
		GB	72.6	91.7	86.7	78.6	82.5	0.746
		ADA-B	73.0	91.7	83.7	81.1	82.4	0.720
	Gait	RF	69.3	91.7	87.0	75.4	80.8	0.530
		GB	68.6	83.3	85.2	75.8	80.2	0.496
		ADA-B	62.4	83.3	76.7	72.9	74.8	0.515
GS	Gait and Cli	RF	78.7	100	90.7	80.1	85.1	0.800
		GB	74.8	91.7	84.5	79.5	81.9	0.771
		ADA-B	69.9	91.7	79.0	75.4	77.2	0.703
	Gait	RF	71.9	91.7	82.7	76.6	79.5	0.721
		GB	68.4	91.7	80.6	73.9	77.1	0.662
		ADA-B	63.2	91.7	73.3	71.1	72.2	0.597
NS	Gait and Cli	RF	76.2	100	61.7	79.0	69.3	0.810
		GB	78.2	92.3	68.1	76.1	71.9	0.810
		ADA-B	74.7	91.7	67.5	72.8	70.0	0.790
	Gait	RF	62.6	83.3	44.7	54.2	49.0	0.621
		GB	59.7	91.7	42.0	51.8	46.4	0.573
		ADA-B	59.0	83.3	46.2	53.1	49.4	0.607

Table 3.22: ML learning models evaluation metrics for the ST, GS, NS, binary classification

Features Select	ST	GS	NS
	Impo [%]	Impo [%]	Impo [%]
<b>Gait+clinical</b>			
Single-Task Gait	37.05	42.84	35.33
Dual-Task Gait	39.75	36.38	37.8
Clinical Data	23.2	20.78	26.87
<b>Gait</b>			
Single-Task Gait	52.47	52.68	50.75
Dual-Task Gait	47.53	47.32	49.25

Table 3.23: Features groups relevance in % for all the RF classification models

Gait + Clinical Features

ST		GS		NS	
neurologicalsequelae	7.30	StepL	8.21	PO_StepWidth	5.89
PO_Cadence	6.81	StrideL	7.95	gaitsymptom	5.69
StepL_ratio	5.63	PO_StrideL	5.98	Speed	5.61
PO_StepL_ratio	5.38	Speed	5.96	StepWidth	5.50
PO_StepWidth	5.31	StepL_ratio	5.84	PO_Speed_ratio	5.42
PO_Speed	5.21	PO_StepL_ratio	5.49	PO_StepL_ratio	5.29
StepL	5.10	PO_StepL	5.05	PO_StepL	5.24
Speed	4.96	musculoskeletalpain	5.00	age	4.79

Table 3.24: Most important features and relative importance in % for all the RF classification models using Gait + Clinical Features Selection

Gait Features

ST		GS		NS	
PO_StepWidth	8.12	StepL	9.24	PO_StepWidth	8.28
PO_Cadence	8.09	StrideL	8.54	StepWidth	7.75
PO_Speed	7.36	StepL_ratio	8.18	Speed	7.74
StepL_ratio	7.20	PO_StrideL	7.82	PO_StepL_ratio	7.01
PO_StepL_ratio	6.66	PO_StepL_ratio	7.48	Cadence	6.57
StepL	6.33	Speed	7.45	PO_StrideL_ratio	6.36
StepWidth	6.24	PO_StepWidth	6.92	PO_Speed_ratio	6.34
PO_StrideL	6.20	PO_StepL	6.33	StepL_ratio	6.32

Table 3.25: Most important features and relative importance in % for all the RF classification models using Gait Features Selection

It is possible to notice that the distribution of importance in percentage of the groups of features was quite similar for all the three classes: the single and dual-task gait had approximately the same half importance if the only gait features were used and the clinical data had an impact between the 20 and 26% in the classification processes using the whole features set (Table 3.23). On the opposite, between each predicted class, the most relevant single features were rather different (Table 3.24 and Table 3.25). Comorbidities like HT, DM, AF, CAD, and DL did not have a high importance for the considered tree-based ML models (data not shown).

To better study the relation between the gait features and the age in the different binary groups of people affected by ST, GS, and NS, scatterplots with regression lines are presented for the most relevant features (Fig. 3.9, 3.10 and 3.11). It is noteworthy that the step width, both for single and dual gait, increased with age in patients with NS while decreased in patients without NS. The same feature (dual gait) had quite a different regression line for patient with and without ST, as it had quite similar values for younger and older subjects. Speed considerably dropped with age in subjects having NS and not having GS, while cadence, despite the high importance in the classification model, did not change much with age. The other features showed a more or less evident decline with age.

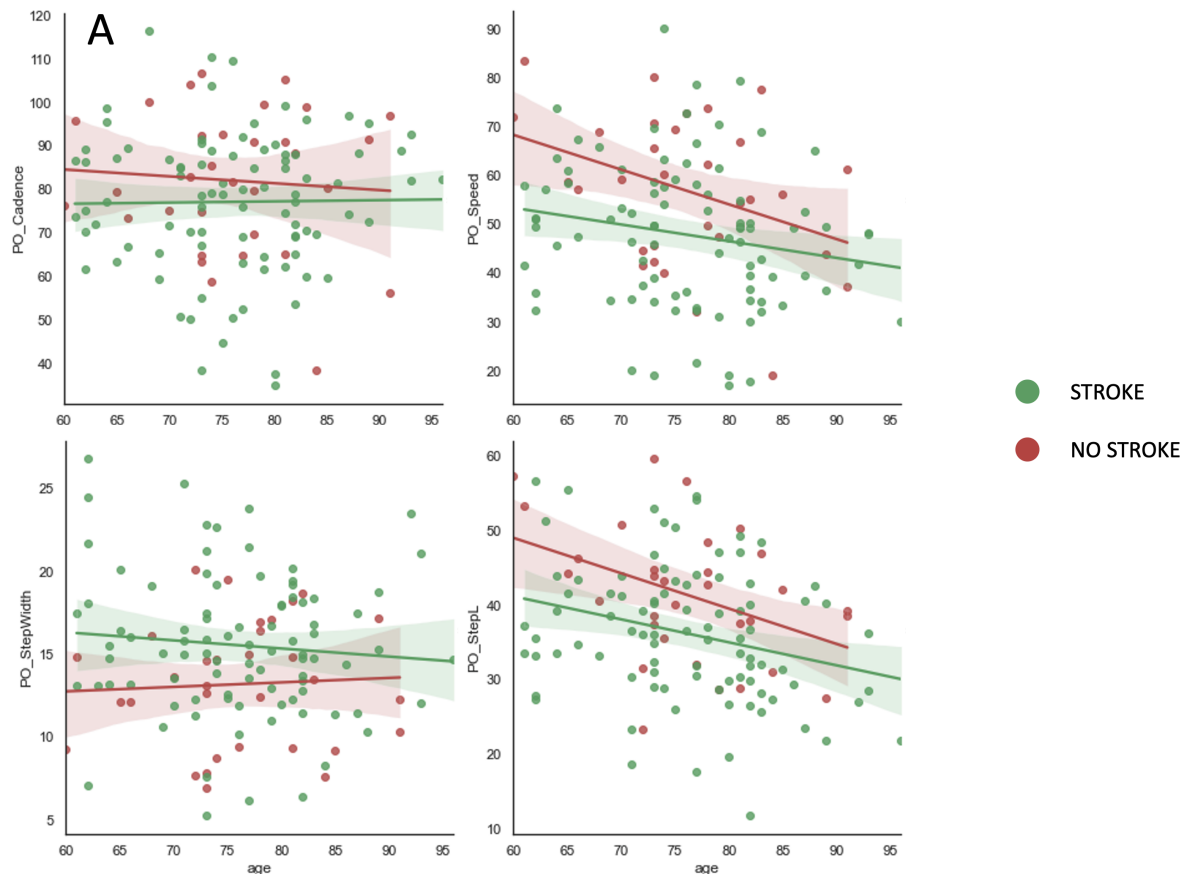


Figure 3.9: Scatterplots with regression lines for four of the most significant features in the classification of subjects with and without an history of strokes in relation to age.

### 3.2.1.3 Discussion

The current study examined Korean older adults performing tree-based ML algorithms to predict stroke and neurological and motor disorders using temporal and spatial gait features both from single and dual-task acquisition. Cognitive dual-task gait features proved their substantial importance in the classification process and ML tree-based algorithms results again of great efficiency when applied on biomedical datasets, also with a reduced number of patients like the present study. In recent research ML was performed to assess neurological disease like Parkinson [258] and Huntington Chorea [259] from gait features, or to predict possible strokes in older population with a gait monitoring system [260]. By the authors knowledge, no studies were published in relation to classification of ST in patients of older age group or with other disorders, using both single and dual gait except for Matsuura et al. [261]: they stated that both single and dual-task features were effective to predict a cognitive dementia score reaching a specificity of 0.799.

From our research's results we can state that dual-task gait assumes a significant relevance in the classification process to distinguish healthy and unhealthy subjects. Dual-task features always contribute for half of the importance with RF algorithms and for the ST classification, four out of six the most important features are from the dual gait analysis. Dual-task gait alone was implemented successfully to improve balance and physical performances in patient with stroke history [262], underlying

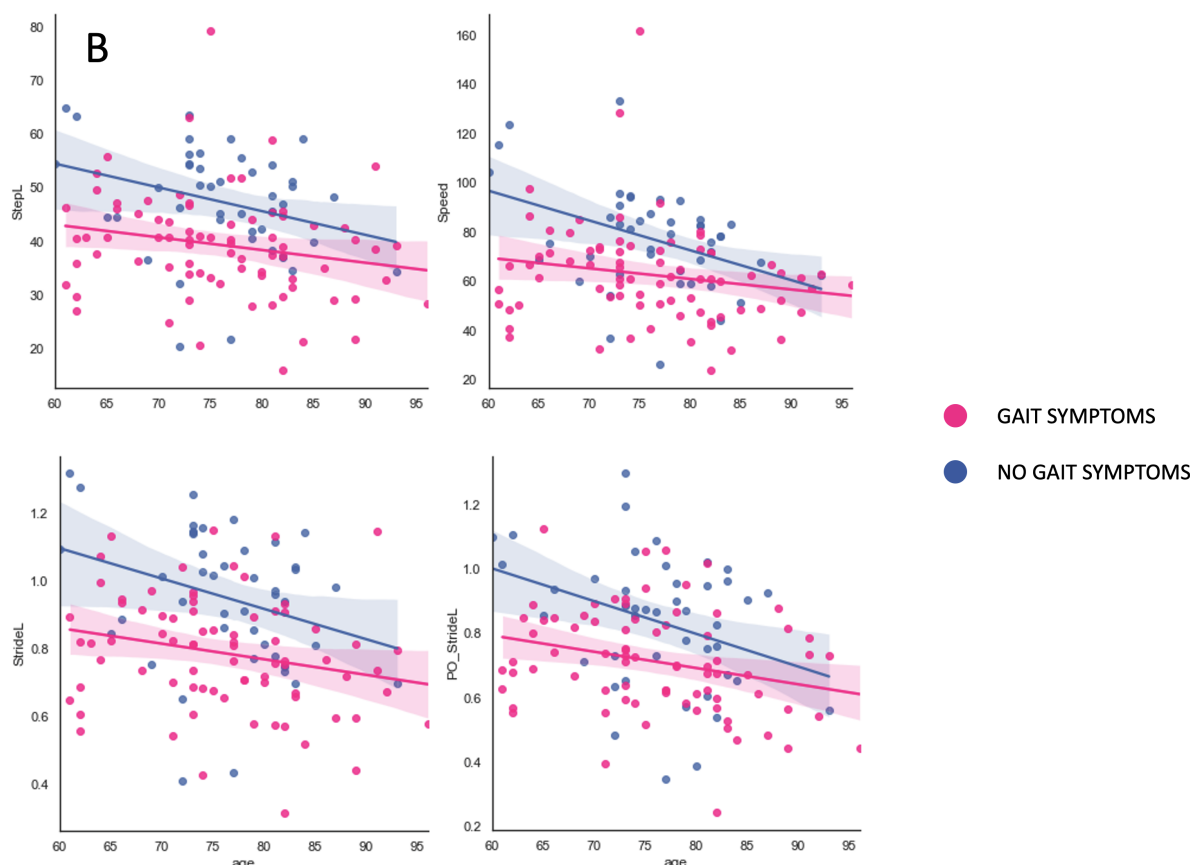


Figure 3.10: Scatterplots with regression lines for four of the most significant features in the classification of subjects with and without gait symptoms in relation to age.

that motor and cognitive task can have different effects [263]. Our results indicate that a cognitive dual-task gait performance is not simply an addition to the normal gait but can be considered as necessary together with the single-task in the analysis of older population, especially for those with an history of stroke. We also observed that clinical data were highly relevant as a group in the overall classification process, but individually they do not assume a high importance compared to the gait features except for the NS status in the ST classification. This would give a stronger importance to the prediction power of both single and dual gait features. From a ML point of view, the classification metrics indicate that tree-based algorithms, in particular RF and GB, are not only good and of easy understanding for clinicians [17], but significantly efficient if performed on biomedical dataset, also with a few numbers of subjects like the present one.

The current study has some limitations which includes the small number of subjects and the absence of healthy-control subjects. Improvements of these limitations would not only increase the meaning of the already significant classification metrics but would also allow physicians to compare healthy and unhealthy older adults. Previous studies highlighted some differences in gait features between Asian and Western population [264] [265] underling peculiar characteristics of Korean population: for example, Ryu et al. [266] found that stride length and speed are lower in Korean subjects. A possible future study can include the acquisition through the same gait system of older people of different ethnicity to understand which can be the abnormal gait differences in patients with strokes or motor and neurological diseases and different genetic characteristics.

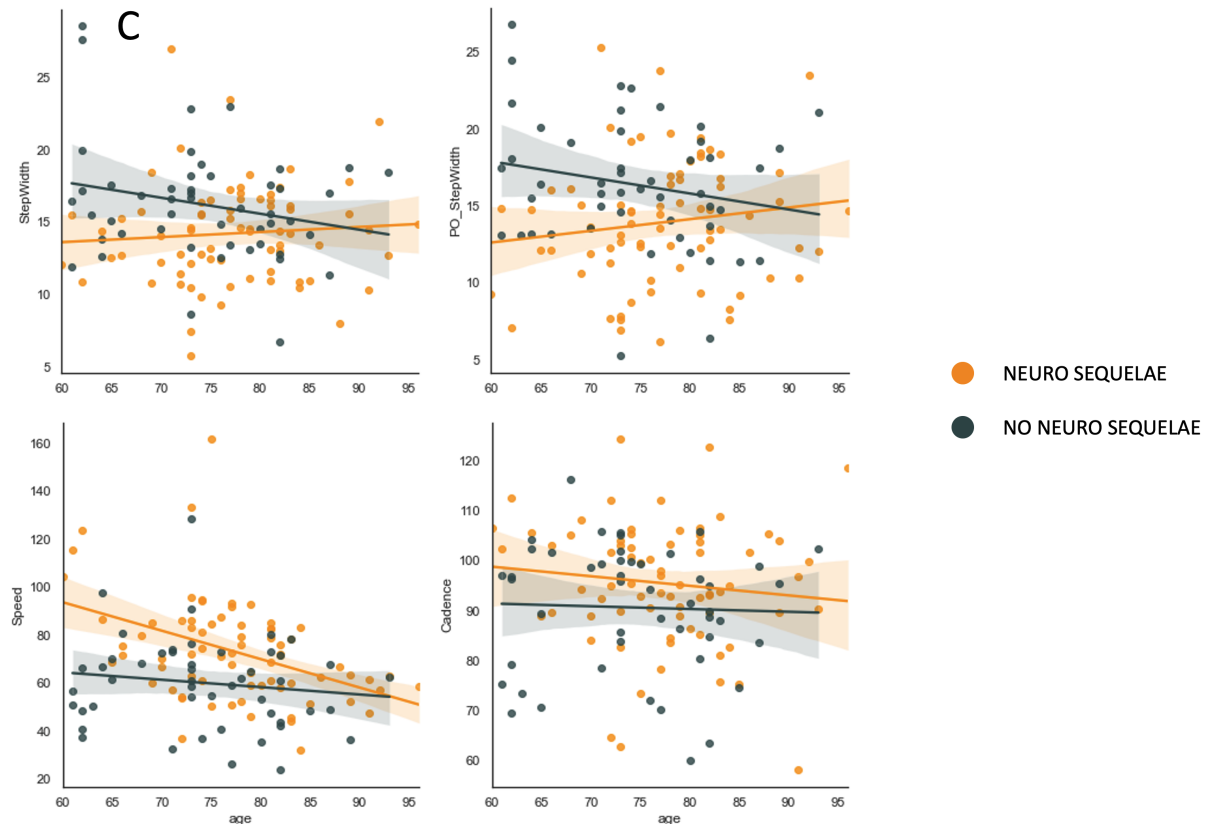


Figure 3.11: Scatterplots with regression lines for four of the most significant features in the classification of subjects with and without focal neurological deficit in relation to age.

In conclusion, the results of present study underline the importance of combining single and dual-task gait for the study of older adults with an history of stroke, motor, or neurological diseases. Furthermore, they can be of help for doctors and medical workers in the cure and diagnosis for these disorders and in the possible rehabilitation for stroke patients. Finally, they can also be of interest from a ML point of view, validating the great efficiency of the tree-based algorithms in the bio-medical dataset analysis.



# Chapter 4

## Knee Osteoarthritis Study

This chapter is dedicated to the medical imaging features analysis focuses on the use of novel and gold-standard parameters elaborated using knee CT Scans and MRI both from a 2D medical image and from a 3D rendering of the scans using advanced medical imaging softwares, in particular MIMICS, Materialize. These gold-standard and novel features were used to classify subjects with different knee cartilage conditions aiming to the development of new therapies for knee osteoarthritis treatment. The research is done in the frame of the European Union funded Horizon 2020 RESTORE Project.

### 4.1 European Union Horizon 2020 RESTORE Project

In the frame of the EU Horizon 2020 project Restore, Reykjavik University developed the 1st European database of chondral lesions morphometric and associated 3D models (<https://restore-project.ru.is/>) with the final scope of developing nanoenabled solution for personalised cartilage regeneration. In collaboration with the Icelandic University Hospital Landspítali in Reykjavik, 47 people's knees have been scanned. Three different categories of subjects have been evaluated: 25 degenerative (D), 14 traumatic (T), and 8 considered healthy as control (C). The database, available for free online, contains information regarding the different type of chondral lesions and the behaviour of the knee cartilage, based on Computed Tomography (CT) and Magnetic Resonance (MRI) images.

Measurements and features where extracted from the 2D images and from the 3D elaboration of the medical images. To have a exhaustive overview of the patient's condition from MRI and CT Scans, multiple 2D features were elaborated using a solid and well known radiological approach. Those includes thickness, grading of the cartilage, as well as the presence of medical pathologies like for instance cysts and bone attrition. Medical images were then exported and elaborated from a 3D point of view with an high performing medical imaging software (Materialise, MIMICS) to segment and dissociate the knee's sections including femur, tibia and patella. These features were then used as input in ML models to classify the different knee cartilage status.

The following section will relate to the study recently published on *Cartilage* [4]. This section will focus mainly on the use of 2D and novel 3D features from knee cartilage analysis to classify, using ML models, the traumatic, degenerative and con-

Category	Degenerative	Traumatic	Control
# Female	12	9	3
Mean age (Std)	66 (12)	39 (11)	29 (5)
# Male	11	7	5
Mean age (Std)	66 (7)	29 (7)	37 (16)

Table 4.1: Description of the patients demographics (age, sex) by group

control subjects. The detailed description of the features extraction can be read on the RESTORE published literature [4][267].

The main scope of ML technologies in this research is to study and understand the predictive potential of the features elaborated from the image analysis and their ability in distinguish degenerative, traumatic, and healthy subjects, with a focus on the features that contribute the most in the classification process. This study compares different cartilage assessment metrics, developing a novel workflow to 3D model bone and cartilage and, moreover, analyzing new features for a more sensitive cartilage assessment, currently required as a support element toward more patient-specific treatment development.

#### 4.1.1 Materials and Methods

Figure 4.1 shows the work done in this manuscript starting from the population recruitment to the data acquisition, analysis, and computation of the feature importance.

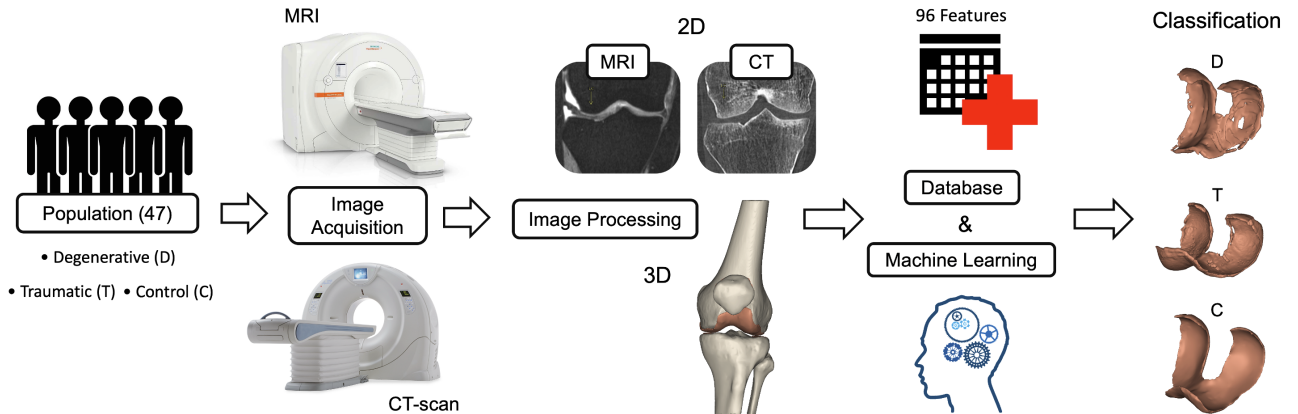


Figure 4.1: Graphical Abstract of the Cartilage Journal manuscript

##### 4.1.1.1 Participants

Participants were recruited as part of the European project RESTORE (<https://restoreproject.eu/>), whose objective is to develop solutions for personalized cartilage regeneration. The aim of our research group is to develop a database of morphometric chondral lesions with associated 3D models (cLDB). The function of the cLDB is to provide accurate 3D models of chondral status, bones, and soft tissue to develop, design, test, and validate 3D printed microtissues that can fit patient-specific lesions.



**Recruitment** 47 subjects (24 females, 23 males, age =  $50 \pm 19$  years) underwent CT and MRI scans of a single knee at Landspítali University Hospital in Reykjavik, Iceland, using standardized acquisition protocols and patient positioning. From the total of patients, 23 subjects (12 females, 11 males, age =  $64 \pm 12$  years) were suffering from degenerative (D) cartilage. They were examined by an orthopedic doctor due to pain from osteoarthritis and were placed on the waiting list for treatment with total knee arthroplasty (TKA). Sixteen (9 females, 7 males, age =  $35 \pm 11$  years) suffered from a knee trauma (T) with possible cartilage injury. The emergency clinic provided an alert when there is a patient with suspected ligament injury and patella dislocation. They underwent plain x-ray to exclude fracture. Then, they were called to exclude any history of knee injuries or problems. The alert was received within a week, and the patients underwent CT and MRI during the second week from the day of the trauma. Finally, 8 subjects (3 females, 5 males, age =  $34 \pm 14$  years) were involved in the study as control (C) subjects (no symptoms of history of knee trauma/degeneration). For D and T group, in addition to the CT and MRI, X-ray data were also available, as a part of the routine clinical evaluation detailed above. The X-rays were not performed for the C group.

Table 4.1 sums up the demographics of the patients.

#### 4.1.1.2 2D features

An exhaustive radiological examination was performed on the bones and articular cartilages of the knee joint to assess their condition. These observations were based on the 3 types of 2D medical imaging aforementioned: X-ray, CT and MRI. The assessment was done on femur and tibia from both the medial (MC) and lateral (LC) compartments as well as on the patella and femoral trochlea within the femoropatellar compartment (FPC) of the scanned knee.

**Bone** Figure 4.2 sums up examples and brief definitions of the pathologies observed on the femur, tibia, and patella. The subfigures on the left (A, C, E, G and I) correspond to CT scans while the right subfigures (B, D, F, H, and J) correspond to MRI scans. As observed in Figure 4.2, some pathologies could be found in both CT and MRI, while others (I, J) had a particular 2D image. Going into more details of the features shown in Figure 4.2, we can describe the multiple bone observations. (A, B) Subchondral bone cysts are typically spherical or ellipsoidal fluid-filled cavities within the subchondral bone region. (C, D) Osteophytes are cartilage-capped bony proliferations (spurs) that most commonly develop at the margins of a synovial joint as a response to articular cartilage damage. (E, F) Bone attrition is the result of flattening or depression of the articular surfaces, probably because of bone remodeling. (G, H) Osteonecrosis is a generic term referring to the ischemic death of the constituents of the bone and is observed as if the bone is missing a piece. (I) Subchondral bone sclerosis is a thickening of the bone seen in joints affected by OA. it is observed as a “whitening” of the bone only in Ct. (J) Subchondral bone edema is a build-up of fluid in the bone marrow as a response to an injury or Oa condition visible on MRI but not on CT.

**Cartilage and Joint Space** Figure 4.3 contains visual examples of the gold standard observations made on the cartilage and the joint space. Except for the Ahlbäck grading,

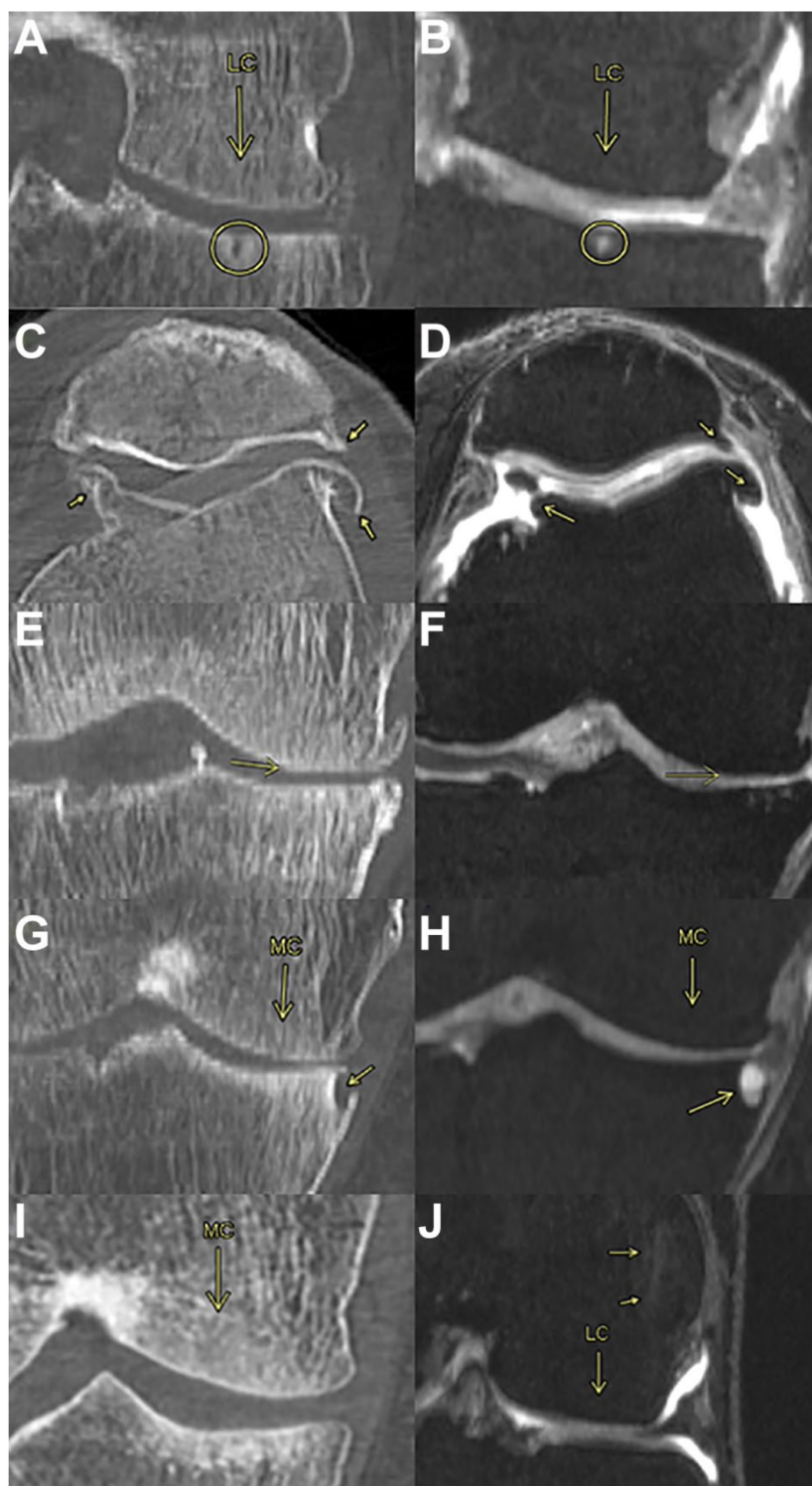


Figure 4.2: Bone observations.

which was observed on an X-ray, the rest of the observations were made on MRI scans. The features from the cartilage and joint space observed in Figure 4.3 are the following. (A, B, C, D) Ahlbäck grading is a classification system that focuses on the reduction of the joint space as an indirect sign of cartilage loss. (A) grade 0: normal. (B) grade 1: joint space narrowing (less than 3 mm). (C) grade 2: joint space obliteration (elimination). (D) grade 3: minor bone attrition (0-5 mm). (E,

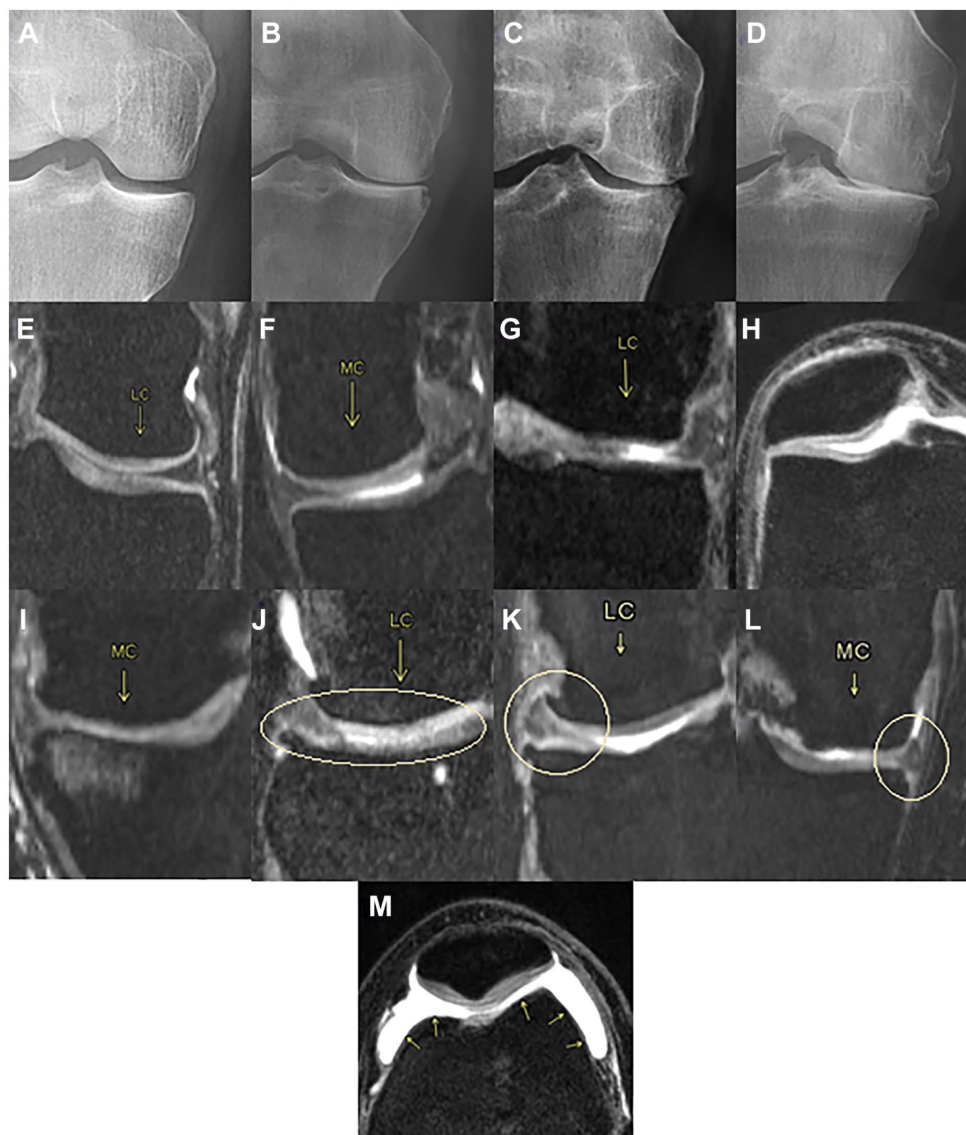


Figure 4.3: Cartilage and joint space observations.

F, G, H, I) iCrS (international Cartilage repair Society) grading is the most used score system for quantification of existing cartilage defects at the knee. (E) grade 0: normal cartilage. (F) grade 1: nearly normal cartilage. Superficial lesions; soft indentation and/or superficial fissures and cracks. (G) grade 2: abnormal cartilage. lesions extending down to <50% of cartilage depth. (H) grade 3: severely abnormal cartilage. Defects extending down to >50% of cartilage depth; down to calcified layer but not through the subchondral bone. Blisters. Defects more visible toward the medial area of the patella. (I) grade 4: severely abnormal. lesions extending down through the subchondral bone. (J, K, L) Meniscal pathology is associated with an elevated prevalence of Mri-detected cartilage damage. there are 3 types of pathology; (J) degeneration: not acute as a tear, this injury is a more gradual onset and tends to occur as we get older. (K) rupture: is a tear in the lateral or medial meniscus due to rotational forces directed to a flexed knee. (L) protrusion: when the location of the outer edge of a meniscus is beyond the tibial articular surface. (M) Synovitis—effusion. While synovitis is the inflammation of the synovium; effusion is when excess synovial

fluid accumulates in or around the knee joint. it is observed generally in the FPC as a white stain.

**Articular Cartilage Thickness** The measure of the articular cartilage thickness (ACT) were manually performed on an MRI scan of the knee following gold-standard methods proposed in literature [268] (Fig. 4.4).

From the same slices used to measure the femoral ACT in the medial and lateral compartments, the tibial ACT was measured (Fig. 4.5).

Femoropatellar compartment cartilage thickness measurements are shown in Figure 4.6.

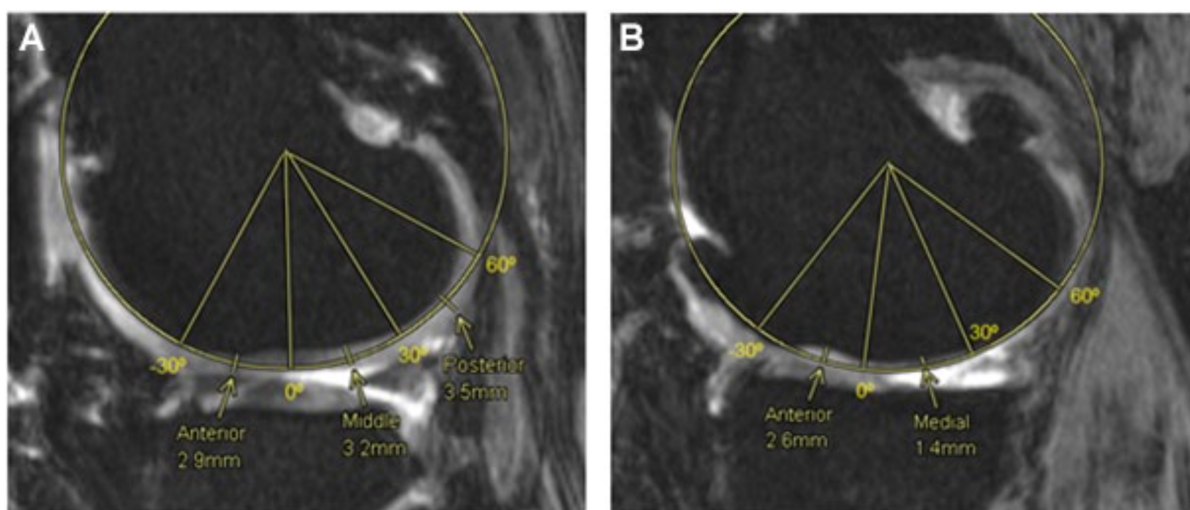


Figure 4.4: Femoral cartilage thickness measurements, medial and lateral compartments. Fitting cylinder method to obtain 3 regions of interest, anterior (-30o-0o), medial (0o-30o) and posterior (30o-60o) in the lateral compartment (A) and the medial compartment (B).

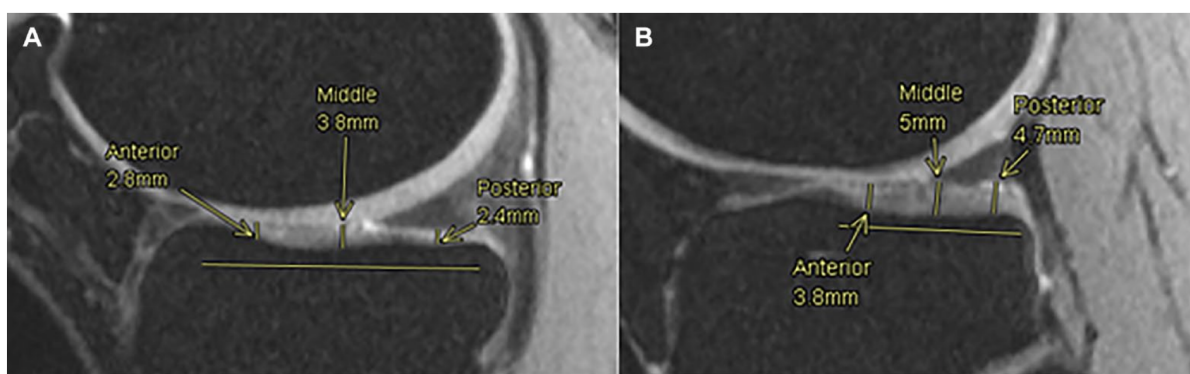


Figure 4.5: tibial cartilage thickness measurements. anterior, middle and posterior points are measured along the tibial cartilage in medial (A) and lateral (B) compartments.

**Cumulative Index Based on Bone Conditions** A cumulative index (CI) from 0 to 6 was used to quantify the bone anomalies present in each bone, that is, subchon-

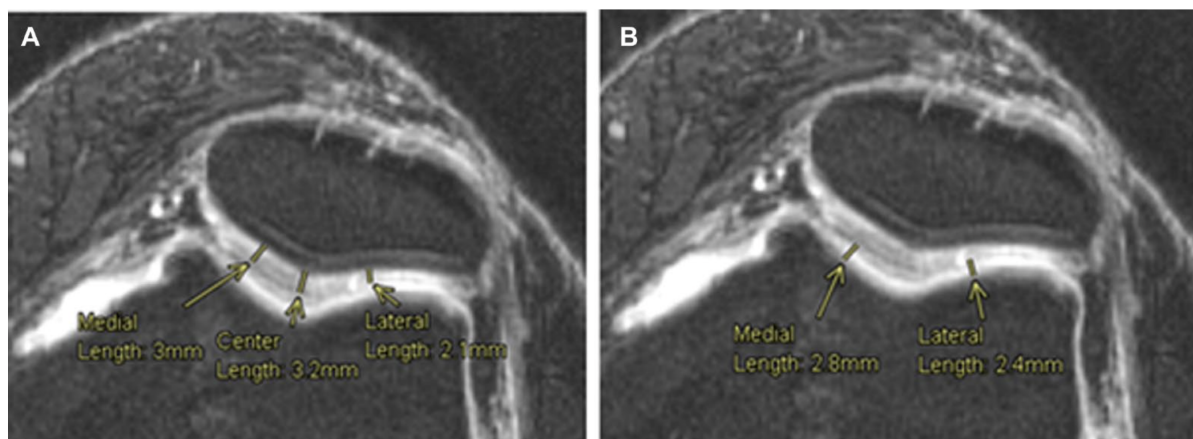


Figure 4.6: Cartilage thickness measurements, femoropatellar compartment. Measurement of the articular cartilage thickness at 3 points on the patella (A) and articular cartilage thickness of the femoral trochlea on 2 different points (B).

dral cysts, subchondral sclerosis, osteophytes, bone attrition, osteonecrosis, and/or subchondral edema, regardless of the compartment for each patient. Hence, for an observed pathology (either in CT or MRI) a 1 was assigned and consequently, the index was the sum of pathologies present in a certain bone of a certain patient. Sometimes, when the CI is compared against observations made within a compartment such as the Ahlbäck grading (AG) or the International Cartilage Repair Society (ICRS) score, the index is then considered for the bone in question within a compartment.

#### 4.1.1.3 3D features

Figure 4.7 describes the 3D features process workflow. It was repeated and evaluated 3 times by 3 biomedical engineers (co-authors of the paper), under the supervision of a senior engineer and an expert radiologist, to obtain the most accurate segmentation. From the CT scans, 4 objects were 3D calculated: the femur, the tibia, the patella, and the fibula (all bones). From the MRI, 4 objects were also 3D calculated: the femoral cartilage, the medial tibia cartilage, the lateral tibia cartilage, and the patellar cartilage.

From the final file of the MRI objects combined with the CT objects, the radiodensity of each part (bone and cartilage), was extracted in Hounsfield Units (HU). The bone mineral density (BMD) (in  $\text{g}/\text{cm}^3$ ) was computed from the radiodensity using a linear formula determined empirically based on phantoms. The cartilage radiodensity was then extracted from a final mask (Fig. 4.8). The volume (in  $\text{mm}^3$ ) and the surface (in  $\text{mm}^2$ ) were also computed from each 3D object.

#### 4.1.1.4 Machine Learning

In the present ML analysis, 2 tree-based algorithms were applied to the multi-classification of the degenerative, traumatic, and healthy (control) patients: random forest (RF) and gradient boosting (GB). RF was performed using the same random seed for every model and the same hyper-parameters (number of trees = 100, split criterion = Information Gain Ratio, maximum 3 depth = 10, and minimum node size = 1). The same was done with GB (number of trees=100, maximum 3 depth = 4, and learning rate = 0.1). The 10-fold cross validation was performed. Accuracy, precision, recall, and F1 have

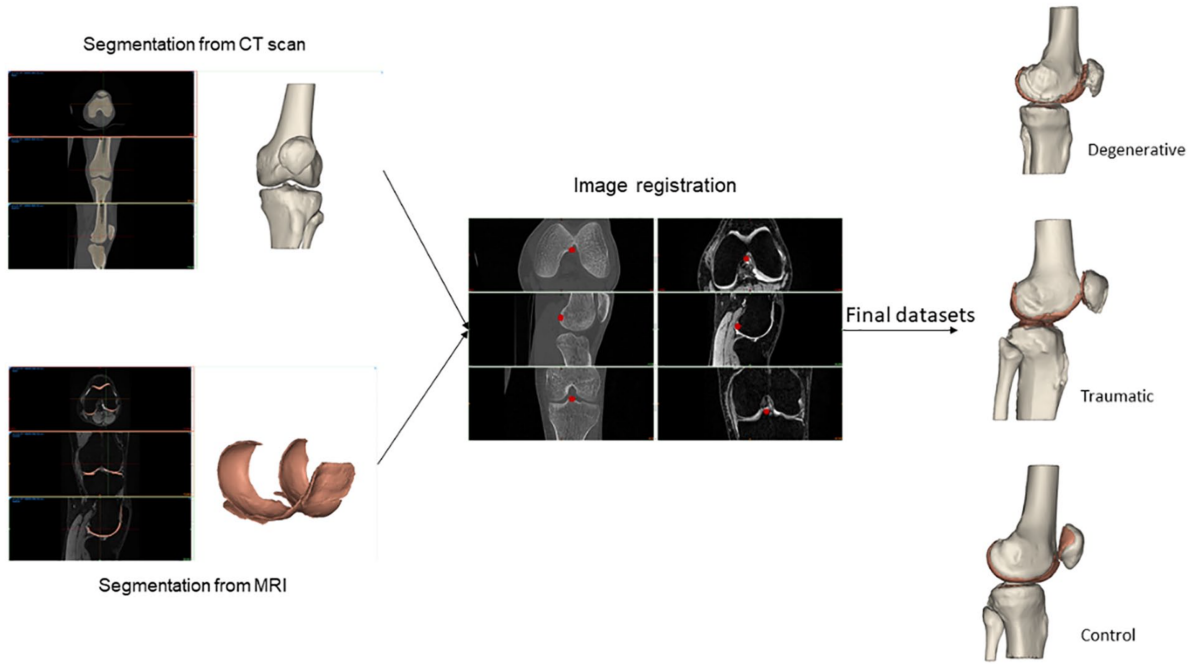


Figure 4.7: Segmentation workflow for 3D analysis

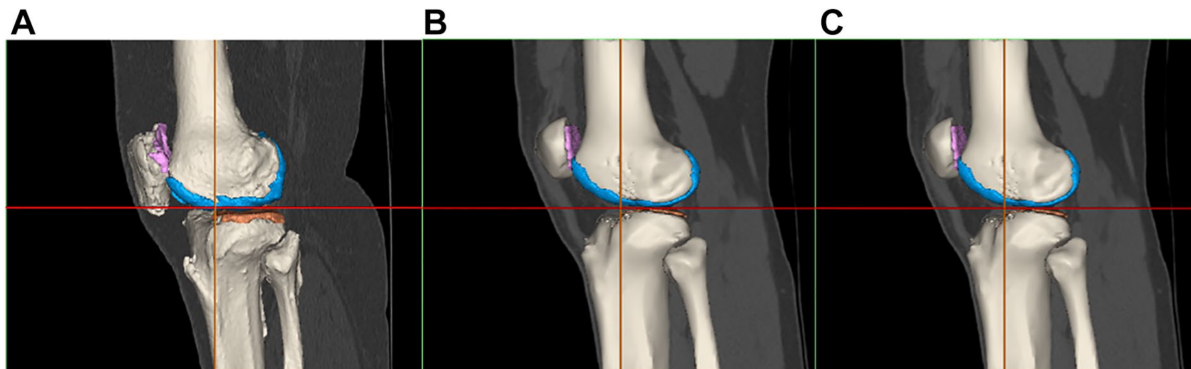


Figure 4.8: 3D model from the registration for the 3 groups of patients (A) Degenerative (B) traumatic (C) Control.

been considered as classification metrics.

Table 4.2 shows the features' sets used as input for the ML algorithms.

## 4.1.2 Results

### 4.1.2.1 2D and 3D Features

Table 4.3 shows the 2D features presence in the subjects as percentage for each group while the results shown in Table 4.4 display the average and standard deviation of the ICRS, cumulative index, and ACT for each group of patients.

The results shown in Table 4.5 display the average bone mineral density, as well as radiodensity, volume, and surface from each cartilage after tissue segmentation. The results were calculated for each group (D, T, and C). T group has the highest density of all the bones. D group and C group have similar values for the femur density, while D group has higher density in tibia, and lower in patella. D group has the lowest

Feat Selection	N of Feat	Source
Tot Feat	96	All the available features from MRI, CT, and 3D elaboration
2D Feat	78	Features from MRI (52 in total) and CT (26 in total)
3D Feat	18	Features of cartilage volume and density (and its standard deviation) from 3D elaboration
Ct - Scan Feat	26	Features from CT – part of the 2D group
MRI Feat	52	Features from MRI – part of the 2D group
Bone Feat	50	Features of Bone from CT and MRI (subchondral bone cysts, sclerosis and edema, osteophytes, osteonecrosis, and bone attrition)
Cartilage Feat	26	Features of Cartilage from MRI (ICRS grades, meniscal pathology, synovitis-effusion, and measurements of thickness)

Table 4.2: Feature Selection Sets used as inputs for the ML analysis

density in the femoral cartilage, the highest in the patella, and a slightly lower but similar values than C group in both lateral and medial tibia cartilage. D group has the highest cartilage volume for every part. C group has a higher volume than T group for the patella and the lateral tibia. D group has the highest cartilage surface for every part. T group has a higher surface than the C group for the patella and the medial tibia. It can be noted that for the patella, D group has the lowest bone density, the lowest cartilage density, the highest volume, and the highest surface. In general, results do not show a significant trend among patient groups due to high interpatient variability; however, patient-specific 3D measurements are used in the machine learning part to enlarge the set of features predicting the subjects' knee cartilage status.

#### 4.1.2.2 Machine Learning

Table 4.6 shows all the results of the ML analysis. The best accuracy value is 89.4, which is obtained with RF using the whole feature set and the 2D measurements feature set. F1 score is high for the degenerative patients (always around 90%), while it is slightly lower for the other 2 groups. These F1 scores are due to the higher number of degenerative patients but also demonstrate that with the selected features, it is efficient to classify patients with degenerative cartilage using RF and GB. The best metrics for the classification of control subjects are obtained with the 2D feature set, while all 96 features give the best F1 score for classifying traumatic patients using RF. In terms of accuracy, good results are obtained with MRI, bone, and cartilage feature selections, especially with the RF algorithm, while 3D selection is the worst, with a maximum of 76.6 accuracy with GB. With RF, the bone features give better results in all 3 classes compared to the cartilage feature set. Using GB and cartilage feature set, only control subjects are classified with higher metrics. We can state that RF is the most efficient of the 2 tree-based algorithms.

Tables 4.7 and 4.8 show respectively the 12 most important features and the percentage of importance of all the different feature groups for the RF classification ML model using the 96 total features as input. This model was selected for the feature importance analysis because it can be considered the most significant in terms of accuracy (89.4 is the highest). It allows for a complete overview of the full set of features and

Pathology	Compartment/ Bone	D	T	C
Subchondral bone cysts	Medial / Femur	34.78%	18.75%	0%
	Medial / Tibia	60.87%	6.25%	0%
	Lateral / Femur	17.39%	6.25%	0%
	Lateral / Tibia	26.09%	12.5%	0%
	Femoropat /Fem. Trochlea	26.09%	18.75%	12.5%
	Femoropat / Patella	30.43%	6.25%	12.5%
Osteophytes	Medial	91.3%	6.25%	12.5%
	Lateral	95.65%	0%	0%
	Femoropat	91.3%	25%	25%
Bone Attrition	Medial / Femur	30.43%	0%	0%
	Medial / Tibia	26.09%	0%	0%
	Lateral / Femur	0%	0%	0%
	Lateral / Tibia	0%	0%	0%
	Femoropat/Fem. Trochlea	0%	0%	0%
	Femoropat/Patella	0%	0%	0%
Osteonecrosis	Medial / Femur	4.35%	25%	0%
	Medial / Tibia	17.39%	12.5%	0%
	Lateral / Femur	21.74%	37.5%	12.5%
	Lateral / Tibia	21.74%	25%	12.5%
Subchondral bone sclerosis	Medial / Femur	86.96%	37.5%	12.5%
	Medial / Tibia	100%	97.35%	87.5%
	Lateral / Femur	34.78%	31.25%	25%
	Lateral / Tibia	21.74%	12.5%	0%
	Femoropat/Fem. Trochlea	13.04%	6.25%	0%
	Femoropat/Patella	69.57%	93.75%	75%
Subchondral bone edema	Medial / Femur	52.27%	31.25%	25%
	Medial / Tibia	65.22%	12.5%	0%
	Lateral / Femur	26.09%	68.75%	25%
	Lateral / Tibia	13.04%	6.25%	0%
	Femoropat/Fem. Trochlea	17.39%	50%	0%
	Femoropat/ Patella	4.35%	50%	0%
Meniscal pathology	Medial	100%	18.75%	0%
	Lateral	26.09%	37.5%	0%
Synovitis - Effusion		95.65%	100%	75%

Table 4.3: 2D measurements: Summary of bone pathologies, meniscal pathology and synovitis-effusion as percentages for each group of patients.



Pathology	Compartment/ Bone	D	T	C
		Avg(SD)	Avg (SD)	Avg (SD)
ICRS	Medial / Femur	3.26 (0.67)	1 (1.41)	0.63 (0.70)
	Medial / Tibia	3.17 (1.31)	1.13 (1.05)	0 (0)
	Lateral / Femur	1.96 (1.37)	1.63 (1.11)	1.38 (0.70)
	Lateral / Tibia	2.22 (1.41)	1.75 (1.25)	0.75 (0.83)
	Femoropat/Fem. Trochlea	2.36 (1.37)	0.94 (1.09)	0.75 (1.09)
	Femoropat/ Patella	3.09 (0.79)	2.69 (0.58)	1.75 (0.66)
Cumulative Index	Femur	3.83 (1.19)	2.38 (0.96)	1.13 (0.83)
	Tibia	4.09 (1.24)	1.69 (0.79)	1.25 (0.46)
	Patella	2.04 (0.93)	1.75 (0.93)	1.13 (0.83)
Articular Cartilage Thickness	Medial / Femur	1.5 (0.71)	2.63 (0.56)	2.64 (0.67)
	Medial / Tibia	1.63 (0.76)	2.3 (0.45)	2.48 (0.42)
	Lateral / Femur	2.44 (0.63)	2.7 (0.54)	2.9 (0.38)
	Lateral / Tibia	2.4 (0.81)	2.9 (0.73)	3.06 (0.65)
	Femoropat/Fem. Trochlea	2.21 (0.77)	1.97 (0.77)	2.31 (0.29)
	Femoropat/Patella	2.29 (0.61)	2.78 (0.7)	2.7 (0.32)

Table 4.4: 2D measurements: Average values of the ICRS grading, CI and ACT for each group (and standard deviation between parentheses) and their location (compartment/bone).

	Degenerative	Traumatic	Control
<b>Bone Mineral Density (g/cm3)</b>			
Femur bone	1,32 (1,13)	1,33 (1,14)	1,32 (1,12)
Tibia bone	1,32 (1,13)	1,35 (1,15)	1,29 (1,16)
Patella bone	1,36 (1,12)	1,40 (1,14)	1,41 (1,11)
<b>Radiodensity (HU)</b>			
Femur cartilage	85,19 (57,47)	88,67 (49,90)	93,53 (54,37)
Lateral Tibia cartilage	87,84 (51,10)	88,69 (44,97)	91,19 (49,21)
Medial Tibia cartilage	98,49 (55,92)	93,63 (44,14)	103,79 (52,82)
Patella cartilage	78,36 (50,68)	81,56 (44,97)	99,09 (55,45)
<b>Volume (mm3)</b>			
Femur cartilage	17303 (5530)	12460 (2710)	11276 (4505)
Lateral Tibia cartilage	2851 (2336)	1100 (439)	1501 (1927)
Medial Tibia cartilage	1915 (1638)	907 (566)	552 (362)
Patella cartilage	2761 (830)	2589 (781)	2703 (705)
<b>Surface (mm2)</b>			
Femur cartilage	14381 (2636)	12610 (1496)	11809 (2791)
Lateral Tibia cartilage	2435 (1503)	1415 (499)	2073 (2530)
Medial Tibia cartilage	2016 (1367)	1301 (550)	967 (407)
Patella cartilage	2602 (760)	2574 (488)	2495 (390)

Table 4.5: 3D measurements: The results show the average variable for each group (with standard deviation between parentheses)

Feat Select	Alg.	Acc.	Re D	Pr D	F1 D	Re T	Pr T	F1 T	Re C	Pr C	F1 C
Tot [96]	RF	89.4	95.9	92.0	93.9	93.3	82.4	87.5	62.5	100	76.9
	GB	87.2	91.7	95.7	93.6	93.3	73.7	82.4	62.5	100	76.9
2D [78]	RF	89.4	91.7	91.7	91.7	86.7	86.7	86.7	87.5	87.5	87.5
	GB	87.2	91.7	91.7	91.7	80.0	80.0	80.0	87.5	87.5	87.5
3D [18]	RF	74.5	83.3	83.3	83.3	66.7	66.7	66.7	62.5	62.5	62.5
	GB	76.6	87.5	84.0	85.7	66.7	76.9	58.8	62.5	55.6	58.8
CT-Scat [26]	RF	80.9	91.7	95.7	93.6	66.7	71.4	69.0	75.0	60.0	66.7
	GB	74.5	87.5	84.0	85.7	60.0	75.0	66.7	62.5	50.0	55.6
MRI [52]	RF	87.2	95.8	95.8	95.8	87.6	76.5	81.2	62.5	83.3	71.4
	GB	87.2	91.7	88.0	89.8	80.0	85.7	82.8	87.5	87.5	87.5
Bone [50]	RF	85.1	91.7	91.7	91.7	86.7	76.5	81.2	62.5	83.3	71.4
	GB	76.6	87.5	95.5	91.3	73.3	64.7	68.8	50.0	50.0	50.0
Cart [26]	RF	83.0	91.7	88.0	89.8	73.3	73.3	73.3	75.0	85.7	80.0
	GB	83.0	91.7	88.0	89.8	67.7	76.9	71.4	87.5	77.8	82.4

Table 4.6: classification metrics (Recall (Re) Precision (Pr) and F1 [%]) for the 2 different tree-based ML algorithms and the seven different features selections (Degenerative (D) - Traumatic (T) - Control (C))

TibCartLatVOL [mm3]	4,804
TibCartMedVOL [mm3]	4,631
CT Lat Osteophytes	4,594
MRI Med Cart Thick FEM [mm] - Med	4,262
MRI Med Menisc Pathol	3,909
MRI Lat Osteophytes	3,805
FemCartVOL [mm3]	3,644
TibCartLatSTD	2,543
MRI Lat Cart Thick FEM [mm] - Ant	2,491
MRI Med Cart Thick FEM [mm] - Post	2,402
MRI Lat Cart Thick TIB [mm] - Med	2,352
CT Med Osteophytes	2,343

Table 4.7: 12 most important features [%] for the RF classification model with 96 Tot Features

their respective importance in the classification process. The highest importance is attributed to 2 features from the 3D collection (the volume of the tibialis cartilage lateralis and medialis). The 3D features set contribute 33% of importance despite being only 18 compared to the 78 2D features. The cartilage set of features has higher importance compared to the bones. In contrast, the CT scan features contribute to only 14.39% of the importance having only 2 of them in the first 12 most important features (lateral and medium osteophytes).

### 4.1.3 Discussion

This work developed a methodology to evaluate cartilage degeneration: it uses a multimodal image approach and 3D models to segment bones and cartilages from the knee area. Indeed, the MRI provides information about pathologies, morphology of the car-

2D	66,08%
MRI (part of 2D features)	51,69%
CT (part of 2D features)	14,39%
3D	33,12%
BONE (from CT and MRI)	28,29%
CARTILAGE (from MRI)	37,79%

Table 4.8: Importance of the groups of features [%] for the RF classification model with 96 Tot Features

tilages, and a geometric representation of the tissue damage, while CT data present a good overview of bone pathologies, especially in boundary regions. The combination of both imaging techniques gives a 3D representation of the knee, and additional information about bone and cartilage. These data overview makes the definition of 96 features possible, which demonstrated various levels of significance with regard to contribution toward the cartilage quality evaluation.

ML results of Table 4.6 underline significant recall and precision, as well as F1, especially on the classification of degenerative patients, reaching a maximum recall value of 95.9 using the total and MRI feature selections, having almost 90% accuracy. The use of all the 96 features gives the best classification metrics and allows a complete feature importance analysis which gives new significant hints for studying the degeneration condition of the knee cartilage.

Noteworthy are the results obtained with the single Bone and Cartilage feature selections. While a good accuracy value is expected for the latter as we are classifying subjects relative to their cartilage status, the classification metrics obtained with the bone selection are of high impact. Cartilage status is highly dependent on the bone's condition, as has been demonstrated by Cai et al., [269] which observed changes in the subchondral bone with OA progression. Moreover, Bonakdari et al. [270] recently used bone features to predict cartilage volume loss obtaining a correlation coefficient of more than 0.78.

Similarly, we demonstrated that bone has high importance in the classification process. Still, if combined with cartilage and 3D features, the metrics significantly increase, indicating that with the contribution of all these sets of features, a more in-depth view of the knee cartilage status can be given. If we consider 3D features or cartilage feature sets alone for the classification process, the metrics are not significant due also to the limited number. But they assume a significant relevance if we consider the whole complete set (the 2 most relevant features, volume of the tibialis cartilage lateralis and medialis, are from the novel 3D group). 3D features contribute one-third of the importance despite the limited number. At the same time, they give the lowest accuracy of 74.5 if considered the only input to the tree-based algorithms. The 26 cartilage features alone can give a decent 83% accuracy but, if taken together with the other 70, contribute to the classification for almost 40% of total importance. We can conclude that the complete set of features gives the best input for future developments of this study: all the 96 bone, cartilage, and 3D features together could be used to develop new clinical solutions like the design of a patient-specific cartilage status profile which will help the clinicians and the researchers in an easier and objective classification of the cartilage status and an evaluation of the degeneration level. This novel methodology, combining 2D and 3D measurements, is of interest to assess cartilage quality.

By designing indexes of pathology and combining it with other parameters such as radiodensity, it is possible to categorize cartilage into a group condition. This study should be pursued with a larger range of subjects to ensure its efficiency.

ML analysis presents some limitations. The number of subjects is not particularly high; this could affect the classification performances and a partial overfitting may occur in some models. Moreover, the multi-class approach can significantly decrease the classification metrics: for future work, a binary classification option can be performed to study the prediction potential of the features to distinguish degenerative patients from all the others. A higher number of control subjects could also potentially be the starting point for a binary classification between degenerative vs healthy or traumatic vs healthy subjects.

In conclusion, we developed a cartilage segmentation and 3D modeling procedure that can be used as benchmark for 3D bioprinting design and to advance cartilage assessment in the frame of the RESTORE project. Based on a cumulative index of bone properties (CI), we demonstrate the importance of bone condition and the sensitivity of these measurements on medial and femoropatellar compartments. Moreover, we show that a combination of 2D radiological measurements and 3D measurements revealed potential biomarkers of cartilage degeneration, especially from medial femur.

This work is a first step toward a patient-specific cartilage profile based on the combination of CT and MRI datasets. This could be crucial for improving cartilage assessment. Indeed, when evaluating patients with knee pain either following trauma or with acute or chronic illness, the patient's symptoms are always the cornerstone in the treatment decision, whether medical or surgical. Following plain x-ray, a CT scan and most often also MR scan are the best tools in elucidating the interior of the knee joint. The CT scan is both easy to get and fast to execute but uses ionizing radiation. It reveals, however, best all the bony structures and injuries. It may also give some clues about the bone marrow and surrounding soft tissues. The MR, however, is the best examination to evaluate the status of both the cartilage and the ligaments. The drawback is both the long time until it can be executed and long running time which can sometimes be impossible in patients with severe pain. When merged, these 2 examinations give the most superior evaluation ever for the knee joint and should always be chosen prior to invasive arthroscopy. Our study shows the feasibility of extending the cartilage assessment using existing and new parameters from both image modalities.

# Chapter 5

## Motions Sickness and Postural Control Study

The following chapter of the thesis explores the study of Postural Control and Motion Sickness begun in summer 2020 with the creation of the Motion Sickness Lab in the Engineering Department of Reykjavik University. BioVRSea system led to several publications relative to the prediction and assessment of MS using multiple bio-metric features from Electroencephalography (EEG), Electromyography (EMG), Center of Pressure (CoP), Heart Rate (HR) combined with data from the Motion Sickness Susceptibility Questionnaire (MSSQ).

The three main researches presented in the present chapter will focus on three different cohorts of healthy subjects who underwent the BioVRSea experiment. They focus on motion sickness prediction, concussion on female athletes and lifestyle influence on MS symptomatology.

### 5.1 Introduction

#### 5.1.1 BioVRSea definition

BioVRSea is a bio-measurement system established in 2020 in the Motion Sickness Lab at Department of Engineering of the Reykjavik University to study PC and MS. This system is a sailing simulator that records, in synchronized fashion, multiple bio-signals from brain, heart and muscle. The participants wear the VR goggles showing a rough sea scenario. The movement of the ship over the waves in the VR scenario is coupled to the moving platform and the frequency and amplitude of the VR wave motion is synchronized with the platform motion. Different "sea sounds" can be heard through the headphones of the VR goggles. Subjective and objective MS levels are assessed by a questionnaire while biosensors measure EEG, electromyography (EMG), and HR of the subject. The creation of a database allows the implementation of various statistical and ML algorithms with the aim of correlating the biometric results with new indexes that combine the various symptoms of MS, having as main novelty the EEG application and interpretation in association with VR and moving platform inducing MS, linked to other biosignals.

The VR software (Virtualis, VR, France) dynamically visualizes a virtual environment as if the subject is out on the open sea on a little boat. A moving platform (Virtualis VR, France) mimics the waves according to the simulated environ-



Figure 5.1: A subject on the BioVRSea platform wearing the VR goggles during the sea virtual simulation in the Motion Sickness Lab at Reykjavik University

ment. The operator can set frequency (between 0.5 and 3 Hz) and amplitude of the waves (from 0 to 2). During the simulation, we vary the amplitude of the platform movements from 0% up to 100%. The platform allows fast (tailored) movements in  $0^\circ$ ,  $-45^\circ$ ,  $-90^\circ$ ,  $-135^\circ$ ,  $-180^\circ$ ,  $-225^\circ$ ,  $-270^\circ$ ,  $-315^\circ$ ,  $-360^\circ$  (linear acc. not available) coupled to synchronized visual VR movements.

Of course, the VR view that the investigated individual visualizes standing on a virtual small vessel is not a true scenario of working environment at sea but is nevertheless capable of creating real MS sensation at least in experienced sailors (verbal statements after being on platform).

Fig. 5.1 shows a subject on the BioVrSea platform during the experiment.

### 5.1.2 BioVRSea Acquisition's Protocol

The whole acquisition, including the preparation before the simulation, can last around 40-50 minutes. The experiment itself has a duration of 320 seconds and it is split in 6 different sections (Fig. 5.2).

The first section is the baseline segment, where the participant only sees a static mountain panorama through the VR goggles. During this time, the platform does not move, and the participant stands with the hands by the side while observing a virtual mountain scene for a total of 120 s. After the Baseline section, the VR scene changes to the sea environment and the subject sees her/himself on a small boat at rough sea. The second section lasts 40 s and it is denominated PRE section. During this phase, the platform does not move, and the participant remains still with hands by the side. In the following third section, the participant holds on to the safety bars and the platform

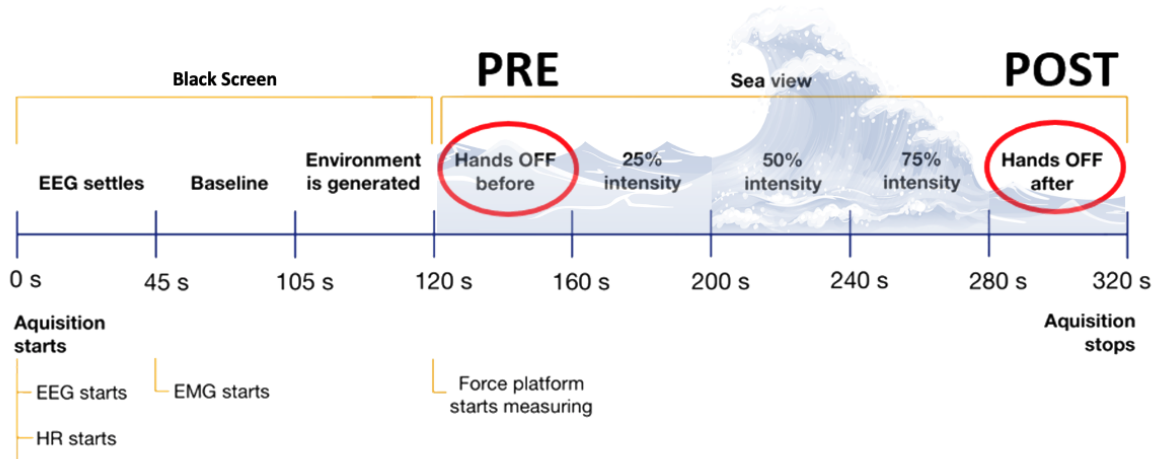


Figure 5.2: BioVrSea Acquisition's Protocol

moves in a synchronized manner with the waves seen in the VR environment. During these 40 s, the platform moves at 25% of the maximum amplitude. In the fourth and fifth section, the platform increases the intensity to 50% and 75%, respectively for 40 s each. In the last section, denominated POST section, the movement of the platform stops while the VR simulation continues. The participant takes their hands off the safety bar and stands quietly trying to maintain equilibrium watching the VR sea scene. After 40 s, this section is over and so is the entire acquisition. The eyes must be open during all the experiment.

Table 5.1 summarizes the VR/platform synchronization protocol.

Time (s)	Section	VR Scene	Hands' Position	Platform
0-120	Baseline	Mountain	By Side	Stationary
120-160	PRE	Sea	By Side	Stationary
160-200	25%	Sea	On Bars	Moving
200-240	50%	Sea	On Bars	Moving
240-280	75%	Sea	On Bars	Moving
280-320	POST	Sea	By Side	Stationary

Table 5.1: VR/platform synchronization protocol

Concerning the waves' frequency and amplitude, two protocols are randomly selected for the experiment:

- Soft: Frequency: 0.1 Hz - Amplitude: 0.6
- Hard: Frequency: 0.3 Hz - Amplitude: 0.5

The selection of these frequencies was based on two main reasons. The first reason is to only act upon one of MS etiologic theory: multiple theories have been listed to explain MS, and the SCT is easily the leading perspective. Frequencies below 1 Hz are not considered because they might act upon the additional Postural Instability

Theory, which is rooted in perception of lower  $<0.5$  Hz frequencies [205]. The second reason is to ensure that an easy scenario (1 Hz) is available to reduce the risk of falling, as well as a harder one (3 Hz) to ensure sufficient movement to trigger MS.

### 5.1.3 Biomedical Data Acquisition



Figure 5.3: BioVRSea Biosignals setup: EEG wet cap and EMG sensors, gastrocnemius lateral, and soleus muscles.

- Brain electrical activity is measured with wet 64-electrode EEG cap (Sampling frequency 4096 Hz, ANTNeuro, Hengelo, The Netherlands) (Fig. 5.3 shows the conductivity gel injection process).
- Muscle electrical activities from the lower limbs is acquired using six wireless EMG sensors (sampling frequency of 1600 Hz) placed on the tibialis anterior (TA), gastrocnemius lateral (GL), and soleus (S) muscles of each leg (Kiso ehf, Reykjavik, Iceland) (Fig. 5.3 shows the application of EMG wireless sensors on GL, and S muscles).
- Heart rate is measured using a chest heart sensor (Polar Electro, Kempele, Finland, sampling frequency 1000 Hz).



- Force Plate measurements are made using 4 sensors located under each foot platform. The sensors give information about the center of mass in the Antero-posterior and Medio-Lateral axis (Virtualis, Clapiers, France, sampling frequency 90 Hz) (Fig. 5.4).

Biomedical-data detailed process description goes beyond the scope of the present thesis and therefore it will not be explained in depth here. More exhaustive information can be found in the BioVRSea current literature [5][6][7][271][272][273][274].



Figure 5.4: Force Plates and VR goggles on the moving platform of BioVRSea

#### 5.1.4 BioVRSea cohorts

From 2020 almost 400 people underwent the BioVRSea experiment and acquisitions are still ongoing at the time of this text drafting. Three different cohorts are considered in this thesis:

- 28 subjects (age:  $23.8 \pm 1.2$ ), 22 women and 6 men. Each participant is measured with both soft (1Hz) and hard (3Hz) protocols as well as with the baseline (denominated in this case 0Hz protocol). This initial cohort is considered in [5] and the relative results are detailed in section 5.2;
- Participants are all female athletes ( $N = 54$ ), competing at the highest level in Iceland in basketball (16.7%), handball (35.2%), soccer (38.9%), ice hockey (5.6%), or martial arts (3.7%). Mean age is 38.4 (SD = 7.7). Mean years since retirement was 4.3 years (SD = 4.9). This second cohort is used for the Concussion study using BioVRSea published in [6] and described in detail in section 5.3;
- The largest cohort considered for the researches this thesis is composed of total of 262 subjects (age  $36.18 \pm 15.66$ ), 152 women and 110 men. It is used to study

lifestyle influence on MS and BioVRSea paradigms: the results were presented at IEEE MetroXRaine Conference in Rome [7] and are detailed in section 5.4.

At the time of writing of this dissertation 355 healthy subjects were measured, plus 11 Early-Parkinson subjects. For the indexes introduced in the following paragraph 5.1.6, all the various numbers refer only to the healthy population of 355 subjects.

### 5.1.5 Motion Sickness Questionnaire

Each participant before and after the acquisition answers a Motion Sickness Questionnaire inspired by the Motion Sickness Susceptibility Questionnaire (MSSQ) proposed by Golding [275], considered a gold standard in the evaluation of symptoms and pronesness relative to MS.

General info like age, gender, weight, and height are initially asked. Consequently Body Mass Index is calculated from weight and height of each subject.

Then, information about everyday life are asked: physical activity, intake of food, caffeine, nicotine, and alcohol.

Numbers relative to the age and gender of the 355 subjects are shown in the pie charts of Fig. 5.5.

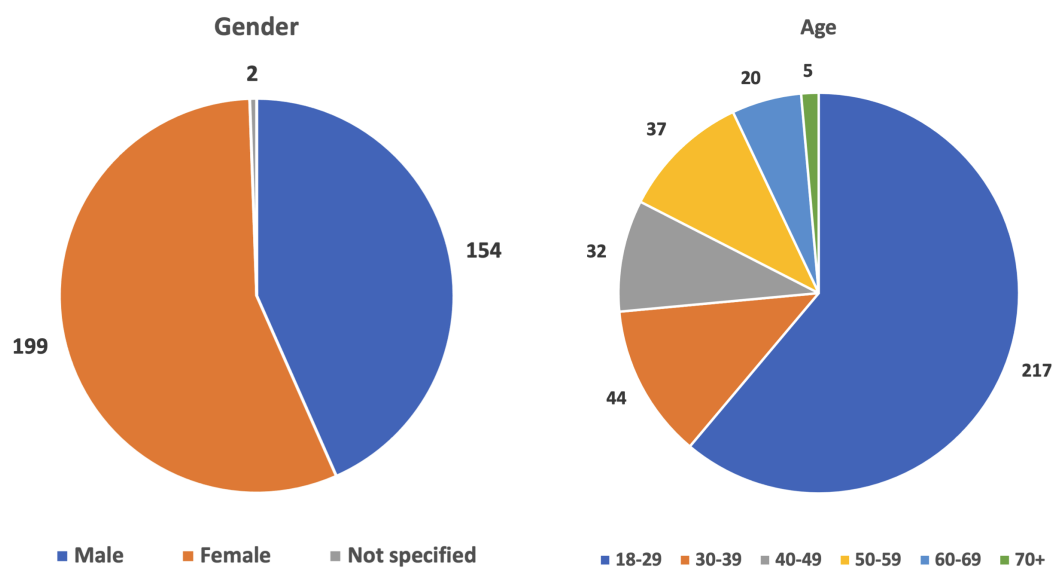


Figure 5.5: Age and Gender piecharts relative to MSSQ answers from 355 subjects

Successively the questionnaire focus on MS susceptibility and predisposition. Participants are asked about their past experiences relative to experience MS on transport and entertainment systems on a scale from 0 to 3 (Never, Rarely, Sometimes, Often felt sick). The transport and entertainment systems are the following:

- Car
- Bus
- Train

- Aircraft
- Small Boat
- Ship
- Swing in playgrounds
- Roundabout in playgrounds
- Funfair ride, big dipper

A fifth option is available in case the subject has never travelled or experienced one of the previous listed systems. In case the subject stated that she/he has never travelled on cars, busses or airplanes, the answers has been considered not valid and substitute with never felt sick. If the participant stated that she/he has never travelled on trains, the answer is considered valid as no railways are present in Iceland and most of the participants are from Iceland.

Afterwards, before to start the virtual experience, the subject will answer about how does she/he feels at the moment rating on a scale from 0 to 3 (None, Slight, Moderate, Severe) the following typical MS symptoms:

- General Discomfort
- Fatigue
- Headache
- Fullness of Head
- Eye Strain
- Difficulty focusing or concentrating
- Increased salivation
- Burping
- Stomach Awareness
- Sweating
- Nausea
- Blurred Vision
- Vertigo or Dizziness

The participant self-evaluate the same symptoms at the end of the sea simulation after the POST segment of the experiment.

The first cohort of participants were able to rate their symptoms on scale from 0 to 2.

### 5.1.6 Motion Sickness and Lifestyle Indexes

Using the data extracted from the MSSQ multiple different indexes are computed relative to lifestyle habits, MS proneness and symptoms.

**Lifestyle Index** The Lifestyle Index (LSI) is a three classes index based on BMI, Sport Activity, Nicotine, Caffeine, and Alcohol assumption.

Participants are assigned 1 point if the behavior is considered healthy, and no points otherwise following these next criteria:

- BMI: Unhealthy [0] if  $\geq 25$  - Healthy [1] if  $< 25$
- SPORT Activity: Unhealthy [0] if  $< 3$  times a week - Healthy [1] if  $\geq$  three times a week
- Nicotine (on daily basis): Unhealthy [0] if YES - Healthy [1] if NO
- Caffeine (today): Unhealthy [0] if YES - Healthy [1] if NO
- Alcohol: Unhealthy [0] if Today or Yesterday - Healthy [1] if NO or more than two days ago

The summation (LS\_Sum) of the points is then computed and the three classes LSI is defined as follow:

- RED if LS\_Sum=0 or 1
- YELLOW if LS\_Sum=2 or 3
- GREEN if LS\_Sum=4 or 5

The results relative to the three classes LSI from the 355 subjects are shown in the piechart of Fig. 5.6.

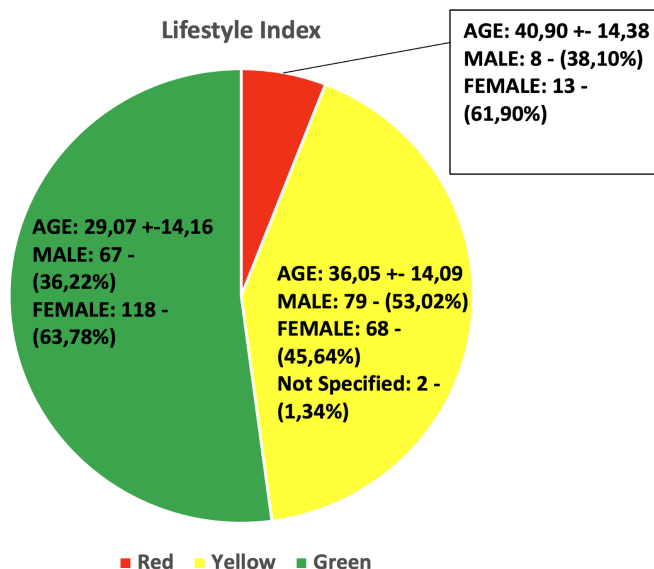


Figure 5.6: Three classes LSI piechart from 355 subjects

A specific Lifestyle Index is developed for the study of paragraph 5.4. It takes in consideration for each patient the BMI, smoking status and physical activities excluding alcohol and caffeine info. A person is considered healthy if two of those characteristics are considered healthy. On the total of 262 participants of the third cohort,

196 has been considered with an healthy lifestyle while the remaining 66 do not have a healthy lifestyle.

**Motion Sickness Proneness Index** The Motion Sickness Proneness Index (MSPI) is a binary index based on the answers relative to MS susceptibility and predisposition on transport and entertainment systems. For each category a point from 0 to 3 is assigned. The summation of these points is computed and divided to the answer given (maximum of 9 categories): the results is defined as (MSP\_Sum).

The MSPI classes are defined as follow and relative percentages are shown in Fig. 5.7:

- MS Prone [1] if  $MSP\_Sum \geq 9$
- Not MS Prone [0] otherwise

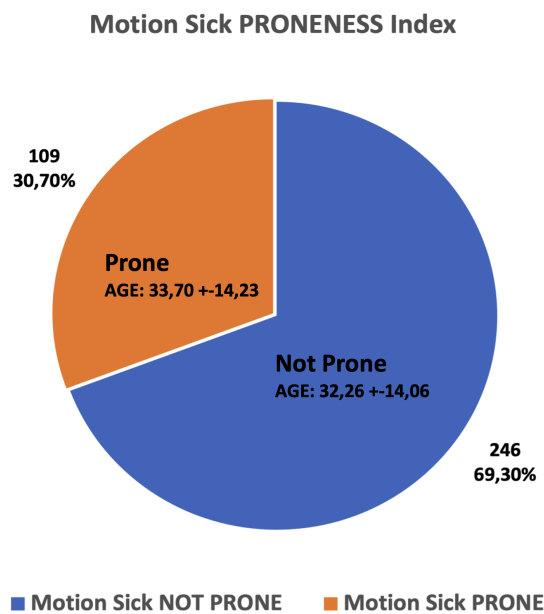


Figure 5.7: MSPI piechart from 355 subjects

**Motion Sickness Index** The Motion Sickness Index (IMS) is a binary index that refers to the 13 MS symptoms listed previously. These symptoms are grouped as follow and computed for the PRE (before the simulation) and the POST (after the simulation). Binary indexes referred to these groups of symptoms are created following these steps: first, we compute the average from the individual responses of each index; second, we calculate the maximum among the averages; and third, we divide the cohort into two groups (below and above 1/3 of the maximum). For IGenDis and IDizz, we apply only steps 2 and 3 using the direct response instead of the average. The symptoms groups are the following:

- Stomach Index (Isto): Avg [Salivation, Sweating, Nausea, Stomach Awareness, Burping]
- Fatigue Index (Ifatig): Avg [Fatigue, Eye Strain, Diff. Focusing]

- Head Index (Ihead): Avg [Headache, Blurred Vision, Fullness of Head]
- General Discomfort (IGenDis): includes only General Discomfort
- Dizziness-Vertigo (IDizz): includes only Dizziness-Vertigo

Moreover, we established two more indexes: Physiological/Vegetative Index (IPV) and Neurological/Muscle Strain Index (INM). IPV is based on the responses from sweating, salivation, nausea, burping, stomach awareness, and general discomfort conditions. Similarly, the INM is based on fatigue, eye strain, difficulty focusing, headache, fullness of head, blurred vision, and again general discomfort conditions. IMS is here defined as the weighted sum (SumMS) of all the MSSQ answers (Eq. 5.1):

$$\begin{aligned}
 SumMS = & (0.2 * GenDisc + 0.2 * Dizz\&Vert + \\
 & + 0.2 * \sum (StomAwe, Nausea, Sweat, Saliv, Burp) + \\
 & + 0.2 * \sum (Fatigue, EyeSt, DiffFocus) + \\
 & + 0.2 * \sum (Headache, FullHead, BlurrVis)) \quad (5.1)
 \end{aligned}$$

**BioVRSea Effect Index** For each symptom group defined before the difference between POST and PRE is computed (if negative is considered 0). If the difference is  $>0$  the Symptom Group Binary Difference (SG\_BinDiff) is =1, otherwise SG\_BinDiff=0. SG\_BinDiff is computed for each of the five groups and then summed up creating MS\_Diff\_Sum. If MS\_Diff\_Sum=1 it means that 1 symptom group has increased from before (PRE) the simulation to after (POST) it. If MS\_Diff\_Sum=5 it means that all the symptom group has increased, if MS\_Diff\_Sum=0, there are no changes in the MS symptoms after the simulation.

BVSEI is binary defined splitting the population in two considering MS\_Diff\_Sum=1. BVSEI differs people who had a single change in the symptoms (or more than one) between before and after the experiment, from people who did not suffer any symptoms.

Out of the 355 subjects, as shown in Fig. 5.8, 233 (65.63%) did have a symptom effect on the BioVRSea experiment (BVSEI=1), 122 (34.37%) did not have any symptom changes (BVSEI=0).

The following tables show the percentages of the 355 subjects when combining the three main indexes: table 5.2 shows the percentages and relative gender and age information related to LSI and MSPI, table 5.3 do the same in relation to LSI and BVSEI, and finally table 5.4 shows the information relative to MSPI and BVSEI.

### 5.1.7 Sport Concussion Assessment Questionnaire

Participants of the second (concussion) cohort, were read before the BioVRSea acquisition a concussion definition and were asked if they had sustained a concussion. The definition was based on the Berlin Consensus statement on concussion in sport from 2016. All participants completed the symptoms scale from The Sport Concussion Assessment Tool 5 (SCAT5) before the experiment. The overall score is calculated by

	<b>PRONE</b>	<b>NOT PRONE</b>
<b>RED</b>	<b>1.97%</b>	<b>3.94%</b>
	AGE: 40,43 +- 14,82 Female: 7 - (100%) Male: 0 - (0,00%)	AGE: 41,14 +- 14,41 Female: 6 - (42,86%) Male: 8 - (57,14%)
<b>YELLOW</b>	<b>12.96%</b>	<b>29.01%</b>
	AGE: 37,74 +- 14,12 Female: 29 - (63,04%) Male: 16 - (34,78%) NotSpec: 1 (2,17%)	AGE: 35,29 +- 14,07 Female: 38 - (36,89%) Male: 64 - (62,14%) NotSpec: 1 (0,97%)
<b>GREEN</b>	<b>15,77%</b>	<b>36,34%</b>
	AGE: 29,54 +- 14,27 Female: 48 - (85,71%) Male: 8 - (14,29%)	AGE: 28,87 +- 14,18 Female: 70 - (54,26%) Male: 59 - (45,74%)

Table 5.2: Percentages of subjects related to Lifestyle Index and Motion Sickness Proneness Index

	<b>BioVRSea Influenced</b>	<b>BioVRSea Not Influenced</b>
<b>RED</b>	<b>3.94%</b>	<b>1.97%</b>
	AGE: 40,50 +- 14,82 Female: 11 - (78,57%) Male: 3 - (21,43%)	AGE: 41,71 +- 14,38 Female: 2 - (28,57%) Male: 5 - (71,43%)
<b>YELLOW</b>	<b>26.19%</b>	<b>17.77%</b>
	AGE: 32,82 +- 14,08 Female: 48 - (51,61%) Male: 44 - (47,31%) Not Spec: 1 (1,08%)	AGE: 41,07 +- 14,16 Female: 20 - (35,71%) Male: 35 - (62,50%) Not Spec: 1 (1,79%)
<b>GREEN</b>	<b>35.49%</b>	<b>16.61%</b>
	AGE: 28,54 +- 14,21 Female: 88 - (69,84%) Male: 38 - (30,16%)	AGE: 29,70 +- 14,18 Female: 30 - (50,85%) Male: 29 - (49,15%)

Table 5.3: Percentages of subjects related to Lifestyle Index and BioVRSea Effect Index

	<b>BioVRSea Influenced</b>	<b>BioVRSea Not Influenced</b>
<b>Prone</b>	<b>25.35%</b>	<b>40.29%</b>
	AGE: 32,01 +- 14,37 Female: 74 - (82,22%) Male: 15 - (16,67%) Not Spec: 1 - (1,11%)	AGE: 30,30 +- 14,09 Female: 73 - (51,05%) Male: 70 - (48,95%)
<b>Not Prone</b>	<b>5.35%</b>	<b>29.01%</b>
	AGE: 39,41 +- 14,38 Female: 11 - (57,89%) Male: 8 - (42,11%)	AGE: 34,62 +- 14,06 Female: 41 - (39,81%) Male: 61 - (59,22%) Not Spec: 1 - (0,98%)

Table 5.4: Percentages of subjects related to Motion Sickness Proneness Index and BioVRSea Effect Index

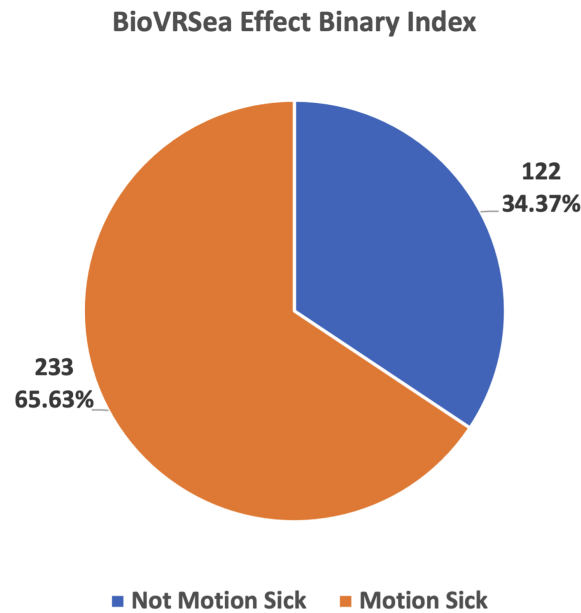


Figure 5.8: BVSEI piechart from 355 subjects

the sum of each participant's responses. The scale has 22 items, each item scoring from 0 to 6, indicating the severity of the symptom. In our study, it was hypothesized that the Icelandic versions of the SCAT5 symptoms checklist, could be used to differentiate between concussed and non-concussed athletes, it was used to validate the self-reported concussion status and assess for each group the changes of some physiological conditions associated with our experiment.

Almost half of the participants had a history of concussion/s, 48.1% ( $n = 26$ ), the remaining 51.9% ( $n = 28$ ) reported no concussion history.

## 5.2 Motion Sickness prediction with first cohort

The present section reports the results from the first significant publication relative to the BioVRSea MS studies. It is adapted from the paper published on *Frontiers in Bioengineering and Biotechnology* [5].

Fig. 5.9 shows the BioVRSea set-up for the present study.

### 5.2.1 BioVRSea Setup for first cohort study

#### 5.2.1.1 Biomedical Features

Fig. 5.10 lists and describes the 19 biometric parameters for each acquisition protocol. For the acquisition process please refer to paragraph 5.1.3 with the exception of EEG measurement: brain electrical activity is here acquired using a 64-channel dry electrode cap (sampling frequency of 500 Hz) from AntNeuro, Hengelo, Netherlands.

Fig. 5.11 shows the objective physiological measurement differences for all the subjects between the first static protocol and the other two, the light one (1 Hz) in green and the hard one (3 Hz) in red. The arrows show how the values of the single EEG, EMG, and HR data rise or fall during the protocols. For example, it is possible to notice how



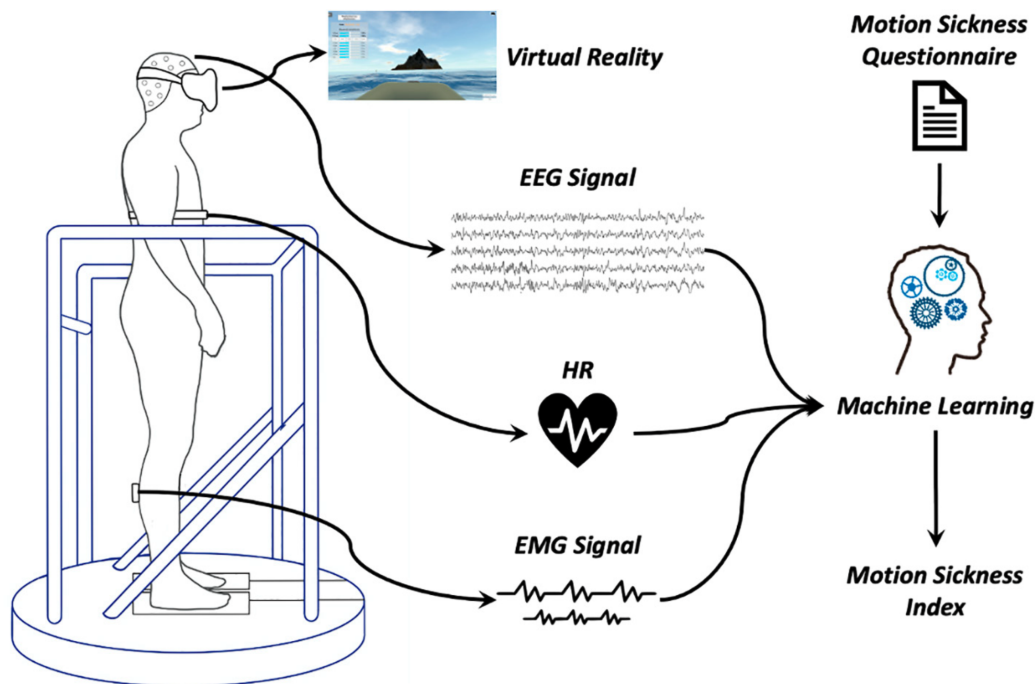


Figure 5.9: BioVRSea structure for the present study: the moving platform, is combined with a rough sea VR scenario and with EEG, EMG, and HR bio-signal acquisition to predict MS using ML.

the EMG values for both legs at low frequencies increase for the most patients, while they decrease at high frequencies. On the opposite, the EEG values do not follow such a regular trend.

In Fig. 5.12, it is possible to see the percentage of MSSQ answers and indexes for the entire cohort. It is possible to identify general discomfort, sweating, nausea, and vertigo as the most significant indexes with over 20% of responses being the highest possible value. Salivation and burping, conversely, are the least significant with a percentage lower than 5% providing a response of highest value. Fig. 5.13, similarly to Fig. 5.11, shows the increase or decrease of the value of the subjective given answer to the questionnaire using the colored arrows. It is possible to see that some patients, like numbers 10, 11, 14, and 22, have an increase of the symptom for both the 1-Hz and the 3-Hz protocols while others do not show any significant difference. Subject number 16 shows an increase only with the 3-Hz protocol, confirming the strong influence of the wave frequency on the body.

### 5.2.1.2 Statistical Analysis and ML tools

All the parameters extracted from EEG, EMG, and HR underwent a non-parametric statistical univariate explorative analysis in order to understand whether there was a statistically different grouping by IGenDis, IDizz, IStom, IHead, IFatig, IPV, INM, and IMS. All the indexes underwent univariate statistical analysis through the Mann-Whitney test.

The following algorithms were implemented: Random Forests (RF), Gradient Boosting tree (GB), Ada-Boosting of decision tree (ADA-B), Support Vector Machine (SVM), K Nearest Neighbor (KNN), and Multilayer Perceptron (MLP).

<b>Biometric parameter</b>	<b>Description</b>
EEG – Delta	Relative power spectra between frequency band 0.5–4 Hz
EEG – Theta	Relative power spectra between frequency band 4–8 Hz
EEG – Alpha	Relative power spectra between frequency band 8–13 Hz
EEG – Beta	Relative power spectra between frequency band 13–35 Hz
EEG – LG	Relative power spectra between frequency band 35–40 Hz
EMG – L area	Integral of the rectified EMG signal of left gastrocnemius divided by the sample size
EMG – R area	Integral of the rectified EMG signal of right gastrocnemius divided by the sample size
EMG – L 40-132	Left gastrocnemius relative PSD in the 40–132 Hz frequency band
EMG – L 132-224	Left gastrocnemius relative PSD in the 132–224 Hz frequency band
EMG – L 224-316	Left gastrocnemius relative PSD in the 224–316 Hz frequency band
EMG – L 316-408	Left gastrocnemius relative PSD in the 316–408 Hz frequency band
EMG – L 408-500	Left gastrocnemius relative PSD in the 408–500 Hz frequency band
EMG – R 40-132	Right gastrocnemius relative PSD in the 40–132 Hz frequency band
EMG – R 132-224	Right gastrocnemius relative PSD in the 132–224 Hz frequency band
EMG – R 224-316	Right gastrocnemius relative PSD in the 224–316 Hz frequency band
EMG – R 316-408	Right gastrocnemius relative PSD in the 316–408 Hz frequency band
EMG – R 408-500	Right gastrocnemius relative PSD in the 408–500 Hz frequency band
HR average	Heart rate average
HR std	Heart rate standard deviation

Figure 5.10: Description of the 19 biometric parameters that compose the database.

The following evaluation metrics were used to assess the performance of the algorithms into the classifications tasks: accuracy, sensitivity, specificity, and Area Under the Curve Receiver Operating Characteristics (AUCROC). All these metrics were computed using the K-Fold Cross Validation with  $k = 10$  using 10 different seeds.

**TABLE 2** | Difference of the objective brain, muscle, and health bio measurements between the first static protocol and the light (1 Hz – green) and the hard one (3 Hz – red) for all the patients (the one that did not perform the 3-Hz protocol is not included).

PATIENT ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
EEG – Delta	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EEG – Theta	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EEG – Alpha	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EEG – Beta	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EEG – LG	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EMG – L area	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EMG – R area	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EMG – L 40-132	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EMG – L 132-224	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EMG – L 224-316	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EMG – L 316-408	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EMG – L 408-500	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EMG – R 40-132	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EMG – R 132-224	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
EMG – R 224-316	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EMG – R 316-408	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
EMG – R 408-500	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
HR average	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
HR std	↓	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

The up arrows show an increase, and the down arrows show a decrease.

Figure 5.11: Difference of the objective brain, muscle, and health bio measurements between the first static protocol and the light (1 Hz – green) and the hard one (3 Hz – red) for all the patients (the one that did not perform the 3-Hz protocol is not included).

MSSQ Symptoms	0 (%)	1 (%)	2 (%)	Index	0 (%)	1 (%)	Index	0 (%)	1 (%)	Index	0 (%)	1 (%)
<b>Gen. Discomfort</b>	43.4	30.1	26.5	<b>I<sub>GenDis</sub></b>	43.4	56.6						
<b>Sweating</b>	43.4	32.5	24.1									
<b>Salivation</b>	79.5	15.7	4.8				<b>I<sub>PV</sub></b>	63.9	36.1			
<b>Nausea</b>	61.4	14.5	24.1	<b>I<sub>Stom</sub></b>	62.6	37.4						
<b>Burping</b>	94.0	6.0	0.0									
<b>Stomach Awern.</b>	66.3	18.1	15.7									
<b>Fatigue</b>	67.5	16.3	13.2							<b>I<sub>MS</sub></b>	61.4	38.6
<b>Eye Strain</b>	53.8	26.9	19.2	<b>I<sub>Fatig</sub></b>	56.6	43.4						
<b>Diff. Focusing</b>	62.7	22.9	14.5									
<b>Headache</b>	55.4	25.3	19.3				<b>I<sub>NM*</sub></b>	59.0	41.0			
<b>Full. Of Head</b>	63.9	22.9	13.2	<b>I<sub>Head</sub></b>	57.8	42.2						
<b>Blurr. Vision</b>	61.4	26.5	12.0									
<b>Dizziness - Vertigo</b>	54.2	24.1	21.7	<b>I<sub>Dizz</sub></b>	54.2	45.8						

\*I<sub>NM</sub> includes also General Discomfort.

Figure 5.12: Percentage of the MSSQ answers for each symptom, and percentage of zeros and ones for the eight computed indexes.

## 5.2.2 Results

### 5.2.2.1 Statistical Analysis

Fig. 5.14 shows the results of the statistical tests that assess the significance of the 19 parameters with the eight binary MSSQ indexes. Interestingly, only 4 out of 19 parameters never show a significance.

The EEG Beta and LG showed significance only for the individuals suffering from headache, fullness of head, and blurred vision (I<sub>Head</sub>), while no other significances were found for an EEG parameter.

The amplitude/area of EMG on both sides achieved a significance for all the conditions except for General Discomfort (I<sub>GenDis</sub>). Similarly, excluding a few cases, the power spectrum of the EMG obtained a significance for almost all the conditions except in the band of 132–224 for the left side, which is never significant.

Patient Id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27		
Gen. Discomfort	↑	-	↑	-	↓	↑	-	↓	↑	↑	↑	↑	↑	↑	↑	↑	-	-	-	↑	↑	↑	↑	↑	↑	↓	↓	-	↑
Sweating	-	-	-	↑	-	-	-	↑	↑	-	↑	↑	↑	↑	↑	↑	↑	-	-	-	↑	↑	↑	↑	↑	↑	↑	↑	
Salivation	-	-	↑	-	-	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	-	-	-	↑	↑	↑	↑	↑	↑	↑	↑	
Nausea	↑	-	↑	↑	-	↑	↑	↑	↑	-	↑	↑	↑	↑	↑	↑	↑	↓	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Burping	-	-	-	↑	-	-	-	↑	↑	-	↑	↑	↑	↑	↑	↑	↑	-	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Stomach Awern.	-	-	-	-	-	-	↑	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Fatigue	-	-	↑	-	↑	↑	↑	↓	↓	↓	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↓	↓	-	↑
Eye Strain	-	↑	-	↑	-	↑	-	↑	↑	↑	↑	↑	-	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Diff. Focusing	-	↑	↑	-	↑	-	-	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Headache	↑	-	↑	↑	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Full. Of Head	-	↑	↑	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Blurr. Vision	-	↑	-	-	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
Dizziness – Vert.	-	-	-	-	-	-	-	-	-	-	↑	-	↑	-	-	-	-	-	-	↓	↓	-	-	-	-	-	↑	↑	

The up arrows show an increase, and the down arrows show a decrease.

Figure 5.13: Difference of the subjective MS symptoms between the first static protocol and the light (1 Hz – green) and the hard (3 Hz – red) for all the patients (the one that did not perform the 3-Hz protocol is not included).

The HR Average was significant according to all indexes excluding IStom, IFatig, and IHead while HR std showed statistical significance only according to IGenDis.

The IGenDis index was the index that showed the least number of significances for the analyzed parameters; only EMG–R 40–132 and 132–4 Hz and HR parameters achieved significant results according to this index. On the other hand, IHead, IDizz, and INM were the indexes according to which the biometric parameters show the greatest number of significant results (respectively, 13 and 12).

Finally, 11 out of 19 parameters show a significant result according to the overall IMS: 10 EMG-related features, 1 HR- related feature, and no EEG feature.

	I <sub>GenDis</sub>	I <sub>Stom</sub>	I <sub>Fatig</sub>	I <sub>Head</sub>	I <sub>Dizz</sub>	I <sub>PV</sub>	I <sub>NM</sub>	I <sub>MS</sub>
EEG – Delta	0.734	0.903	0.790	0.383	0.841	0.529	0.993	0.667
EEG – Theta	0.934	0.880	0.388	0.613	0.400	0.927	0.184	0.181
EEG – Alpha	0.713	0.865	0.769	0.620	0.577	0.560	0.971	0.888
EEG – Beta	0.393	0.510	0.236	0.050*	0.194	0.105	0.130	0.335
EEG – LG	0.508	0.247	0.229	0.029*	0.070	0.087	0.081	0.217
EMG – L area	0.114	<b>0.001***</b>	0.004**	<b>0.001***</b>	<b>0.001***</b>	<b>0.001***</b>	<b>0.001***</b>	<b>0.001***</b>
EMG – R area	0.157	0.004**	<b>0.001***</b>	<b>0.001***</b>	<b>0.001***</b>	0.004**	<b>0.001***</b>	<b>0.001***</b>
EMG – L 40–132	0.274	0.029*	0.006**	<b>0.001***</b>	<b>0.001***</b>	0.040*	0.004**	0.011**
EMG – L 132–224	0.941	0.492	0.274	0.079	0.165	0.444	0.285	0.449
EMG – L 224–316	0.247	0.031*	0.006**	<b>0.001***</b>	<b>0.001***</b>	0.090	0.010**	0.013**
EMG – L 316–408	0.236	0.025*	0.012**	0.003**	<b>0.001***</b>	0.043*	0.003**	0.011**
EMG – L 408–500	0.286	0.023*	0.029*	0.003**	<b>0.001***</b>	0.032*	0.002**	0.009**
EMG – R 40–132	0.040*	0.006**	0.003**	<b>0.001***</b>	<b>0.001***</b>	0.002**	<b>0.001***</b>	0.008**
EMG – R 132–224	0.044*	0.027*	0.051*	0.035*	0.015**	0.005**	0.008**	0.100
EMG – R 224–316	0.079	0.026*	0.012**	0.007**	0.002**	0.037*	0.007**	0.040*
EMG – R 316–408	0.139	0.030*	0.006**	<b>0.001***</b>	<b>0.001***</b>	0.039*	0.003**	0.012**
EMG – R 408–500	0.118	0.020*	0.003**	<b>0.001***</b>	<b>0.001***</b>	0.018*	<b>0.001***</b>	0.003**
HR average	<b>0.001***</b>	0.219	0.082	0.314	0.012**	0.037*	0.010**	0.042*
HR std	0.040*	0.451	0.213	0.149	0.251	0.264	0.249	0.136

\*Significant at 0.05. \*\*Significant at 0.01. \*\*\*Significant at 0.001.

Figure 5.14: Significance of the 19 biometric parameters calculated with the univariate statistical analysis (Mann–Whitney test) for all the eight indexes.

Index	Alg	Acc	Sens	Spec	AUCROC
IPV	RF	<b>75.9</b>	77.5	74.4	<b>0.815</b>
	GB	74.7	75.0	74.4	0.705
	ADA-B	<b>75.9</b>	85.0	67.4	<b>0.781</b>
	SVM	60.2	62.5	58.1	0.603
	KNN	57.8	60.0	55.8	0.573
	MLP	49.4	77.5	23.3	0.642
INM	RF	<b>79.5</b>	74.3	83.3	<b>0.832</b>
	GB	69.9	71.4	68.8	0.711
	ADA-B	73.5	80.0	68.8	0.746
	SVM	60.2	51.4	66.7	0.590
	KNN	67.5	65.7	68.8	0.694
	MLP	45.8	80.0	20.8	0.737
IMS	RF	<b>74.7</b>	59.4	84.3	<b>0.801</b>
	GB	72.3	68.8	74.5	<b>0.803</b>
	ADA-B	71.1	71.9	70.6	<b>0.765</b>
	SVM	55.4	43.8	62.7	0.532
	KNN	67.5	59.4	72.5	0.670
	MLP	53.0	40.6	60.8	0.681

Table 5.5: Evaluation metrics [%] after the classification ML analysis for IPV, INM, and IMS. (The bold values are the most significant results)

	IPV	INM	IMS
<b>EEG – Delta</b>	3,74	2,37	2,47
<b>EEG – Theta</b>	5,46	4,84	5,87
<b>EEG – Alpha</b>	2,78	1,97	2,16
<b>EEG – Beta</b>	7,18	6,71	4,29
<b>EEG – LG</b>	2,68	1,48	2,16
<b>EMG – L area</b>	5,56	8,88	4,23
<b>EMG – R area</b>	9,39	12,54	14,77
<b>EMG – L 40-132</b>	6,51	6,12	6,04
<b>EMG – L 132-224</b>	3,07	2,76	4,06
<b>EMG – L 224-316</b>	2,49	5,92	3,56
<b>EMG – L 316-408</b>	7,85	4,24	4,18
<b>EMG – L 408-500</b>	5,17	3,95	6,16
<b>EMG – R 40-132</b>	8,52	7,7	7,9
<b>EMG – R 132-224</b>	7,28	3,26	5,36
<b>EMG – R 224-316</b>	2,11	7,5	4,96
<b>EMG – R 316-408</b>	6,9	5,92	4,78
<b>EMG – R 408-500</b>	5,36	3,85	2,57
<b>HR average</b>	4,89	4,54	7,72
<b>HR std</b>	3,07	5,43	6,76

Table 5.6: Feature importance (%) for IPV, INM, and IMS using Random Forest algorithm. (The marked values are the three most significant features for each index classification)

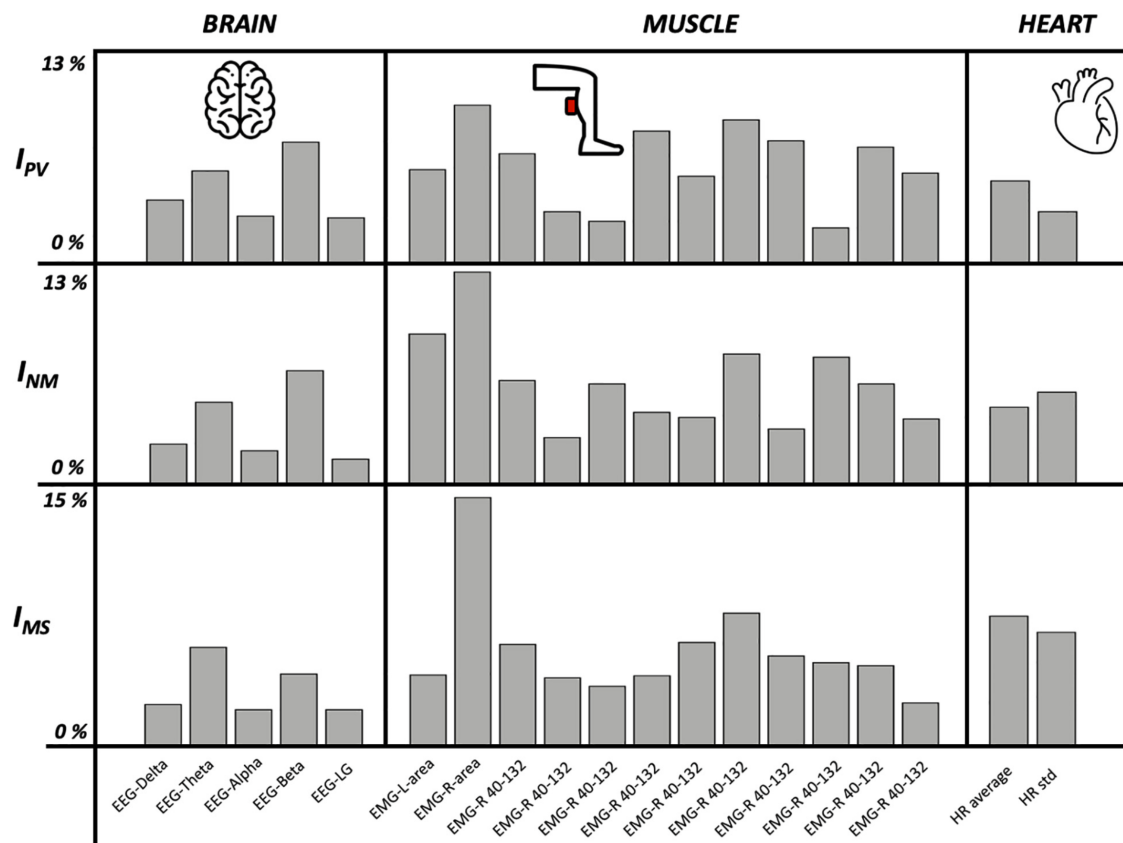


Figure 5.15: Feature importance (%) for IPV, INM, and IMS using Random Forest algorithm.

### 5.2.2.2 Machine Learning

The ML analysis focuses on the binary classification of physiological, neurological, and general MS conditions based on the MSSQ responses. We performed the classification of the following index previously defined:

- (1) ThePhysiological/VegetativeIndex(IPV),
- (2) The Neurological/Muscle Strain Index (INM),
- (3) The MS Index (IMS).

We assessed these conditions using six different algorithms, finding RF to yield the best results (Table 5.5).

As regards IPV and INM, the RF was the best algorithm for classifying both indexes; an accuracy of 75.9% with an AUCROC of 0.815 was achieved for IPV while an accuracy of 79.5% with an AUCROC of 0.832 was obtained. The highest sensitivity (85.0%) was obtained by the ADA-B for the physiological index, while the highest sensitivity was achieved by RF (74.4%). As regards the neurological index, the best sensitivity (80.0%) was achieved by the ADA-B while the best specificity (83.3%) was obtained by RF. The MLP was the worst algorithm to perform both the classifications since it reached the lowest accuracy (respectively, 49.4 and 45.8%) while the lowest

AUCROC was reached by KNN for the physiological index (0.573) and by SVM for the neurological index (0.590).

The feature importance analysis (Fig. 5.15) shows that parameters extracted from EMG were the most important ones by far for the classification of both indexes. The first EEG-based features can be found in the 5th place in the ranking while the first HR-based features can be found after the 10th place. Moreover, it has to be highlighted that the top three features in Fig. 5.15 for these indexes are all statistically significant also in the previous univariate analysis (Fig. 5.14).

Concerning IMS, the overall model for the indexes is good enough considering accuracies greater than 70.0%, AUCROC greater than 0.800, and the number of trials (equal to 83) that does not allow us to analyze a large dataset; indeed, a greater number of subjects would allow the improvement of the evaluation metrics of the models.

The feature importance analysis (Fig. 5.15 and Table 5.6) highlighted novel results for IMS: the seasickness can be strongly linked to features extracted from EMG (the top two were area and frequency analysis in the range 40–132) and HR-based (average and standard deviation were at the third and fourth place). On the other hand, another important and surprising result is the low importance of all the features extracted from the EEG, they were below the seventh place in the final ranking (this also for the other indexes except EEG-Beta which is quite relevant for IPV and INM). This can be explained by the fact that a dry cap EEG was used for the acquisition. More noise was detected and led to a lower signal quality. Channels had to be rejected and could not be interpolated, leading to an averaged PSD on less channels. This can be one of the reasons of the low significance related to EEG features.

### 5.2.3 Discussion

PC is central in governing upright posture in humans. PC failure is dual, firstly pathological disruption leading to clinical difficulties where symptoms of vertigo, dizziness, imbalance, and falling are prominent [276]. Secondly, an overstimulation of the PC system may precipitate a series of symptoms of discomfort known as MS [213]. As in PC diseases, there are many objective measurements to be used in the diagnosis of these diseases. On the other hand, there are limited ways to objectively measure MS. Questionnaires are used to evaluate the incidence of subjective symptoms associated with MS, most often nausea, pallor, vomiting, sweating, headache, lightheadedness, and body discomfort [277]. There is an urgent need for objective measurement to evaluate MS, as it threatens human well-being when one is situated in a motion-rich environment. It is also critical to objectively distinguish people into MS-prone and non-MS-prone individuals. This is possible using questionnaires [275], but having an objective way to discriminate these two groups is of great value when comes to genetic research.

In this study, we use the BioVRSea research setup and focus on EEG, EMG, and HR bio-signals associated with subjective MS symptoms. Our EEG-coupled results show significant difference in brain neural networking in individuals indicating subjective symptoms of headache, fullness of the head, and blurred vision (IHead). In an earlier study, we showed that open eyes trials reflect a greater number of significant differences in EEG absolute spectral power across all bands during both adaptation and habituation. This suggests that following both acute and prolonged proprioceptive perturbation, cortical activity may be up-regulated with the availability of visual

feedback [278]. These results generally support our prior hypothesis that the visual recognition of instability may play a critical role in governing cortical processes requisite for PC [279]. These results underline the importance of visual information in PC and simultaneously open up the VR afferent link in PC perturbations. Being able to couple these subjective symptoms, i.e., headache, fullness of the head, and blurred vision, to objective intracranial activity is crucial in clinical context and opens up ways for VR-coupled biosignal evaluation of PC pathologies [280]. This is in keeping with many other studies performed on motion and CNS triggers of head-related symptoms. Jang et al. [281] identified that the alpha band was linked to VR sickness, with a decrease of the absolute power during the experiment, followed by an increase during the recovery, highlighting a negative correlation with the MSSQ score. Kim et al. [282] detected that in the case of cybersickness, the severity of the symptoms was positively correlated with the delta wave, and negatively with the beta waves. It is interesting to see that, in our study, despite the low significance of EEG regarding the different indexes chosen, the power associated to the beta band is the parameter presenting the most importance in IPV and INM. This corroborates the fact that beta band is related to MS symptoms and is a feature that should be investigated in MS studies. Our results do not enable to drawing of hypotheses regarding the other power bands. Our observations have indicated that definite vection does not necessarily result in visually induced MS (you can have very compelling vection but no visually induced MS), but at the same time, most participants who get sick also report vection. Our intentions are to verify the relationship between MS and visually induced MS, although there are some participants that get sick on the platform but never experience MS in the real world. They probably have not experienced enough rough waters and therefore we do not expect false positive/negative.

To be able to evaluate the relationship between BioVRSea biosignals and subjective MS symptoms, the use of ML was necessary. This study clearly shows the benefit of ML; indeed, it allowed us to achieve two aims: first, the possibility to model several biometric parameters extracted from three types of signals (EEG, EMG, and HR) in order to be able to classify/distinguish patients suffering from seasickness according to these features; second, the feature importance analysis allowed us to further confirm the statistical results by ranking the features according to their contribution to the classification task. Moreover, as regards the ML models, the RF was the most reliable among all the implemented ones (Table 5.5).

The amplitude of the EMG signals in both legs showed significant difference regarding all conditions except general discomfort (IGenDis). That indicates that almost all subjective symptoms of MS showed correlation with changes in EMG. This is in context with the fact that all human efforts initiated to prevent falls, i.e., acute or long-term vertigo and dizziness, are mediated through postural stabilizing muscles [283]. Some studies used EMG measurements to analyze the behavior related to MS, with sensors placed on the abdominal muscles [284] and EMG combined with cervical vestibular myogenic potential to study the effect of scopolamine for the seasickness treatment [285]. As far as we know, no study found significance related to EMG in the lower limbs to quantify MS. This is an important piece of information regarding our BioVRSea research setup and is a promising single tool to objectively extract MS sufferers. On the other hand, this is not surprising as the prime effector in PC is aimed at muscles maintaining the upright posture and simultaneously avoiding falls. Our BioVRSea research setup might answer several clinical questions related to strategies used to prevent falls in patients with PC pathologies.



The HR parameters were significantly associated with the symptom of General Discomfort (IGenDis). The General Discomfort symptom is general in its nature and does not specifically point to MS. On the other hand, the triggered MS discomfort relates to an escalated sense of generalized panic in severe MS conditions, which is well capable of creating extreme cardiovascular deviations [286]. This is expected as HR is probably the best-known biosignal associated with numerous physical as well as pathological conditions, particularly of a PC nature.

These results can be used to have quite a whole vision of the body reaction to induced MS. This total vision can be used to help the pathological patients and the people that are more prone to MS planning an eventual rehabilitative therapy. Future ideas are to use more physiological measurements like blood oxygenation, skin sweating, and force used on the legs for the equilibrium. All these actual and future body parameters coupled with the BioVRSea system and ML are of value in further evaluation of PC disruptions, which are probably the most disturbing and costly health conditions affecting humans.

Of course, the study has some limitations. The first is the small number of subjects of this first cohort that limits the possibility of obtaining higher evaluation metrics in the ML analysis. The second is the type of population that has been analyzed in this research because it was limited regarding mainly age; all the subjects were young students. Further studies will increase the number of subjects, which would allow improvements in the performance of ML and include in the population more diverse subjects. We use for this cohort dry electrodes in the EEG acquisition process resulting in high noise signal, which we believe has limited the value of the associated EEG parameters in both statistical significance and ML. The use of a wet cap EEG will show significant improvements in the quality of the signals.

### 5.3 Female Athletes Concussion Cohort

A second relevant study about BioVRSea was published on *Scientific Reports* [6]. It is focused on defining biomarkers from the BioVRSea features for the evaluation of sport-concussion in female athletes. The article ranges between psychological and biological studies done in collaboration with the Psychology Department of the Reykjavik University.

This paper aimed to validate concussion/non-concussion classification and quantitatively assess different physiological responses during postural control tasks associated with concussion symptoms. The study was conducted on a homogeneous cohort of female athletes with a background in sports with high impact contact. The athletes' self-reported concussion history was used to divide them in two groups: concussions and non-concussion. We first validated concussion history by asking them about concussion symptoms, using the SCAT5 questionnaire. Next, concussion groups were assessed with BioVRSea (Fig. 5.16). We computed the different measurements for each individual using ML techniques with the aim of classifying concussions based on self-reported and measured parameters. We hypothesized that (1) The Icelandic versions of the SCAT5 symptoms checklist, although not a diagnosis tool, can be used to differentiate between concussed and non-concussed athletes (2) Changes of CoP, heart rate, EMG and EEG data can quantitatively measure concussion and concus-

	<b>Athletes with a history of concussion (N=26)</b>	<b>Athletes with no history of concussion (N=28)</b>
<b>Age mean (s.d.)</b>	30.5 (6.9)	29.8 (8.2)
<b>Retired %(n)</b>	69.2 (18)	57.1 (16) *
<b>Active %(n)</b>	30.8 (8)	39.3 (11)

Table 5.7: Age and athletic status of participants. \*One athlete without a history of concussion did not clarify if she was retired or still active (3.6%).

sion symptoms, (3) ML techniques using SCAT5 and neurophysiological parameters can improve assessment of concussion.

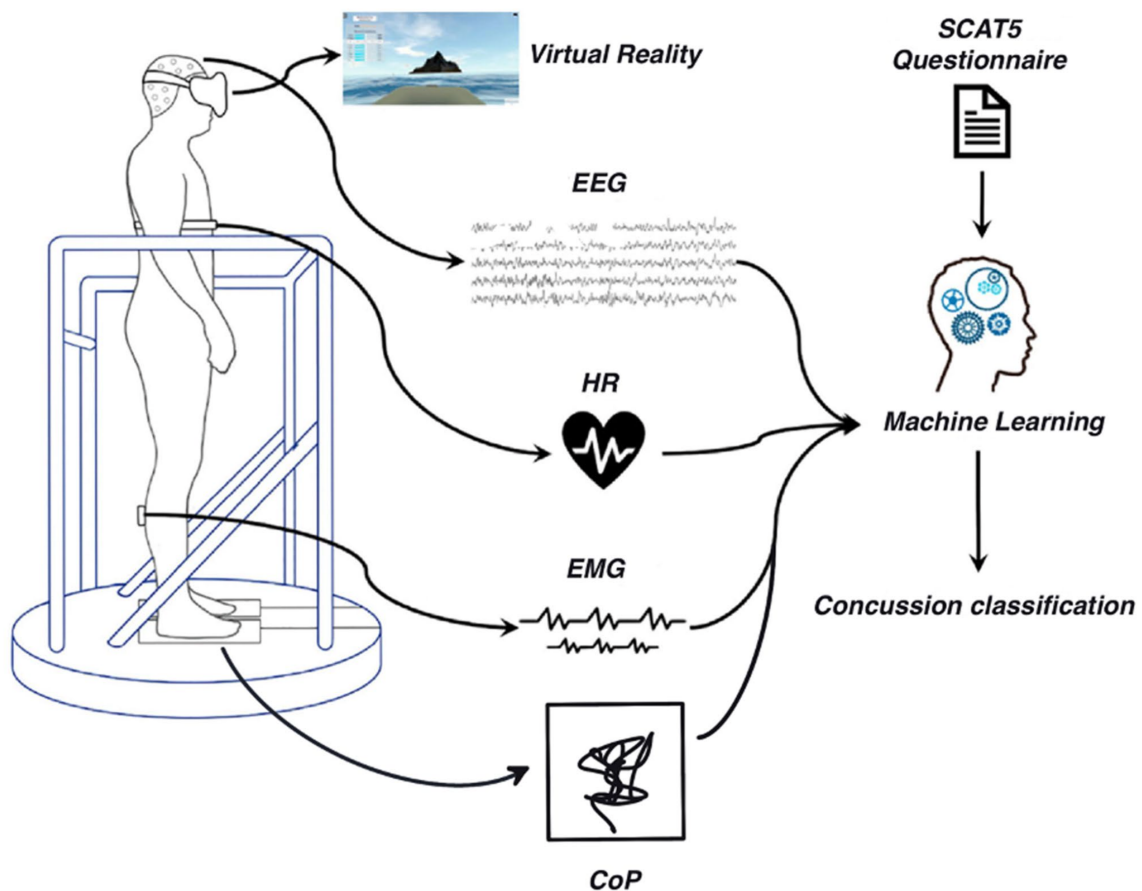


Figure 5.16: BioVRSea experimental Setup for the Present Concussion Study

### 5.3.1 BioVRSea Concussion Experiment Setup

Participants info previously listed in paragraph 5.1.4 are shown in Table 5.7. BioVRSea protocol is the one described in paragraph 5.1.2. During each protocol, muscle, brain, heart, and CoP data were acquired. The data from each measurement were divided into 6 segments, corresponding to each stage of the protocol. Data for the EEG, EMG and CoP were analyzed by calculating POST-PRE (POST minus PRE) paradigm.

Signal	Phase	Description	Tot N
HR	Pre	Average Heart Rate	5
	25		
	50		
	75		
	Post		
EMG	Pre	6 Muscles	12
	Post	Areas	
	Pre	6 Muscles	12
	Post	Median Frequency	
EEG	Pre	Alpha – Beta – Theta – Delta – Low-Gamma	10
	Post		
CoP	Pre	TOTEX – MDIST_AP- MVELO –	12
	Post	Ellipse_Main_Axis – SD_AP- Postero_Magnitude	

Table 5.8: The 51 selected features from the BioVRSea measurements.

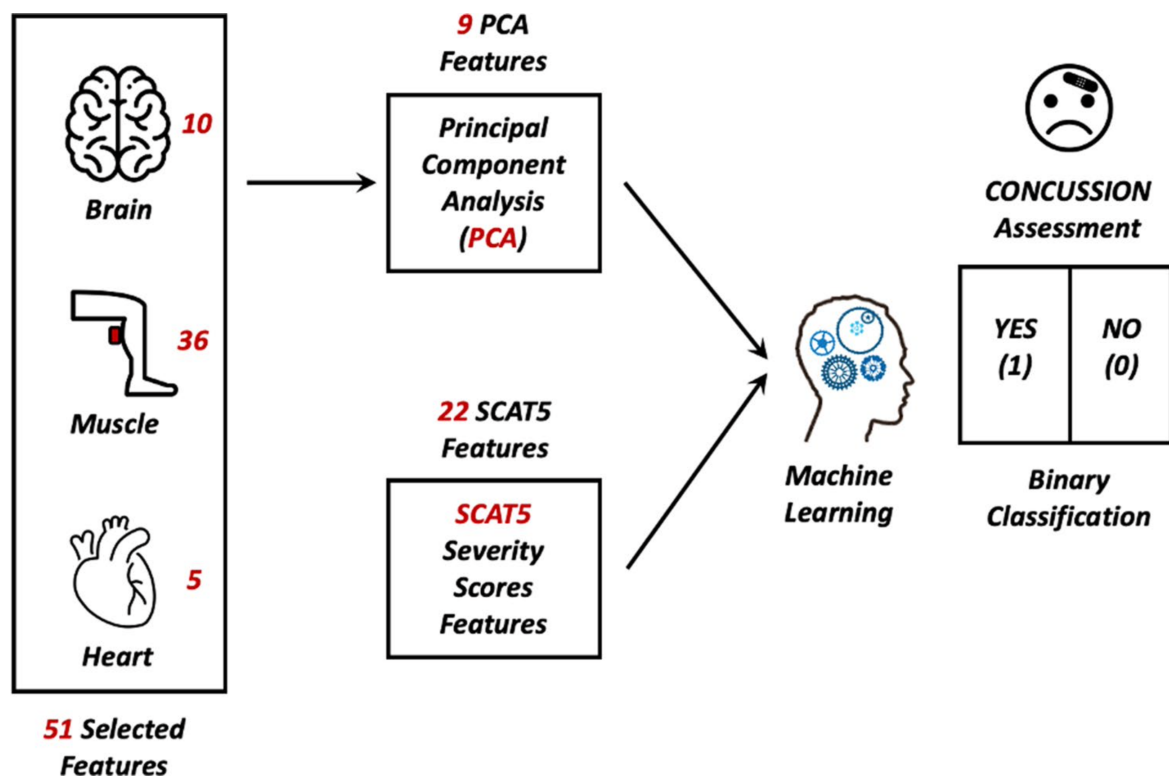


Figure 5.17: Feature Selection and Machine Learning Workflow. 51 features are selected from Brain, Muscles and Heart signal and on them, the PCA is performed to obtain 9 PCA features used to the binary classification of the concussion assessment using different ML algorithms; 22 SCAT5 features are then used as well as features for the concussion classification; the 22 SCAT5 and 9 PCA are then combined and used together to classify concussed and non-concussed subjects.

The Machine Learning analysis was performed to binary classify the participants who self-reported concussion and those who had not (Fig. 5.17). The feature extracted from the BioVRSea measurement provides a total of 51 features (Table 5.8). They are

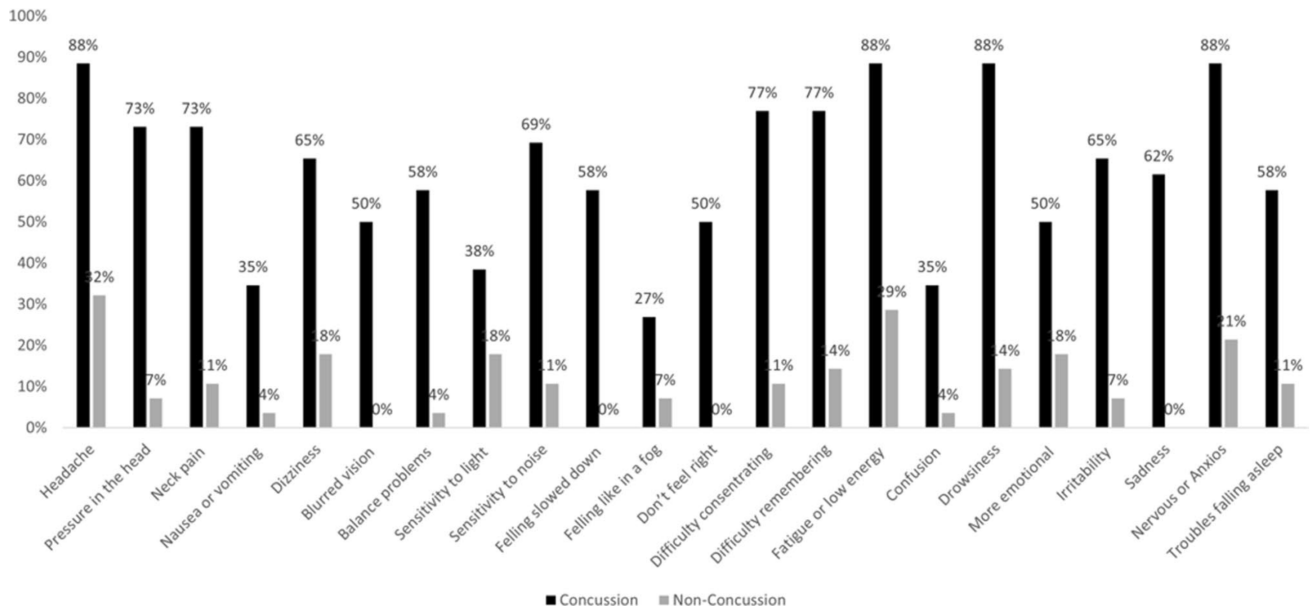


Figure 5.18: Symptoms (SCAT5) reported by all participants before and after VR acquisition. Overview of concussion symptoms reported by all participants before VR acquisition.

the 5 HR average for each experiment phase, the 10 EEG features (Delta, Theta, Alpha, Beta and Low Gamma for the PRE and POST phase), the 24 EMG features (PRE and POST areas for each of the six muscles sensors and the PRE and POST median frequencies in each muscle), and the 6 most significant CoP features (Table 5.8). On these selected 51 features, the Principal Component Analysis (PCA) is performed: this algorithm generates 9 PCA features used as input for the binary classification of the concussion participants (Fig. 5.17). The 22 Severity Score Features, SCAT5 (Fig. 5.18), are also used for the concussion classification. Finally, the 22 SCAT5 and the 9 PCA are combined, obtaining in total three different feature selections as input to the binary classification ML models:

- 1) 9 PCA features from brain, muscles, heart, and CoP signals from the BioVRSea measurements
- 2) 22 SCAT5 features from the concussion questionnaire
- 3) 31 PCA + SCAT5 features

Tree-based, linear-based and simplified artificial neural networks algorithms were used to consider different algorithmic strategies for the classification (same of paragraph 5.2) with a 10 k-fold cross validation.

### 5.3.2 Results

Table 5.9 shows the accuracy, sensitivity and specificity obtained with the 9 PCA features: the highest accuracy of 72.7 was reached with a RF model. In contrast, the

<b>PCA Features (9)</b>			
<b>Alg</b>	<b>Acc</b>	<b>Sens</b>	<b>Spec</b>
<b>RF</b>	72.7	68.4	76.0
<b>SVM</b>	61.4	30.8	92.0
<b>ADA-B</b>	68.2	57.9	76.0
<b>MLP</b>	70.5	63.2	76.0
<b>GB</b>	63.6	73.7	56.0
<b>SCAT5 Features (22)</b>			
<b>RF</b>	88.6	84.2	92.0
<b>SVM</b>	88.6	84.2	92.0
<b>ADA-B</b>	84.1	73.7	92.0
<b>MLP</b>	81.8	89.5	76.0
<b>GB</b>	77.3	78.9	76.0
<b>PCA + SCAT5 Features (31)</b>			
<b>RF</b>	90.9	89.5	92.0
<b>SVM</b>	95.5	94.7	96.0
<b>ADA-B</b>	93.2	94.7	92.0
<b>MLP</b>	90.9	96.0	86.0
<b>GB</b>	79.5	78.9	80.0

Table 5.9: Concussion classification evaluation metrics for the three different feature selections proposed: The 9 PCA features; the 22 SCAT5 features; the 31 combined features of PCA and SCAT5.

most heightened sensitivity of 57.9 was achieved with GB, another tree-based model. Table 5.9 shows the results from the SCAT5 features models, with RF and the linear-based SVM model having the highest accuracy of 88.6 and the simplified artificial neural networks of the MLP model having the highest sensitivity of 89.5. The best results were achieved by merging the two features sets (Table 5.9) with all the models exceeding 90 in accuracy except (GB), getting a significant 95.5 with SVM and 93.2 ADA-B (Fig. 5.19). All the sensitivity and specificity results for all the models with the combined feature selection are noteworthy. It is also worthy of underlining that the linear model SVM is the one from which the best accuracy is obtained with the 31 features. At the same time, the algorithm based on the artificial neural network MLP gives the best sensitivity of 96.0. Tree-based models work with good results apart from the GB, which, in all cases, is the one with the least significant accuracy results.

### 5.3.3 Discussion

The results of this study support a novel method in concussion assessment by evaluating self-reported concussion symptoms and history against neural and postural responses acquired in a BioVRSea environment, with ML used to demonstrate the classification ability of this model. We hypothesized that (1) The Icelandic versions of the SCAT5 symptoms checklist, although not a diagnosis tool, can be used to differentiate between concussed and non-concussed athletes, (2) changes of CoP, heart rate, EMG and EEG data can quantitatively measure concussion and concussion symptoms, and (3) machine learning techniques using SCAT5, and neurophysiological parameters can improve assessment of concussion.

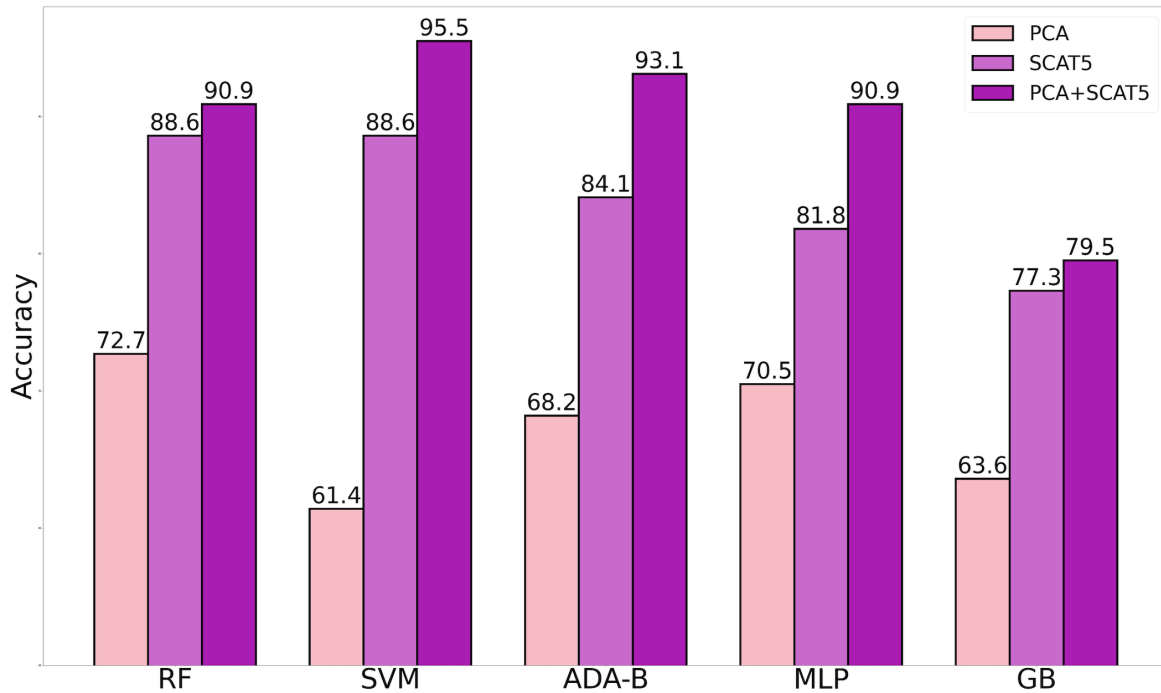


Figure 5.19: The results of 5 different algorithms (RF- Random Forest, SVM—Support Vector Machine, ADA-B— Adaptive Boosting , MLP- Multilayer Perceptron, GB- Gradient Boosted) and combinations of features. Highest accuracy in all algorithms is achieved using both the PCA and SCAT5 features, but highest accuracy overall (95.5%) when computed with an SVM algorithm.

ML results are promising and demonstrate that concussion can be assessed by the biometric measurements from BioVRSea and SCAT5, especially if combined. The use of only neurophysiological measures allows a decent classification accuracy of almost 73% with RF, which demonstrate the ability of the BioVRSea multi-biometric system not only to evaluate sick and not motion sick people as previously discussed 5.2 but also to individuate concussed and not concussed female athletes. These results increase if the BioVRSea features are combined with the SCAT5 features. The latter alone can produce an 88.6% accuracy, which is remarkable but not innovative. The novelty of the results is the successful combination of measured and self-reported parameters as seen in Fig. 5.9, with an accuracy of 95.5%. Furthermore, the combination of these two different measurement approaches provides a novel tool that can be implemented also for monitoring effect of treatment, to develop rehabilitation strategies, or even to support insurance assessment.

Discussing the different algorithmic approaches, we can state that the linear system works better when the SCAT5 features are involved. In [238], SVM was also used for a similar purpose with a larger but unbalanced dataset. Our dataset is smaller and includes fewer features, but they are quicker to assess, and the balanced number of concussed and not-concussed subjects strengthens the obtained classification accuracy. SVM was also successfully performed in [287] for an individual-level concussion detection starting from only EEG features: this can suggest that as a probable future development of the ML concussion analysis, a focus on more specific EEG features can be performed to understand better how the BioVRSea system can eventually mit-

igate or improve concussion influence on the brain. The tree-based algorithmic models confirm that the BioVRSea multimetric measurements can be better processed with this approach, like shown in the previous chapter. The simplified neural network MLP suggests us that a more complicated neural network model can be, with a larger population, worth trying to increase the classification ability of the model.

As participants were all actively or historically involved in contact sports and as such, this is a group at elevated risk for receiving sub-concussive injuries throughout their careers [288][289]. The comparison between the concussion and non-concussion groups is not a comparison between a concussion group and a normal population. Both groups will likely have received repeated head and body impacts, with possible sub-concussive blows.

The limited number of participants limits the ML algorithm's predictive capabilities, but the balanced number of concussed and not-concussed subject is a strength. Recruitment of more participants in the future could improve the results in terms of accuracy, and more complicated algorithmic models can be implemented.

We demonstrated that we can discriminate between concussion and non-concussion groups using the BioVRSea setup and particularly, symptoms associated with concussion, especially with balance problems, follow a pattern that can be quantified. This study shows the value of a subject-specific multi-faceted PC assessment. This study also proves that the ML application on BioVRSea paradigms can be of great value not only for the MS classification but also for different studies related to PC disorders, like concussion classification or the lifestyle assessment discussed in the following chapter.

## 5.4 Lifestyle prediction with EEG and CoP signals

The present section is adapted from a further study relative to MS using BioVRSea which was presented at the *IEEE Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE) International Conference* in Rome, October 2022 [7]. The paper discusses the influence of positive and negative lifestyle on the MS symptoms using BioVRSea paradigms and the cohort of 262 people described in paragraph 5.1.4.

The paper was awarded as "*Best PhD Contribution*" by the *IEEE MetroXRINE* committee.

In this paper, following the results of [241][245], we study the differences of healthy and not healthy subjects on BioVRSea predicting, using ML techniques, a lifestyle index (section 5.1.6 based on the one previously introduced in [2][290][291] where we explored the lifestyle and physical activity influence using the AGES Icelandic database (chapter 3.1.3)). A similar approach, using as input features the EMG and CoP data from BioVRSea, is here implemented to understand which of these features are the most relevant to distinguish healthy and unhealthy subjects (Fig. 5.20).

### 5.4.1 BioVRSea Features for MS Lifestyle study

For this study EMG and CoP data are taken into consideration.

The muscle electrical activity has as final output seven features computed in the fre-

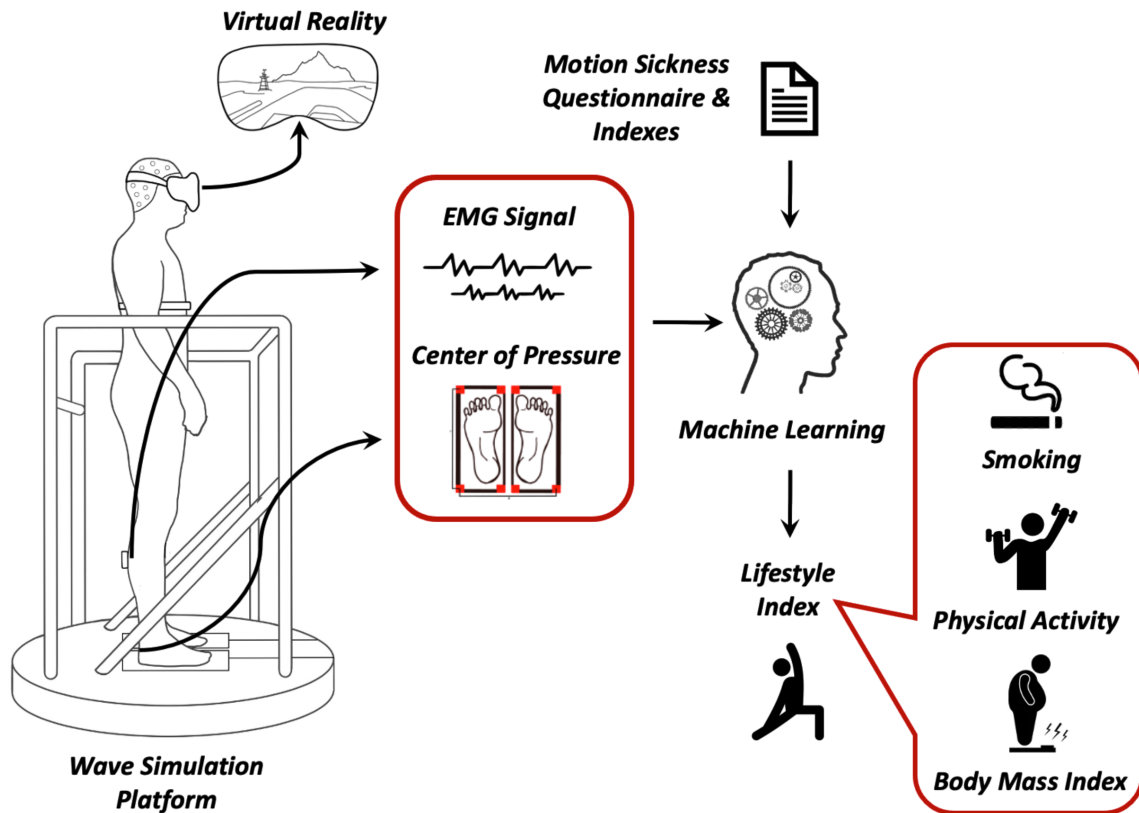


Figure 5.20: BioVRSea: workflow for the study presented in Rome at IEEE MetroX-RAINE Conference, October 2022.

quency domain and thirty-five features in the time domain (a total of forty-two (42)) for each muscle and each experiment's segment.

The sensors to measure the CoP produce Thirty-five (35) parameters extracted from the stabilogram to evaluate the PC response of the subject before (PRE), during (25%, 50%, 75%) and after (POST) the movements of the simulation platform and the visual input provided by the VR goggles.

Considering all the 6 different segments of the experiment plus the difference between POST and PRE, POST and BASELINE, and PRE and BASELINE, all the EMG sensors with their respective 42 features extracted and the 35 features calculated from the CoP sensors a total of almost 2500 features are present in the database. Therefore, a feature analysis with respective selection of the most relevant features for the lifestyle index classification is needed.

Three different groups of features are considered to perform the binary lifestyle index classification: the first comprehends all the EMG features, the second comprehends all the CoP features and the third one combines the first two groups considering all the available features (EMG+CoP).

For the feature selection the Analysis of Variance (ANOVA) f-test is performed: it consists in a parametric statistical hypothesis test for determining whether the means from two samples of data (can be also three or more) come from the same or a different distribution. The scikit-learn function GridSearchCV with stratified 5 folds is



Feat. Selection	Alg	Acc Max	Acc Mean	Rec	AUCROC
<b>CoP</b>	<b>RF</b>	0.83	0.72	0.40	0.59
	<b>SVM-sig</b>	0.78	0.72	0.36	0.39
	<b>KNN</b>	0.73	0.67	0.60	0.58
	<b>MLP</b>	0.77	0.55	0.42	0.46
	<b>GB</b>	0.72	0.68	0.35	0.61
<b>EMG</b>	<b>RF</b>	0.79	0.73	0.36	0.55
	<b>SVM-sig</b>	0.77	0.75	0.09	0.55
	<b>KNN</b>	0.77	0.68	0.45	0.54
	<b>MLP</b>	0.77	0.58	0.36	0.54
	<b>GB</b>	0.74	0.62	0.33	0.52
<b>CoP+EMG</b>	<b>RF</b>	0.80	0.76	0.22	0.55
	<b>SVM-sig</b>	0.81	0.77	0.16	0.57
	<b>KNN</b>	0.74	0.65	0.44	0.52
	<b>MLP</b>	0.79	0.58	0.40	0.49
	<b>GB</b>	0.73	0.67	0.32	0.62

Table 5.10: ML Results on Different Feature Groups with Tuning Based on Recall

Feat. Selection	Alg	Acc Max	Acc Mean	Rec	AUCROC
<b>CoP</b>	<b>RF</b>	0.80	0.76	0.13	0.64
	<b>SVM-sig</b>	0.78	0.77	0.00	0.38
	<b>KNN</b>	0.83	0.78	0.13	0.61
	<b>MLP</b>	0.79	0.77	0.05	0.42
	<b>GB</b>	0.74	0.69	0.34	0.60
<b>EMG</b>	<b>RF</b>	0.81	0.76	0.13	0.58
	<b>SVM-sig</b>	0.77	0.76	0.00	0.50
	<b>KNN</b>	0.83	0.71	0.11	0.55
	<b>MLP</b>	0.79	0.77	0.11	0.54
	<b>GB</b>	0.79	0.67	0.18	0.45
<b>CoP+EMG</b>	<b>RF</b>	0.83	0.77	0.18	0.57
	<b>SVM-sig</b>	0.79	0.78	0.00	0.45
	<b>KNN</b>	0.79	0.78	0.00	0.49
	<b>MLP</b>	0.79	0.77	0.09	0.51
	<b>GB</b>	0.76	0.74	0.26	0.57

Table 5.11: ML Results on Different Feature Groups with Tuning Based on Accuracy

performed to select the best 5,10 and 15 features. For the lifestyle index classification task a stratified 5-fold cross validation is performed using four different algorithms: RF, SVM with sigmoid kernels, KNN, MLP, and GB. As evaluation metrics accuracy, recall (reference class is the not-healthy group) and AUCROC are calculated. For the tuning of the hyperparameters both accuracy and recall are taken in consideration.

## 5.4.2 Results

Table 5.10 shows the results of the lifestyle index classification with the tuning of the hyperparameters on the recall while Table 5.11 is similar to Table 5.10, but the tuning

Ellipse Main Axis Length PRE
MVELO TOTAL PRE
TOTEX TOTAL PRE
MVELO TOTAL 25
MVELO AP PRE
TOTEX ML 25
MVELO ML 25
Magnitudo Antero 25
TOTEX ML 50

Table 5.12: Most Significant CoP Features for the highest accuracy and recall with KNN algorithm

is done on the accuracy.

Table 5.10 shows the best recall result (0.60) and a significant AUCROC of 0.58 with KNN and the CoP features. KNN is the best algorithms in terms of recall, while tuning the parameters on the recall the maximum accuracy is obtained with RF (0.83 with CoP features) and the highest mean accuracy is with SVM algorithm and CoP and EMG features together (0.77).

In table 5.11 it is possible to see that the accuracy max and mean is clearly better than in Table 5.10 at the expense of the recall which is at its max 0.34 and sometimes also null. The best algorithm is here KNN reaching a max average accuracy of 0.78 with the CoP and CoP+EMG features. The max accuracy is again with KNN and CoP (0.83). It results the best approach in term of accuracy mean and max together with RF and all the features together. The big difference between table 5.10 and 5.11 is mainly the accuracy mean which increase if the parameters are tuned on it, at the opposite the recall values significantly decrease in table 5.11. The best AUCROC value is reached with the CoP feat selection and RF (Table 5.11 - 0.64). MLP in both tables can be considered the algorithm with the worst metrics.

Table 5.12 shows the most important features for the highest accuracy combination of algorithms and features which is for both Tables 5.10 and 5.11 KNN algorithm with CoP features. The feature from the POST phase are rarely selected as highly important for the lifestyle index binary classification while PRE and 25% segments contribute with higher relevance in the ML model. TOTEX is the summation over the elementary movement between consecutive samples on the support plane. MVELO is the mean velocity of point on the support plane computed as TOTEX/Total Time. Magnitudo is the furthest point in the plane, in the case of Table 5.12 in the anterior portion of the plane. AP means Anterior-Posterior axis while ML is for Medio-Lateral axis.

The box-plot in Fig 5.21 shows the distribution of the 3 first features included in Table 5.12 underling the differences between the healthy and unhealthy people who underwent the BioVRSea experiment.

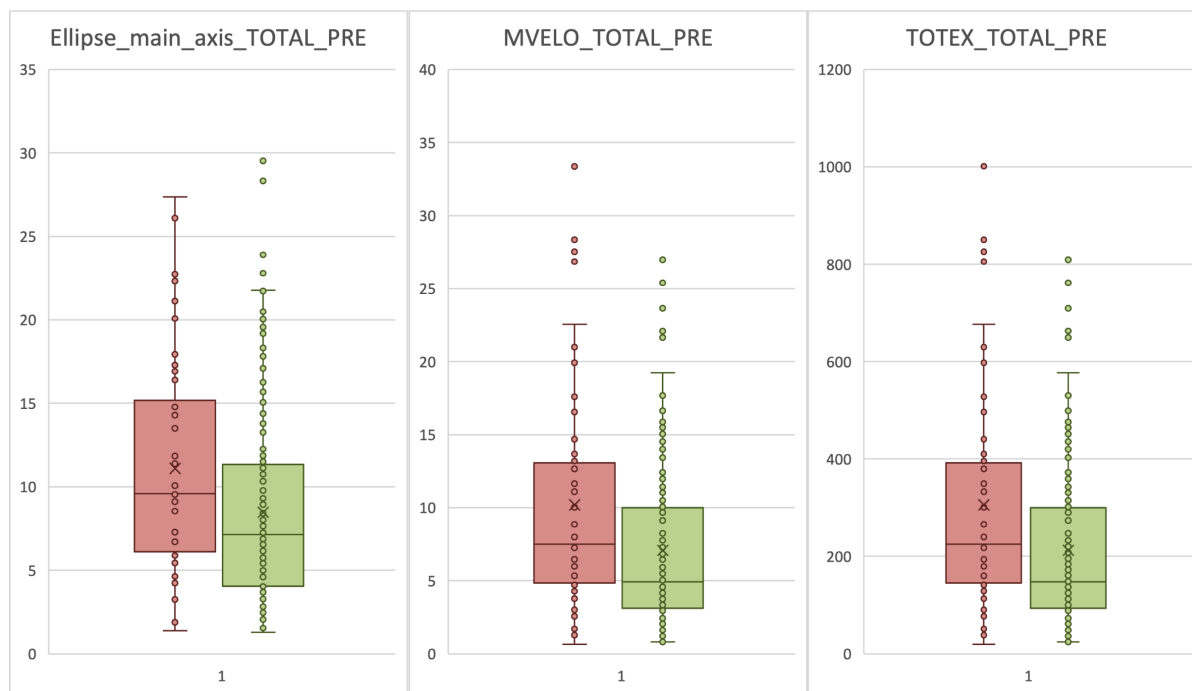


Figure 5.21: Boxplot for the 3 most important features for the lifestyle index classification. The red indicates the unhealthy people while green is referred to healthy participants.

### 5.4.3 Discussion

The results from BioVRSea are again highly satisfactory and we can say that these features, originally thought to study MS, are of high potential also to identify people with an healthy or unhealthy lifestyle. CoP features in particular, shows relevance in the classification model and it is clear, also from Fig. 5.21, that their distribution is different for the two classes of people. The most relevant features for the classification are from the PRE and 25% segment of the experiment: this means that it is the lifestyle itself that influences the performances on the platform and not only the virtual simulation. From these results we can state that a bad lifestyle will have a relevant negative influence on the final MS symptoms results.

The present study can be used as preliminary work to study the influence of lifestyle behavior not only in physical tasks on the virtual platform, but also in terms of MS symptoms understanding if smoking, physical activity, and BMI differs in people with strong nausea or dizziness at the end of the different BioVRSea segments.

A possible future improvement to this research can be studying the lifestyle influence on BioVRSea biometric paradigms on Parkinsonian people, as 11 people affected by early stage Parkinson disease already underwent the BioVRSea protocol. It is known in literature that PC task can show high difference between healthy and Parkinsonian people [292] [293], but the influence of lifestyle was not studied yet in relation to induced MS. A clear difference in behavior can be assumed between healthy and neuro-degenerative patients starting from the results proposed here.

We can then conclude from our results, as similarly mentioned in literature by Golding et al.[241] and Caillet et al.[245], that an unhealthy lifestyle influences, in a negative way, the performances of a person in terms of equilibrium and balance in an induced

MS task.

# Chapter 6

## Conclusion and Future Development

In the clinical context, biomedical images and signals have been demonstrated to be essential tools for diagnostic and clinical investigations: their features, combined with modern AI and ML tools, have had a significant impact on the scientific community [294][295]. This dissertation shows the work and the impact of the medical images and biosignal features. Firstly, in the aging study, we examined soft tissue radiodensitometric features extracted from mid-thigh CT scans of elderly populations and used them to train ML models to classify cardiac pathophysiological and comorbidities such as diabetes and hypertension. Single and dual gait features were also examined and found to be highly relevant in classifying elderly Korean patients with a history of stroke or motor and neurological deficits. Secondly, 2D and 3D features from multiple knee medical images were used to assess cartilage degeneration in the knee osteoarthritis study. Finally, the novel BioVRSea system and its brain, muscle, and postural stability biosignals extracted from different cohorts of healthy and concussed subjects raised significant impact for clinical advances in motion sickness and postural control studies. However, from these studies, some limitations emerge.

**Dataset Dimension** One limitation of the studies discussed is the size of the datasets, including both the number of features and the number of subjects. In the aging study, although the AGES dataset has a significant number of subjects, only eleven features are considered in the models to classify a single comorbidity. Increasing the number of soft tissue features, possibly with the aid of ANN applied to the CT scans, while maintaining the high number of subjects, could provide further clinical insights into the effects of muscle, fat, and connective tissue on comorbidities such as cardiac status, diabetes, and hypertension, and potentially allow for feature selection.

In contrast, the MS study faces a different dimensional limitation. Although the BioVRSea dataset has thousands of features per acquisition when considering all the biosignals and MSSQ information, the number of subjects is still limited compared to AGES. To overcome this limitation, ongoing data acquisition for the BioVRSea study aims to include approximately 1,000 subjects, both healthy and pathological, to make it one of the most representative databases for PC and MS studies. With a larger number of subjects, it would be possible to obtain a more population-representative view of the symptoms of an induced MS environment, and potentially identify patterns that could have a significant scientific impact in quantifying MS levels or in identifying PC pathologies and other disorders.

In the knee osteoarthritis study, the 96 total features are a considerable number, given

the time-consuming extraction process, but the high cost of CT and MRI acquisitions partially limits the potential of the study as it may be difficult to obtain images from thousands or even hundreds of degenerative, traumatic, or healthy subjects.

**Dataset Imbalance** The imbalance of the dataset classes can be considered a further limitation. It can only be partially addressed by data augmentation techniques such as SMOTE. The potential of SMOTE has been widely discussed in the literature and is currently a trending topic in the field [296][297]. During training, SMOTE generates a larger set that balances the minority class, leading to higher evaluation metrics. However, these metrics may not necessarily reflect the actual values of the original biomedical data since they are based on an augmented set. Therefore, to effectively apply these techniques, the metrics from the "real" dataset and the "augmented" one must be compared. The evaluations of SMOTE metrics must be scrutinized scientifically, as discussed in the lifestyle study on NTRA in Chapter 3.1.3. In that study, the use of SMOTE was necessary to achieve the clinical goal of correlating the radiodensitometric features with the classified comorbidities. The results without the augmentation technique demonstrated typical outcomes of an unbalanced set (high accuracy and sensitivity, very low specificity), which was not useful in obtaining a significant view of the impact of ML technologies on the clinical goal. However, in all other study cases, the use of SMOTE was unnecessary or would have led to results that were not truly representative of significant clinical impact. Although the RESTORE dataset had a small number of control patients, it could have been applied, but the total number of patients and the goal of multiclassification would have led to excessively high results that were not scientifically significant. Data augmentation is an efficient tool for studying unbalanced datasets, but when working with biomedical data, caution must be taken, and evaluation metrics should be compared with the "original" ones to obtain a clear view of the effect of the oversampling.

**Deep Learning Potential** Additionally, DL techniques were not considered in this thesis. In the aging study, NTRA features were extracted indirectly from the CT-Scan, as the 128 densitometric HU bins were used to generate distribution curves (refer to section 3.1.1.1). Directly working with the CT itself presents an opportunity for employing DL algorithms to derive new features from soft tissues and incorporate them with the NTRA in the development of new ML classification models. In the investigation of knee osteoarthritis, DL could also prove to be a valuable tool for generating highly precise 2D and 3D features. Nevertheless, a significant increase in the number of patients is required in this scenario to thoroughly evaluate the impact of the ANN on knee cartilage image elaboration.

**Future Development** NTRA radiodensitometric features were initially developed as a tool to provide insight into the optimal diagnosis and quantification of soft tissue degeneration in sarcopenia [142]. This thesis demonstrates their strong predictive value and their effect on comorbidities such as diabetes, hypertension, and cardiovascular outcomes that may not, at first glance, appear related to the soft tissue status of a mid-thigh CT scan. The results reveal significant predictive potential and a high

impact in aging studies. Similarly, research utilizing a radiodensitometry approach from thigh and abdominal CT scans with the same AGES population demonstrated that skeletal muscle attenuation decreases while adipose tissue attenuation increases with aging, confirming the robust clinical impact of the radiodensitometric approach [298].

Ongoing and future developments of the NTRA research include the extraction of radiodensitometric features from both legs of the AGES subjects, as the AGES dataset contains mid-thigh CT-scans of both the left and right leg. The main goal is to study how the asymmetry between soft tissues in the legs can be related to BMI or comorbidities that have already been predicted by NTRA, such as cardiac pathophysiologies, diabetes, and hypertension. Asymmetry can also influence gait performance and has been shown to deteriorate with age. Ongoing research is analyzing the direction of the asymmetry (left or right) to understand if it follows the dominant leg trend or if it is determined by any life factors and eventually changes with age. While muscle asymmetry has been studied in top professional athletes [299] or to improve sports performance [300], there are no relevant studies in the scientific literature focused on aging, mobility, and asymmetry studied on soft tissue. The novelty of this asymmetry study could have a significant impact on aging and gerontology, strengthening the scientific impact of the NTRA features and the single predictive value of muscle, fat, and connective tissue radiodensitometric parameters.

The NTRA approach has potential for application to other CT scan images, such as those included in different datasets like the one created for the RESTORE project. This European project, which is continuing in the SinPain consortium, aims to develop a patient-specific multiscale and sensitive knee bone/cartilage profile by analyzing specific features of the 2D and 3D set, including a larger number of subjects, especially in the healthy control group. ML technologies will make a substantial contribution to the feature analysis, and unsupervised techniques could also be implemented to cluster patients according to the cartilage status described by the novel profile.

Furthermore, gait features extracted from the Seoul University Hospital dataset have demonstrated promising potential. The acquisition protocol used in this study could be adapted and implemented on Icelandic subjects at Lanspítali University Hospital to investigate potential differences with the Korean population. Additionally, a motor dual-task component could be incorporated into the existing cognitive dual-task protocol to enable further analysis on the significance of dual-task gait in the classification of pathologies that commonly affect elderly individuals.

Regarding biosignal analysis in MS studies, the BioVRSea approach has shown significant impact in the field and has been recognized by Keshavarz and Golding as one of the most significant examples in the literature of ML application to predict the presence of MS [301]. A possible future development is the inclusion of subjects with anxiety, and furthermore, ML techniques could be used to identify the most significant features for quantifying the reaction of symptoms in an induced MS virtual environment. Ongoing studies are also focusing on the use of EMG, EEG, and CoP features to classify early-stage Parkinson's disease or subjects with neurodegenerative diseases. Professional athletes, with or without concussion, could also be included in the study to understand how possible PC disorders or biosignal-related abnormalities can affect sports performance. In addition, a clustering unsupervised approach could be developed to study the potential of biosignal features to distinguish two or more clusters of subjects that hypothetically correspond to MS index groups and to healthy and pathological subject groups. The involvement of seamen could also be a future

implementation with a significant impact on the improvement of health and safety during long sea voyages. Additionally, it could be useful in preventing the symptomatic effects of "mal du débarquement" (sickness of disembarkment) [302].

The impact of the digital revolution on the healthcare system is very extensive, and it is leading to significant cultural and technological progress and substantial changes in clinical approaches across various healthcare domains. At the core of the success and efficacy of these approaches lies a fundamental shift in the data acquisition mentality, which emphasizes the attainment of precise clinical goals, offering a broader view of the problem by analyzing a more diverse group of individuals. This translates in design and use of health data structure that immediately allows optimal implementation of ML algorithms. Therefore, the interdisciplinary synergistic collaboration between engineers and medical professionals is vital in achieving effective research planning that prioritizes the long-term vision of the clinical problem at hand, rather than focusing solely on engineering objectives to attain high evaluation metrics. Disregarding the real clinical problem can lead to a significant loss of scientific value in research projects. Ultimately, the goal is to employ digital technologies and foster multidisciplinary synergies within the newly digitized society to improve clinical evaluation and diagnostics, leading to a tangible impact on people's daily lives.



# Chapter 7

## List of Publications from the Author

Marco Recenti's list of peer-reviewed publications as main or contributor author in scientific journals or conferences, presented in chronological order from the most recent:

- Aubonnet, R., Ramos, J., Recenti, M., Jacob, D., Ciliberti, F., Guerrini, L., ... & Gargiulo, P. (2022). Toward New Assessment of Knee Cartilage Degeneration. *Cartilage*, 19476035221144746; DOI: 10.1177/19476035221144746.
- Stehle, S. A., Aubonnet, R., Hassan, M., Recenti, M., Jacob, D., Petersen, H., & Gargiulo, P. (2022). Predicting postural control adaptation measuring EEG, EMG, and center of pressure changes: BioVRSea paradigm. *Frontiers in Human Neuroscience*, 16; DOI: 10.3389/fnhum.2022.1038976
- Recenti, M., Jacob, D., Aubonnet, R., Burgunder, B., Escalona, I. M., Gunnarsson, A. E., ... & Gargiulo, P. (2022, October). Predicting lifestyle using BioVRSea multi-biometric paradigms. In *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)* (pp. 329-334). IEEE; DOI: 10.1109/MetroXRINE54828.2022.9967685. - (*Winner of Best PhD Contribution Award*)
- Jacob, D., Aubonnet, R., Recenti, M., Audardottir, S. A., Ivarsdottir, T. I., Burgunder, B., ... & Gargiulo, P. (2022, October). Assessing Early-stage Parkinson's Disease Using BioVRSea. In *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)* (pp. 271-276). IEEE; DOI: 10.1109/MetroXRINE54828.2022.9967502 - (*Winner of Best Contribution from a Female Researcher Award*)
- Ciliberti, F. K., Cesarelli, G., Guerrini, L., Gunnarsson, A. E., Forni, R., Aubonnet, R., ... & Gargiulo, P. (2022). The role of bone mineral density and cartilage volume to predict knee cartilage degeneration. *European Journal of Translational Myology*, 32(2); DOI: 10.4081/ejtm.2022.10678
- Jacob, D., Unnsteinsdóttir Kristensen, I. S., Aubonnet, R., Recenti, M., Donisi, L., Ricciardi, C., ... & Gargiulo, P. (2022). Towards defining biomarkers to evaluate concussions using virtual reality and a moving platform (BioVRSea). *Scientific Reports*, 12(1), 1-22; DOI: 10.1038/s41598-022-12822-0.
- Ciliberti, F. K., Guerrini, L., Gunnarsson, A. E., Recenti, M., Jacob, D., Cangianno, V., ... & Aubonnet, R. (2022). CT-and MRI-Based 3D Reconstruction

of Knee Joint to Assess Cartilage and Bone. *Diagnostics*, 12(2), 279; DOI: 10.3390/diagnostics12020279.

- Recenti, M., Ricciardi, C., Edmunds, K., Jacob, D., Gambacorta, M., & Gargiulo, P. (2021). Testing soft tissue radiodensity parameters interplay with age and self-reported physical activity. *European Journal of Translational Myology*, 31(3); DOI: 10.4081/ejtm.2021.9929
- Recenti, M., Ricciardi, C., Aubonnet, R., Picone, I., Jacob, D., Svansson, H. Á., ... & Gargiulo, P. (2021). Toward predicting motion sickness using virtual reality and a moving platform assessing brain, muscles, and heart signals. *Frontiers in Bioengineering and Biotechnology*, 9, 635661; DOI: 10.3389/fbioe.2021.635661.
- Recenti, M., Ricciardi, C., Monet, A., Jacob, D., Ramos, J., Gíslason, M., ... & Gargiulo, P. (2021). Predicting body mass index and isometric leg strength using soft tissue distributions from computed tomography scans. *Health and Technology*, 11(1), 239-249; DOI: 10.1007/s12553-020-00498-3.
- Recenti, M., Ricciardi, C., Edmunds, K. J., Gíslason, M. K., Sigurdsson, S., Carraro, U., & Gargiulo, P. (2020). Healthy aging within an image: Using muscle radiodensitometry and lifestyle factors to predict diabetes and hypertension. *IEEE Journal of Biomedical and Health Informatics*, 25(6), 2103-2112; DOI: 10.1109/JBHI.2020.3044158.
- Ricciardi, C., Jónsson Jr, H., Jacob, D., Improta, G., Recenti, M., Gíslason, M. K., ... & Gargiulo, P. (2020). Improving prosthetic selection and predicting BMD from biometric measurements in patients receiving total hip arthroplasty. *Diagnostics*, 10(10), 815; DOI: 10.3390/diagnostics10100815
- Recenti, M., Ricciardi, C., Aubonnet, R., Esposito, L., Jónsson, H., & Gargiulo, P. (2020, June). A regression approach to assess bone mineral density of patients undergoing total hip arthroplasty through gait analysis. In *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 1-6). IEEE; DOI: 10.1109/MeMeA49120.2020.9137182
- Recenti, M., Ricciardi, C., Edmunds, K., Gíslason, M. K., & Gargiulo, P. (2020). Machine learning predictive system based upon radiodensitometric distributions from mid-thigh CT images. *European Journal of Translational Myology*, 30(1); DOI: 10.4081/ejtm.2019.8892
- Ricciardi, C., Edmunds, K. J., Recenti, M., Sigurdsson, S., Gudnason, V., Carraro, U., & Gargiulo, P. (2020). Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions. *Scientific reports*, 10(1), 1-13; DOI: 10.1038/s41598-020-59873-9.
- Recenti, M., Gíslason, M. K., Edmunds, K. J., & Gargiulo, P. (2020). Aging health behind an image: Quantifying sarcopenia and associated risk factors from advanced CT analysis and machine learning technologies. In *International Symposium on Computer Methods in Biomechanics and Biomedical Engineering* (pp. 188-197). Springer, Cham; DOI: 10.1007/978-3-030-43195-2\_15

- Recenti, M., Ricciardi, C., Gislason, M., Edmunds, K., Carraro, U., & Gargiulo, P. (2019, September). Machine learning algorithms predict body mass index using nonlinear trimodal regression analysis from computed tomography scans. In *Mediterranean Conference on Medical and Biological Engineering and Computing* (pp. 839-846). Springer, Cham; DOI: 10.1007/978-3-030-31635-8\_100



# Bibliography

- [1] C. Ricciardi, K. J. Edmunds, M. Recenti, *et al.*, “Assessing cardiovascular risks from a mid-thigh ct image: A tree-based machine learning approach using radiodensitometric distributions”, *Scientific reports*, vol. 10, no. 1, pp. 1–13, 2020.
- [2] M. Recenti, C. Ricciardi, K. J. Edmunds, *et al.*, “Healthy aging within an image: Using muscle radiodensitometry and lifestyle factors to predict diabetes and hypertension”, *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2103–2112, 2020.
- [3] M. Recenti, P. Gargiulo, M. Chang, S. B. Ko, T. J. Kim, and S. U. Ko, “Predicting stroke, neurological and movement disorders using single and dual-task gait in korean older population”, *Gait & Posture - UNDER REVIEW*, 2023.
- [4] R. Aubonnet, J. Ramos, M. Recenti, *et al.*, “Toward new assessment of knee cartilage degeneration”, *Cartilage*, p. 19 476 035 221 144 746, 2022.
- [5] M. Recenti, C. Ricciardi, R. Aubonnet, *et al.*, “Toward predicting motion sickness using virtual reality and a moving platform assessing brain, muscles, and heart signals”, *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 635 661, 2021.
- [6] D. Jacob, I. S. Unnsteinsdóttir Kristensen, R. Aubonnet, *et al.*, “Towards defining biomarkers to evaluate concussions using virtual reality and a moving platform (biovrsea)”, *Scientific Reports*, vol. 12, no. 1, pp. 1–22, 2022.
- [7] M. Recenti, D. Jacob, R. Aubonnet, *et al.*, “Predicting lifestyle using biovrsea multi-biometric paradigms”, in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroX-RAINE)*, IEEE, 2022, pp. 329–334.
- [8] A. Turing, *Computing machinery and intelligence*. 1950.
- [9] T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997, vol. 1.
- [10] M. W. Berry, A. Mohamed, and B. W. Yap, *Supervised and unsupervised learning for data science*. Springer, 2019.
- [11] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.
- [12] R. Kohavi *et al.*, “A study of cross-validation and bootstrap for accuracy estimation and model selection”, in *Ijcai*, Montreal, Canada, vol. 14, 1995, pp. 1137–1145.
- [13] T.-T. Wong, “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation”, *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015.

- [14] S. Purushotham and B. Tripathy, “Evaluation of classifier models using stratified tenfold cross validation techniques”, in *Global Trends in Information Systems and Software Applications: 4th International Conference, ObCom 2011, Vellore, TN, India, December 9-11, 2011. Proceedings, Part II*, Springer, 2012, pp. 680–690.
- [15] M. M. Ghiasi and S. Zendehboudi, “Application of decision tree-based ensemble learning in the classification of breast cancer”, *Computers in Biology and Medicine*, vol. 128, p. 104089, 2021.
- [16] L. Breiman, *Classification and regression trees*. Routledge, 2017.
- [17] M. Banerjee, E. Reynolds, H. B. Andersson, and B. K. Nallamothu, “Tree-based analysis: A practical approach to create clinical decision-making tools”, *Circulation: Cardiovascular Quality and Outcomes*, vol. 12, no. 5, e004879, 2019.
- [18] S. B. Kotsiantis, “Decision trees: A recent overview”, *Artificial Intelligence Review*, vol. 39, pp. 261–283, 2013.
- [19] T. K. Ho, “Random decision forests”, in *Proceedings of 3rd international conference on document analysis and recognition*, IEEE, vol. 1, 1995, pp. 278–282.
- [20] L. Breiman, “Random forests”, *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [21] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees”, *Machine learning*, vol. 63, pp. 3–42, 2006.
- [22] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting”, *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [23] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, “Advance and prospects of adaboost algorithm”, *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013.
- [24] J. H. Friedman, “Greedy function approximation: A gradient boosting machine”, *Annals of statistics*, pp. 1189–1232, 2001.
- [25] T. Cover and P. Hart, “Nearest neighbor pattern classification”, *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [26] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.
- [27] C. Cortes and V. Vapnik, “Support-vector networks”, *Machine learning*, vol. 20, pp. 273–297, 1995.
- [28] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.
- [29] S. Ghosh, A. Dasgupta, and A. Swetapadma, “A study on support vector machine based linear and non-linear pattern classification”, in *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, 2019, pp. 24–28.
- [30] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.
- [31] S. Haykin, *Neural networks and learning machines, 3/E*. Pearson Education India, 2009.

- [32] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [33] F. Murtagh, “Multilayer perceptrons for classification and regression”, *Neurocomputing*, vol. 2, no. 5-6, pp. 183–197, 1991.
- [34] M. Riedmiller and H. Braun, “A direct adaptive method for faster backpropagation learning: The rprop algorithm”, in *IEEE international conference on neural networks*, IEEE, 1993, pp. 586–591.
- [35] A. G. Asuero, A. Sayago, and A. González, “The correlation coefficient: An overview”, *Critical reviews in analytical chemistry*, vol. 36, no. 1, pp. 41–59, 2006.
- [36] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance”, *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.
- [37] T. Chai and R. R. Draxler, “Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature”, *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [38] M. Hossin and M. N. Sulaiman, “A review on evaluation metrics for data classification evaluations”, *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.
- [39] V. Bewick, L. Cheek, and J. Ball, “Statistics review 13: Receiver operating characteristic curves”, *Critical care*, vol. 8, no. 6, pp. 1–5, 2004.
- [40] L. Yang and A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice”, *Neurocomputing*, vol. 415, pp. 295–316, 2020.
- [41] P. Probst, A.-L. Boulesteix, and B. Bischl, “Tunability: Importance of hyperparameters of machine learning algorithms”, *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1934–1965, 2019.
- [42] G. E. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data”, *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.
- [43] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique”, *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [44] G. Chandrashekar and F. Sahin, “A survey on feature selection methods”, *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [45] N. Sánchez-Marroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, “Filter methods for feature selection—a comparative study”, *Lecture notes in computer science*, vol. 4881, pp. 178–187, 2007.
- [46] N. El Aboudi and L. Benhlila, “Review on wrapper feature selection approaches”, in *2016 International Conference on Engineering & MIS (ICEMIS)*, IEEE, 2016, pp. 1–5.
- [47] L. Melkumova and S. Y. Shatskikh, “Comparing ridge and lasso estimators for data analysis”, *Procedia engineering*, vol. 201, pp. 746–755, 2017.

- [48] K. Pearson, “Liii. on lines and planes of closest fit to systems of points in space”, *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559–572, 1901.
- [49] I. T. Jolliffe, *Principal component analysis for special types of data*. Springer, 2002.
- [50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [51] D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification”, in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 2012, pp. 3642–3649.
- [52] D. Chicco, P. Sadowski, and P. Baldi, “Deep autoencoder neural networks for gene ontology annotation predictions”, in *Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics*, 2014, pp. 533–540.
- [53] N. F. Greenwald, G. Miller, E. Moen, *et al.*, “Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning”, *Nature biotechnology*, vol. 40, no. 4, pp. 555–565, 2022.
- [54] R. Abdollahi-Arpanahi, D. Gianola, and F. Peñagaricano, “Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes”, *Genetics Selection Evolution*, vol. 52, pp. 1–15, 2020.
- [55] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep learning applications in medical image analysis”, *Ieee Access*, vol. 6, pp. 9375–9389, 2017.
- [56] B. D. Mittelstadt and L. Floridi, “The ethics of big data: Current and foreseeable issues in biomedical contexts”, *The ethics of biomedical big data*, pp. 445–480, 2016.
- [57] F. F. Costa, “Big data in biomedicine”, *Drug discovery today*, vol. 19, no. 4, pp. 433–440, 2014.
- [58] J. Mathaiyan, A. Chandrasekaran, and S. Davis, “Ethics of genomic research”, *Perspectives in clinical research*, vol. 4, no. 1, p. 100, 2013.
- [59] D. H. Huson, S. Beier, I. Flade, *et al.*, “Megan community edition-interactive exploration and analysis of large-scale microbiome sequencing data”, *PLoS computational biology*, vol. 12, no. 6, e1004957, 2016.
- [60] T. B. Murdoch and A. S. Detsky, “The inevitable application of big data to health care”, *Jama*, vol. 309, no. 13, pp. 1351–1352, 2013.
- [61] M. M. Rodgers, V. M. Pai, and R. S. Conroy, “Recent advances in wearable sensors for health monitoring”, *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3119–3126, 2014.
- [62] C. Sudlow, J. Gallacher, N. Allen, *et al.*, “Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age”, *PLoS medicine*, vol. 12, no. 3, e1001779, 2015.
- [63] P. Gopalan, W. Hao, D. M. Blei, and J. D. Storey, “Scaling probabilistic models of genetic variation to millions of humans”, *Nature genetics*, vol. 48, no. 12, pp. 1587–1590, 2016.
- [64] L. L. Cavalli-Sforza, “The human genome diversity project: Past, present and future”, *Nature Reviews Genetics*, vol. 6, no. 4, pp. 333–340, 2005.



- [65] I. G. P. Consortium *et al.*, “An integrated map of genetic variation from 1,092 human genomes”, *Nature*, vol. 491, no. 7422, p. 56, 2012.
- [66] I. Lazaridis, N. Patterson, A. Mittnik, *et al.*, “Ancient human genomes suggest three ancestral populations for present-day europeans”, *Nature*, vol. 513, no. 7518, pp. 409–413, 2014.
- [67] D. Klarin, S. M. Damrauer, K. Cho, *et al.*, “Genetics of blood lipids among ~300,000 multi-ethnic participants of the million veteran program”, *Nature genetics*, vol. 50, no. 11, pp. 1514–1523, 2018.
- [68] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis, “Deep learning: New computational modelling techniques for genomics”, *Nature Reviews Genetics*, vol. 20, no. 7, pp. 389–403, 2019.
- [69] M. W. Libbrecht and W. S. Noble, “Machine learning applications in genetics and genomics”, *Nature Reviews Genetics*, vol. 16, no. 6, pp. 321–332, 2015.
- [70] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, “Predicting the sequence specificities of dna-and rna-binding proteins by deep learning”, *Nature biotechnology*, vol. 33, no. 8, pp. 831–838, 2015.
- [71] J. Listgarten, S. Damaraju, B. Poulin, *et al.*, “Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms”, *Clinical cancer research*, vol. 10, no. 8, pp. 2725–2737, 2004.
- [72] T. B. Harris, L. J. Launer, G. Eiriksdottir, *et al.*, “Age, gene/environment susceptibility–reykjavik study: Multidisciplinary applied phenomics”, *American journal of epidemiology*, vol. 165, no. 9, pp. 1076–1087, 2007.
- [73] P. Suetens, *Fundamentals of medical imaging*. Cambridge university press, 2017.
- [74] G. Choy, O. Khalilzadeh, M. Michalski, *et al.*, “Current applications and future impact of machine learning in radiology”, *Radiology*, vol. 288, no. 2, pp. 318–328, 2018.
- [75] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis”, *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [76] K. Suzuki, “Overview of deep learning in medical imaging”, *Radiological physics and technology*, vol. 10, no. 3, pp. 257–273, 2017.
- [77] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, “Deep learning for brain mri segmentation: State of the art and future directions”, *Journal of digital imaging*, vol. 30, pp. 449–459, 2017.
- [78] A. D. Weston, P. Korfiatis, T. L. Kline, *et al.*, “Automated abdominal segmentation of ct scans for body composition analysis using deep learning”, *Radiology*, vol. 290, no. 3, pp. 669–679, 2019.
- [79] Y. Ding, J. H. Sohn, M. G. Kawczynski, *et al.*, “A deep learning model to predict a diagnosis of alzheimer disease by using 18f-fdg pet of the brain”, *Radiology*, vol. 290, no. 2, pp. 456–464, 2019.
- [80] T. Fujioka, K. Kubota, M. Mori, *et al.*, “Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network”, *Japanese journal of radiology*, vol. 37, pp. 466–472, 2019.

- [81] K. Fritscher, P. Raudaschl, P. Zaffino, M. F. Spadea, G. C. Sharp, and R. Schubert, “Deep neural networks for fast segmentation of 3d medical images”, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, Springer, 2016, pp. 158–165.
- [82] J. Schaefer, M. Lehne, J. Schepers, F. Prasser, and S. Thun, “The use of machine learning in rare diseases: A scoping review”, *Orphanet journal of rare diseases*, vol. 15, pp. 1–10, 2020.
- [83] S. Roy, W. Menapace, S. Oei, *et al.*, “Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound”, *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2676–2687, 2020.
- [84] H.-t. Zhang, J.-s. Zhang, H.-h. Zhang, *et al.*, “Automated detection and quantification of covid-19 pneumonia: Ct imaging analysis by a deep learning-based software”, *European journal of nuclear medicine and molecular imaging*, vol. 47, pp. 2525–2532, 2020.
- [85] H. R. Roth, C. T. Lee, H.-C. Shin, *et al.*, “Anatomy-specific classification of medical images using deep convolutional nets”, in *2015 IEEE 12th international symposium on biomedical imaging (ISBI)*, IEEE, 2015, pp. 101–104.
- [86] W. Zhang, R. Li, H. Deng, *et al.*, “Deep convolutional neural networks for multi-modality isointense infant brain image segmentation”, *NeuroImage*, vol. 108, pp. 214–224, 2015.
- [87] U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, *et al.*, “Automated detection of alzheimer’s disease using brain mri images—a study with various feature extraction techniques”, *Journal of Medical Systems*, vol. 43, pp. 1–14, 2019.
- [88] C. Parmar, P. Grossmann, D. Rietveld, M. M. Rietbergen, P. Lambin, and H. J. Aerts, “Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer”, *Frontiers in oncology*, vol. 5, p. 272, 2015.
- [89] L. Macyszyn, H. Akbari, J. M. Pisapia, *et al.*, “Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques”, *Neuro-oncology*, vol. 18, no. 3, pp. 417–425, 2015.
- [90] B. Zhang, X. He, F. Ouyang, *et al.*, “Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma”, *Cancer letters*, vol. 403, pp. 21–27, 2017.
- [91] Radiopaedia.org. “Normal brain cases”. (2020), [Online]. Available: <https://doi.org/10.53347/rID-80429> (visited on 02/13/2023).
- [92] M. Akay, *Biomedical signal processing*. Academic press, 2012.
- [93] L. G. Tassinary, J. T. Cacioppo, and E. J. Vanman, “The skeletomotor system: Surface electromyography.”, 2007.
- [94] B. Rodriguez-Tapia, I. Soto, D. M. Martinez, and N. C. Arballo, “Myoelectric interfaces and related applications: Current state of emg signal processing—a systematic review”, *IEEE Access*, vol. 8, pp. 7792–7805, 2020.
- [95] S. Rissanen, M. Kankaanpää, M. P. Tarvainen, *et al.*, “Analysis of surface emg signal morphology in parkinson’s disease”, *Physiological measurement*, vol. 28, no. 12, p. 1507, 2007.

- [96] K. R. Mills, “Detecting fasciculations in amyotrophic lateral sclerosis: Duration of observation required”, *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 82, no. 5, pp. 549–551, 2011.
- [97] M. Atzori, M. Cognolato, and H. Müller, “Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands”, *Frontiers in neurorobotics*, vol. 10, p. 9, 2016.
- [98] M. Knaflitz and F. Molinari, “Assessment of muscle fatigue during biking”, *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 11, no. 1, pp. 17–23, 2003.
- [99] T. Tokuyasu, S. Kushizaki, S. Matsumoto, and T. Kitawaki, “Development of automatic positioning system for bicycle saddle based on lower limb’s emg signals during pedaling motion”, in *2013 IEEE 6th International Workshop on Computational Intelligence and Applications (IWCIA)*, IEEE, 2013, pp. 27–32.
- [100] J. Yu, S. Park, S.-H. Kwon, C. M. B. Ho, C.-S. Pyo, and H. Lee, “Ai-based stroke disease prediction system using real-time electromyography signals”, *Applied Sciences*, vol. 10, no. 19, p. 6791, 2020.
- [101] W. J. Brady, *Electrocardiogram in clinical medicine*. John Wiley & Sons, 2020.
- [102] M. Wasimuddin, K. Elleithy, A.-S. Abuzneid, M. Faezipour, and O. Abuzaghlleh, “Stages-based ecg signal analysis from traditional signal processing to machine learning approaches: A survey”, *IEEE Access*, vol. 8, pp. 177 782–177 803, 2020.
- [103] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, “Heart-beat classification fusing temporal and morphological information of ecgs via ensemble of classifiers”, *Biomedical Signal Processing and Control*, vol. 47, pp. 41–48, 2019.
- [104] G. Sannino and G. De Pietro, “A deep learning approach for ecg-based heartbeat classification for arrhythmia detection”, *Future Generation Computer Systems*, vol. 86, pp. 446–455, 2018.
- [105] A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, “An integrated machine learning approach to stroke prediction”, in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 183–192.
- [106] K. Rathakrishnan, S.-N. Min, and S. J. Park, “Evaluation of ecg features for the classification of post-stroke survivors with a diagnostic approach”, *Applied Sciences*, vol. 11, no. 1, p. 192, 2020.
- [107] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2013.
- [108] M. X. Cohen, “It’s about time”, *Frontiers in human neuroscience*, vol. 5, p. 2, 2011.
- [109] M. X. Cohen, *Analyzing neural time series data: theory and practice*. MIT press, 2014.
- [110] M. Teplan *et al.*, “Fundamentals of eeg measurement”, *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.
- [111] M.-P. Hosseini, A. Hosseini, and K. Ahi, “A review on machine learning for eeg signal processing in bioengineering”, *IEEE reviews in biomedical engineering*, vol. 14, pp. 204–218, 2020.

- [112] W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik, “A machine learning framework involving eeg-based functional connectivity to diagnose major depressive disorder (mdd)”, *Medical & biological engineering & computing*, vol. 56, pp. 233–246, 2018.
- [113] Y. Wang, W. Chen, K. Huang, and Q. Gu, “Classification of neonatal amplitude-integrated eeg using random forest model with combined feature”, in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, IEEE, 2013, pp. 285–290.
- [114] V. Tuyisenge, L. Trebault, M. Bhattacharjee, *et al.*, “Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning”, *Clinical Neurophysiology*, vol. 129, no. 3, pp. 548–554, 2018.
- [115] A. Neuro. “Eegotm mylab acquisition system - ant neuro”. (2023), [Online]. Available: [https://www.ant-neuro.com/products/eego\\_mylab](https://www.ant-neuro.com/products/eego_mylab) (visited on 02/14/2023).
- [116] D. Levine, J. Richards, and M. W. Whittle, *Whittle’s gait analysis*. Elsevier health sciences, 2012.
- [117] C. Ricciardi, M. Amboni, C. De Santis, *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)*, IEEE, 2020, pp. 1–6.
- [118] R. LeMoyne, F. Heerinckx, T. Aranca, R. De Jager, T. Zesiewicz, and H. J. Saal, “Wearable body and wireless inertial sensors for machine learning classification of gait for people with friedreich’s ataxia”, in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, IEEE, 2016, pp. 147–151.
- [119] R. LeMoyne, T. Mastroianni, A. Hessel, and K. Nishikawa, “Implementation of machine learning for classifying prosthesis type through conventional gait analysis”, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 202–205.
- [120] C. Prakash, R. Kumar, and N. Mittal, “Recent developments in human gait research: Parameters, approaches, applications, machine learning techniques, datasets and challenges”, *Artificial Intelligence Review*, vol. 49, pp. 1–40, 2018.
- [121] S. B. Kwon, H.-S. Han, M. C. Lee, H. C. Kim, Y. Ku, *et al.*, “Machine learning-based automatic classification of knee osteoarthritis severity using gait data and radiographic images”, *IEEE Access*, vol. 8, pp. 120 597–120 603, 2020.
- [122] R. Z. U. Rehman, S. Del Din, Y. Guan, A. J. Yarnall, J. Q. Shi, and L. Rochester, “Selecting clinically relevant gait characteristics for classification of early parkinson’s disease: A comprehensive machine learning approach”, *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.
- [123] B. H. Goodpaster, C. L. Carlson, M. Visser, *et al.*, “Attenuation of skeletal muscle and strength in the elderly: The health abc study”, *Journal of applied physiology*, vol. 90, no. 6, pp. 2157–2165, 2001.
- [124] E. Volpi, R. Nazemi, and S. Fujita, “Muscle tissue changes with aging”, *Current opinion in clinical nutrition and metabolic care*, vol. 7, no. 4, p. 405, 2004.

- [125] K.-i. Sasaki and Y. Fukumoto, "Sarcopenia as a comorbidity of cardiovascular disease", *Journal of Cardiology*, vol. 79, no. 5, pp. 596–604, 2022.
- [126] R. R. Kalyani, M. Corriere, and L. Ferrucci, "Age-related and disease-related muscle loss: The effect of diabetes, obesity, and other diseases", *The lancet Diabetes & endocrinology*, vol. 2, no. 10, pp. 819–829, 2014.
- [127] S. V. Brooks and J. A. Faulkner, "Skeletal muscle weakness in old age: Underlying mechanisms.", *Medicine and science in sports and exercise*, vol. 26, no. 4, pp. 432–439, 1994.
- [128] R. Maughan, J. S. Watson, and J. Weir, "Strength and cross-sectional area of human skeletal muscle.", *The Journal of physiology*, vol. 338, no. 1, pp. 37–49, 1983.
- [129] A. M. Campos, F. A. Moura, S. N. Santos, W. M. Freitas, A. C. Sposito, *et al.*, "Sarcopenia, but not excess weight or increased caloric intake, is associated with coronary subclinical atherosclerosis in the very elderly", *Atherosclerosis*, vol. 258, pp. 138–144, 2017.
- [130] R. L. Reed, L. Pearlmutter, K. Yochum, K. E. Meredith, and A. D. Mooradian, "The relationship between muscle mass and muscle strength in the elderly", *Journal of the American Geriatrics Society*, vol. 39, no. 6, pp. 555–561, 1991.
- [131] M. Zamboni, A. P. Rossi, F. Corzato, C. Bambace, G. Mazzali, and F. Fantin, "Sarcopenia, cachexia and congestive heart failure in the elderly", *Endocrine, Metabolic & Immune Disorders-Drug Targets (Formerly Current Drug Targets-Immune, Endocrine & Metabolic Disorders)*, vol. 13, no. 1, pp. 58–67, 2013.
- [132] A. Collamati, E. Marzetti, R. Calvani, *et al.*, "Sarcopenia in heart failure: Mechanisms and therapeutic strategies", *Journal of geriatric cardiology: JGC*, vol. 13, no. 7, p. 615, 2016.
- [133] I. Janssen, S. B. Heymsfield, and R. Ross, "Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability", *Journal of the American Geriatrics Society*, vol. 50, no. 5, pp. 889–896, 2002.
- [134] J. Butler, A. Kalogeropoulos, V. Georgiopoulou, *et al.*, "Incident heart failure prediction in the elderly: The health abc heart failure score", *Circulation: Heart Failure*, vol. 1, no. 2, pp. 125–133, 2008.
- [135] B. H. Goodpaster, D. E. Kelley, F. L. Thaete, J. He, and R. Ross, "Skeletal muscle attenuation determined by computed tomography is associated with skeletal muscle lipid content", *Journal of applied physiology*, vol. 89, no. 1, pp. 104–110, 2000.
- [136] S. Jubrias, I. R. Odderson, P. C. Esselman, and K. E. Conley, "Decline in isokinetic force with age: Muscle cross-sectional area and specific force", *Pflügers Archiv*, vol. 434, no. 3, pp. 246–253, 1997.
- [137] A. Young, M. Stokes, and M. Crowe, "The size and strength of the quadriceps muscles of old and young men.", *Clinical physiology (Oxford, England)*, vol. 5, no. 2, pp. 145–154, 1985.

- [138] U. Carraro, K. J. Edmunds, and P. Gargiulo, “3d false color computed tomography for diagnosis and follow-up of permanent denervated human muscles submitted to home-based functional electrical stimulation”, *European Journal of Translational Myology*, vol. 25, no. 2, 2015.
- [139] P. Gargiulo, H. Kern, U. Carraro, *et al.*, “Quantitative color three-dimensional computer tomography imaging of human long-term denervated muscle”, *Neurological research*, vol. 32, no. 1, pp. 13–19, 2010.
- [140] G. E. Hicks, E. M. Simonsick, T. B. Harris, *et al.*, “Cross-sectional associations between trunk muscle composition, back pain, and physical function in the health, aging and body composition study”, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 60, no. 7, pp. 882–887, 2005.
- [141] K. J. Edmunds, Í. Árnadóttir, M. K. Gíslason, U. Carraro, and P. Gargiulo, “Nonlinear trimodal regression analysis of radiodensitometric distributions to quantify sarcopenic and sequelae muscle degeneration”, *Computational and Mathematical Methods in Medicine*, vol. 2016, 2016.
- [142] K. Edmunds, M. Gíslason, S. Sigurðsson, *et al.*, “Advanced quantitative methods in correlating sarcopenic muscle degeneration with lower extremity function biometrics and comorbidities”, *PloS one*, vol. 13, no. 3, e0193241, 2018.
- [143] J. P. Reis, C. M. Loria, P. D. Sorlie, Y. Park, A. Hollenbeck, and A. Schatzkin, “Lifestyle factors and risk for new-onset diabetes: A population-based cohort study”, *Annals of internal medicine*, vol. 155, no. 5, pp. 292–299, 2011.
- [144] C. I. Neutel and N. R. Campbell, “Changes in lifestyle after hypertension diagnosis in canada”, *Canadian Journal of Cardiology*, vol. 24, no. 3, pp. 199–204, 2008.
- [145] P. Anand, A. B. Kunnumakara, C. Sundaram, *et al.*, “Cancer is a preventable disease that requires major lifestyle changes”, *Pharmaceutical research*, vol. 25, no. 9, pp. 2097–2116, 2008.
- [146] L. Djoussé, J. A. Driver, and J. M. Gaziano, “Relation between modifiable lifestyle factors and lifetime risk of heart failure”, *Jama*, vol. 302, no. 4, pp. 394–400, 2009.
- [147] C. Thoma, C. P. Day, and M. I. Trenell, “Lifestyle interventions for the treatment of non-alcoholic fatty liver disease in adults: A systematic review”, *Journal of hepatology*, vol. 56, no. 1, pp. 255–266, 2012.
- [148] J. Twisk, B. Staal, M. Brinkman, H. Kemper, and W. Van Mechelen, “Tracking of lung function parameters and the longitudinal relationship with lifestyle”, *European Respiratory Journal*, vol. 12, no. 3, pp. 627–634, 1998.
- [149] B. Stengel, M. E. Tarver-Carr, N. R. Powe, M. S. Eberhardt, and F. L. Brancati, “Lifestyle factors, obesity and the risk of chronic kidney disease”, *Epidemiology*, pp. 479–487, 2003.
- [150] R. M. Van Dam, T. Li, D. Spiegelman, O. H. Franco, and F. B. Hu, “Combined impact of lifestyle factors on mortality: Prospective cohort study in us women”, *Bmj*, vol. 337, 2008.
- [151] A. D. Flouris, B. E. Faight, and P. Klentrou, “Cardiovascular disease risk in adolescent smokers: Evidence of smoker lifestyle”, *Journal of Child Health Care*, vol. 12, no. 3, pp. 221–231, 2008.

- [152] K. Liu, M. L. Daviglius, C. M. Loria, *et al.*, “Healthy lifestyle through young adulthood and the presence of low cardiovascular disease risk profile in middle age: The coronary artery risk development in (young) adults (cardia) study”, *Circulation*, vol. 125, no. 8, pp. 996–1004, 2012.
- [153] L. Ferrucci, G. Izmirlian, S. Leveille, *et al.*, “Smoking, physical activity, and active life expectancy”, *American journal of epidemiology*, vol. 149, no. 7, pp. 645–653, 1999.
- [154] C. M. Burchfiel, D. S. Sharp, J. D. Curb, *et al.*, “Physical activity and incidence of diabetes: The honolulu heart program”, *American journal of epidemiology*, vol. 141, no. 4, pp. 360–368, 1995.
- [155] F. B. Hu, “Sedentary lifestyle and risk of obesity and type 2 diabetes”, *Lipids*, vol. 38, no. 2, pp. 103–108, 2003.
- [156] K. Patja, P. Jousilahti, G. Hu, T. Valle, Q. Qiao, and J. Tuomilehto, “Effects of smoking, obesity and physical activity on the risk of type 2 diabetes in middle-aged finnish men and women”, *Journal of internal medicine*, vol. 258, no. 4, pp. 356–362, 2005.
- [157] R. J. Dougherty, J. B. Lindheimer, A. J. Stegner, S. Van Riper, O. C. Okonkwo, and D. B. Cook, “An objective method to accurately measure cardiorespiratory fitness in older adults who cannot satisfy widely used oxygen consumption criteria”, *Journal of Alzheimer’s Disease*, vol. 61, no. 2, pp. 601–611, 2018.
- [158] S. A. Prince, K. B. Adamo, M. E. Hamel, J. Hardt, S. C. Gorber, and M. Tremblay, “A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review”, *International journal of behavioral nutrition and physical activity*, vol. 5, no. 1, pp. 1–24, 2008.
- [159] N. Šarabon, D. Smajla, Ž. Kozinc, and H. Kern, “Speed-power based training in the elderly and its potential for daily movement function enhancement”, *European Journal of Translational Myology*, vol. 30, no. 1, 2020.
- [160] M. W. Whittle, *Gait analysis: an introduction*. Butterworth-Heinemann, 2014.
- [161] L. Rochester, V. Hetherington, D. Jones, *et al.*, “Attending to the task: Interference effects of functional tasks on walking in parkinson’s disease and the roles of cognition, depression, fatigue, and balance”, *Archives of physical medicine and rehabilitation*, vol. 85, no. 10, pp. 1578–1585, 2004.
- [162] Y.-R. Yang, S.-J. Cheng, Y.-J. Lee, Y.-C. Liu, and R.-Y. Wang, “Cognitive and motor dual task gait training exerted specific training effects on dual task gait performance in individuals with parkinson’s disease: A randomized controlled pilot study”, *PloS one*, vol. 14, no. 6, e0218180, 2019.
- [163] O. Bock, “Dual-task costs while walking increase in old age for some, but not for other tasks: An experimental study of healthy young and elderly persons”, *Journal of neuroengineering and rehabilitation*, vol. 5, no. 1, pp. 1–9, 2008.
- [164] 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.
- [165] D. Goyal, K. R. Jerripathula, and A. Mittal, “Detection of gait abnormalities caused by neurological disorders”, in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, 2020, pp. 1–6.

- [166] B. Balaban and F. Tok, “Gait disturbances in patients with stroke”, *Pm&r*, vol. 6, no. 7, pp. 635–642, 2014.
- [167] A. Holzinger, “Machine learning for health informatics”, in *Machine learning for health informatics*, Springer, 2016, pp. 1–24.
- [168] C. Ricciardi, M. Amboni, C. De Santis, *et al.*, “Using gait analysis’ parameters to classify parkinsonism: A data mining approach”, *Computer methods and programs in biomedicine*, vol. 180, p. 105 033, 2019.
- [169] P.-H. Chen, C.-W. Lien, W.-C. Wu, L.-S. Lee, and J.-S. Shaw, “Gait-based machine learning for classifying patients with different types of mild cognitive impairment”, *Journal of medical systems*, vol. 44, no. 6, pp. 1–6, 2020.
- [170] B. Pogorelc, Z. Bosnić, and M. Gams, “Automatic recognition of gait-related health problems in the elderly using machine learning”, *Multimedia tools and applications*, vol. 58, no. 2, pp. 333–354, 2012.
- [171] Y. Zhou, R. Romijnders, C. Hansen, *et al.*, “The detection of age groups by dynamic gait outcomes using machine learning approaches”, *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [172] Y. Zhang and J. M. Jordan, “Epidemiology of osteoarthritis”, *Clinics in geriatric medicine*, vol. 26, no. 3, pp. 355–369, 2010.
- [173] M. Cross, E. Smith, D. Hoy, *et al.*, “The global burden of hip and knee osteoarthritis: Estimates from the global burden of disease 2010 study”, *Annals of the rheumatic diseases*, vol. 73, no. 7, pp. 1323–1330, 2014.
- [174] L. Murphy, T. A. Schwartz, C. G. Helmick, *et al.*, “Lifetime risk of symptomatic knee osteoarthritis”, *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, vol. 59, no. 9, pp. 1207–1213, 2008.
- [175] T. Neogi, “The epidemiology and impact of pain in osteoarthritis”, *Osteoarthritis and cartilage*, vol. 21, no. 9, pp. 1145–1153, 2013.
- [176] R. Karvonen, W. Negendank, R. Teitge, A. Reed, P. Miller, and F. Fernandez-Madrid, “Factors affecting articular cartilage thickness in osteoarthritis and aging.”, *The Journal of rheumatology*, vol. 21, no. 7, pp. 1310–1318, 1994.
- [177] F. W. Roemer, S. Demehri, P. Omoumi, *et al.*, “State of the art: Imaging of osteoarthritis—revisited 2020”, *Radiology*, vol. 296, no. 1, pp. 5–21, 2020.
- [178] M. D. Crema, F. W. Roemer, M. D. Marra, *et al.*, “Articular cartilage in the knee: Current mr imaging techniques and applications in clinical practice and research”, *Radiographics*, vol. 31, no. 1, p. 37, 2011.
- [179] R. J. Schmitz, H.-M. Wang, D. R. Polprasert, R. A. Kraft, and B. G. Pietrosimone, “Evaluation of knee cartilage thickness: A comparison between ultrasound and magnetic resonance imaging methods”, *The Knee*, vol. 24, no. 2, pp. 217–223, 2017.
- [180] A. F. M. Hani, D. Kumar, A. S. Malik, R. Razak, and A. Kiffie, “Fusion of multinuclear magnetic resonance images of knee for the assessment of articular cartilage”, in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2013, pp. 6466–6469.
- [181] T. M. Link, J. Neumann, and X. Li, “Prestructural cartilage assessment using mri”, *Journal of Magnetic Resonance Imaging*, vol. 45, no. 4, pp. 949–965, 2017.



- [182] J. Johnston, B. Masri, and D. Wilson, “Computed tomography topographic mapping of subchondral density (ct-tomasd) in osteoarthritic and normal knees: Methodological development and preliminary findings”, *Osteoarthritis and cartilage*, vol. 17, no. 10, pp. 1319–1326, 2009.
- [183] M. A. Bowes, K. Kacena, O. A. Alabas, *et al.*, “Machine-learning, mri bone shape and important clinical outcomes in osteoarthritis: Data from the osteoarthritis initiative”, *Annals of the rheumatic diseases*, vol. 80, no. 4, pp. 502–508, 2021.
- [184] B. G. Ashinsky, M. Bouhrara, C. E. Coletta, *et al.*, “Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative”, *Journal of Orthopaedic Research*, vol. 35, no. 10, pp. 2243–2250, 2017.
- [185] F. Liu, Z. Zhou, A. Samsonov, *et al.*, “Deep learning approach for evaluating knee mr images: Achieving high diagnostic performance for cartilage lesion detection”, *Radiology*, vol. 289, no. 1, p. 160, 2018.
- [186] N. Bien, P. Rajpurkar, R. L. Ball, *et al.*, “Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of mrnet”, *PLoS medicine*, vol. 15, no. 11, e1002699, 2018.
- [187] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, “Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach”, *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.
- [188] A. A. Tolpadi, J. J. Lee, V. Padoia, and S. Majumdar, “Deep learning predicts total knee replacement from magnetic resonance images”, *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [189] Y. Du, R. Almajalid, J. Shan, and M. Zhang, “A novel method to predict knee osteoarthritis progression on mri using machine learning methods”, *IEEE transactions on nanobioscience*, vol. 17, no. 3, pp. 228–236, 2018.
- [190] A. Tiulpin, S. Klein, S. Bierma-Zeinstra, *et al.*, “Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data”, *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.
- [191] J. Massion, “Postural control system”, *Current opinion in neurobiology*, vol. 4, no. 6, pp. 877–887, 1994.
- [192] J. F. Golding, “Motion sickness susceptibility”, *Autonomic Neuroscience*, vol. 129, no. 1-2, pp. 67–76, 2006.
- [193] J. Bos, D. Damala, C. Lewis, A. Ganguly, and O. Turan, “Susceptibility to seasickness”, *Ergonomics*, vol. 50, no. 6, pp. 890–901, 2007.
- [194] D. Huppert, E. Grill, and T. Brandt, “Survey of motion sickness susceptibility in children and adolescents aged 3 months to 18 years”, *Journal of neurology*, vol. 266, no. 1, pp. 65–73, 2019.
- [195] H. Petersen, “Seasickness”, *Laeknabladid*, vol. 98, no. 12, pp. 653–659, 2012.
- [196] G. H. Crampton, *Motion and space sickness*. CRC Press, 1990.
- [197] J. J. LaViola Jr, “A discussion of cybersickness in virtual environments”, *ACM Sigchi Bulletin*, vol. 32, no. 1, pp. 47–56, 2000.

- [198] M. B. Flanagan, J. G. May, and T. G. Dobie, “Sex differences in tolerance to visually-induced motion sickness”, *Aviation, space, and environmental medicine*, vol. 76, no. 7, pp. 642–646, 2005.
- [199] A. Lawther and M. Griffin, “A survey of the occurrence of motion sickness amongst passengers at sea.”, *Aviation, space, and environmental medicine*, vol. 59, no. 5, pp. 399–406, 1988.
- [200] M. Turner and M. J. Griffin, “Motion sickness in public road transport: The effect of driver, route and vehicle”, *Ergonomics*, vol. 42, no. 12, pp. 1646–1664, 1999.
- [201] J. T. Reason, “Motion sickness adaptation: A neural mismatch model”, *Journal of the Royal Society of Medicine*, vol. 71, no. 11, pp. 819–829, 1978.
- [202] C. M. Oman and K. E. Cullen, “Brainstem processing of vestibular sensory exafference: Implications for motion sickness etiology”, *Experimental brain research*, vol. 232, no. 8, pp. 2483–2492, 2014.
- [203] B. Cohen, M. Dai, S. B. Yakushin, and C. Cho, “The neural basis of motion sickness”, *Journal of neurophysiology*, vol. 121, no. 3, pp. 973–982, 2019.
- [204] B. J. Yates, M. F. Catanzaro, D. J. Miller, and A. A. McCall, “Integration of vestibular and emetic gastrointestinal signals that produce nausea and vomiting: Potential contributions to motion sickness”, *Experimental brain research*, vol. 232, no. 8, pp. 2455–2469, 2014.
- [205] G. E. Riccio and T. A. Stoffregen, “An ecological theory of motion sickness and postural instability”, *Ecological psychology*, vol. 3, no. 3, pp. 195–240, 1991.
- [206] S. V. G. Cobb, “Measurement of postural stability before and after immersion in a virtual environment”, *Applied ergonomics*, vol. 30, no. 1, pp. 47–57, 1999.
- [207] D. Tal, R. Bar, Z. Nachum, A. Gil, and A. Shupak, “Postural dynamics and habituation to seasickness”, *Neuroscience letters*, vol. 479, no. 2, pp. 134–137, 2010.
- [208] A. Koohestani, D. Nahavandi, H. Asadi, *et al.*, “A knowledge discovery in motion sickness: A comprehensive literature review”, *IEEE access*, vol. 7, pp. 85 755–85 770, 2019.
- [209] A. Thurrell and A. Bronstein, “Vection increases the magnitude and accuracy of visually evoked postural responses”, *Experimental Brain Research*, vol. 147, no. 4, pp. 558–560, 2002.
- [210] S. Weech, J. Moon, and N. F. Troje, “Influence of bone-conducted vibration on simulator sickness in virtual reality”, *PLoS one*, vol. 13, no. 3, e0194137, 2018.
- [211] R. Laboissiere, J.-C. Letievant, E. Ionescu, P.-A. Barraud, M. Mazzuca, and C. Cian, “Relationship between spectral characteristics of spontaneous postural sway and motion sickness susceptibility”, *PLoS One*, vol. 10, no. 12, e0144466, 2015.
- [212] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness”, *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.
- [213] J. F. Golding, “Motion sickness”, *Handbook of clinical neurology*, vol. 137, pp. 371–390, 2016.

- [214] A. Mazloumi Gavgani, F. R. Walker, D. M. Hodgson, and E. Nalivaiko, “A comparative study of cybersickness during exposure to virtual reality and “classic” motion sickness: Are they different?”, *Journal of Applied Physiology*, vol. 125, no. 6, pp. 1670–1680, 2018.
- [215] X. Li, C. Zhu, C. Xu, J. Zhu, Y. Li, and S. Wu, “Vr motion sickness recognition by using eeg rhythm energy ratio based on wavelet packet transform”, *Computer Methods and Programs in Biomedicine*, vol. 188, p. 105266, 2020.
- [216] L.-W. Ko, C.-S. Wei, T.-P. Jung, and C.-T. Lin, “Estimating the level of motion sickness based on eeg spectra”, in *International Conference on Foundations of Augmented Cognition*, Springer, 2011, pp. 169–176.
- [217] Y. Li, A. Liu, and L. Ding, “Machine learning assessment of visually induced motion sickness levels based on multiple biosignals”, *Biomedical signal processing and control*, vol. 49, pp. 202–211, 2019.
- [218] Y. Wang, J.-R. Chardonnet, and F. Merienne, “Vr sickness prediction for navigation in immersive virtual environments using a deep long short term memory model”, in *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, IEEE, 2019, pp. 1874–1881.
- [219] S. Hell and V. Argyriou, “Machine learning architectures to predict motion sickness using a virtual reality rollercoaster simulation tool”, in *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, IEEE, 2018, pp. 153–156.
- [220] P. McCrory, N. Feddermann-Demont, J. Dvořák, *et al.*, “What is the definition of sports-related concussion: A systematic review”, *British journal of sports medicine*, vol. 51, no. 11, pp. 877–887, 2017.
- [221] J. A. Langlois, W. Rutland-Brown, and M. M. Wald, “The epidemiology and impact of traumatic brain injury: A brief overview”, *The Journal of head trauma rehabilitation*, vol. 21, no. 5, pp. 375–378, 2006.
- [222] M. B. Clay, K. L. Glover, and D. T. Lowe, “Epidemiology of concussion in sport: A literature review”, *Journal of chiropractic medicine*, vol. 12, no. 4, pp. 230–251, 2013.
- [223] P. McCrory, W. Meeuwisse, J. Dvorak, *et al.*, “Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in berlin, october 2016”, *British journal of sports medicine*, vol. 51, no. 11, pp. 838–847, 2017.
- [224] T. Covassin, R. Elbin, A. Bleecker, A. Lipchik, and A. P. Kontos, “Are there differences in neurocognitive function and symptoms between male and female soccer players after concussions?”, *The American journal of sports medicine*, vol. 41, no. 12, pp. 2890–2895, 2013.
- [225] S. E. Chancellor, E. S. Franz, O. V. Minaeva, and L. E. Goldstein, “Pathophysiology of concussion”, in *Seminars in Pediatric Neurology*, Elsevier, vol. 30, 2019, pp. 14–25.
- [226] J. K. Yue, R. R. Phelps, A. Chandra, E. A. Winkler, G. T. Manley, and M. S. Berger, “Sideline concussion assessment: The current state of the art”, *Neurosurgery*, vol. 87, no. 3, pp. 466–475, 2020.

- [227] R. J. Echemendia, W. Meeuwisse, P. McCrory, *et al.*, “The sport concussion assessment tool 5th edition (scat5): Background and rationale”, *British journal of sports medicine*, vol. 51, no. 11, pp. 848–850, 2017.
- [228] S. P. Broglio and T. W. Puetz, “The effect of sport concussion on neurocognitive function, self-report symptoms and postural control”, *Sports Medicine*, vol. 38, no. 1, pp. 53–67, 2008.
- [229] K. Dams-O’Connor, J. B. Cantor, M. Brown, M. P. Dijkers, L. A. Spielman, and W. A. Gordon, “Screening for traumatic brain injury: Findings and public health implications”, *The Journal of head trauma rehabilitation*, vol. 29, no. 6, p. 479, 2014.
- [230] M. W. Collins, A. P. Kontos, E. Reynolds, C. D. Murawski, and F. H. Fu, “A comprehensive, targeted approach to the clinical care of athletes following sport-related concussion”, *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 22, no. 2, pp. 235–246, 2014.
- [231] S. Slobounov, W. Sebastianelli, and M. Hallett, “Residual brain dysfunction observed one year post-mild traumatic brain injury: Combined eeg and balance study”, *Clinical Neurophysiology*, vol. 123, no. 9, pp. 1755–1761, 2012.
- [232] G. Dumkrieger, C. D. Chong, K. Ross, V. Berisha, and T. J. Schwedt, “Differentiating between migraine and post-traumatic headache using a machine learning classifier”, *Neurology*, vol. 98, no. 1 Supplement 1, S5–S6, 2022.
- [233] Y. Chu, G. Knell, R. P. Brayton, S. O. Burkhart, X. Jiang, and S. Shams, “Machine learning to predict sports-related concussion recovery using clinical data”, *Annals of physical and rehabilitation medicine*, vol. 65, no. 4, p. 101626, 2022.
- [234] C. K. Rosenblatt, A. Harriss, A.-N. Babul, and S. A. Rosenblatt, “Machine learning for subtyping concussion using a clustering approach”, *Frontiers in human neuroscience*, vol. 15, 2021.
- [235] R. M. Visscher, N. Feddermann-Demont, F. Romano, D. Straumann, and G. Bertolini, “Artificial intelligence for understanding concussion: Retrospective cluster analysis on the balance and vestibular diagnostic data of concussion patients”, *PloS one*, vol. 14, no. 4, e0214525, 2019.
- [236] Y. Cai, S. Wu, W. Zhao, Z. Li, Z. Wu, and S. Ji, “Concussion classification via deep learning using whole-brain white matter fiber strains”, *PloS one*, vol. 13, no. 5, e0197992, 2018.
- [237] D. E. Fleck, N. Ernest, R. Asch, *et al.*, “Predicting post-concussion symptom recovery in adolescents using a novel artificial intelligence”, *Journal of neurotrauma*, vol. 38, no. 7, pp. 830–836, 2021.
- [238] J. Castellanos, C. P. Phoo, J. T. Eckner, *et al.*, “Predicting risk of sport-related concussion in collegiate athletes and military cadets: A machine learning approach using baseline data from the care consortium study”, *Sports medicine*, vol. 51, no. 3, pp. 567–579, 2021.
- [239] F. Whalen, J. Sprung, C. M. Burkle, D. R. Schroeder, and D. O. Warner, “Recent smoking behavior and postoperative nausea and vomiting”, *Anesthesia & Analgesia*, vol. 103, no. 1, pp. 70–75, 2006.

- [240] C. Apfel, F. Heidrich, S. Jukar-Rao, *et al.*, “Evidence-based analysis of risk factors for postoperative nausea and vomiting”, *British journal of anaesthesia*, vol. 109, no. 5, pp. 742–753, 2012.
- [241] J. F. Golding, O. Prosyaniukova, M. Flynn, and M. A. Gresty, “The effect of smoking nicotine tobacco versus smoking deprivation on motion sickness”, *Autonomic Neuroscience*, vol. 160, no. 1-2, pp. 53–58, 2011.
- [242] P. Perrin, C. Perrot, D. Deviterne, B. Ragaru, and H. Kingma, “Dizziness in discus throwers is related to motion sickness generated while spinning”, *Acta oto-laryngologica*, vol. 120, no. 3, pp. 390–395, 2000.
- [243] Y.-C. Chen, T.-C. Tseng, T.-H. Hung, C. C. Hsieh, F.-C. Chen, and T. A. Stoffregen, “Cognitive and postural precursors of motion sickness in adolescent boxers”, *Gait & Posture*, vol. 38, no. 4, pp. 795–799, 2013.
- [244] R. Elbin, A. P. Kontos, A. Sufrinko, *et al.*, “Motion sickness susceptibility and baseline vestibular and ocular-motor performance in adolescent athletes”, *Journal of Athletic Training*, vol. 54, no. 9, pp. 939–944, 2019.
- [245] G. Caillet, G. Bosser, G. C. Gauchard, N. Chau, L. Benamghar, and P. P. Perrin, “Effect of sporting activity practice on susceptibility to motion sickness”, *Brain research bulletin*, vol. 69, no. 3, pp. 288–293, 2006.
- [246] Þ. Pétursson, K. J. Edmunds, M. K. Gíslason, *et al.*, “Bone mineral density and fracture risk assessment to optimize prosthesis selection in total hip replacement”, *Computational and mathematical methods in medicine*, vol. 2015, 2015.
- [247] M. Recenti, C. Ricciardi, M. Gíslason, K. Edmunds, U. Carraro, and P. Gargiulo, “Machine learning algorithms predict body mass index using nonlinear trimodal regression analysis from computed tomography scans”, in *Mediterranean Conference on Medical and Biological Engineering and Computing*, Springer, 2019, pp. 839–846.
- [248] M. Recenti, C. Ricciardi, A. Monet, *et al.*, “Predicting body mass index and isometric leg strength using soft tissue distributions from computed tomography scans”, *Health and Technology*, vol. 11, no. 1, pp. 239–249, 2021.
- [249] H. Drexler, U. Riede, T. Münzel, H. König, E. Funke, and H. Just, “Alterations of skeletal muscle in chronic heart failure.”, *Circulation*, vol. 85, no. 5, pp. 1751–1759, 1992.
- [250] M. J. Haykowsky, E. J. Kouba, P. H. Brubaker, B. J. Nicklas, J. Eggebeen, and D. W. Kitzman, “Skeletal muscle composition and its relation to exercise intolerance in older patients with heart failure and preserved ejection fraction”, *The American journal of cardiology*, vol. 113, no. 7, pp. 1211–1216, 2014.
- [251] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, “Heart failure: Diagnosis, severity estimation and prediction of adverse events through machine learning techniques”, *Computational and structural biotechnology journal*, vol. 15, pp. 26–47, 2017.
- [252] E. K. Oikonomou, M. C. Williams, C. P. Kotanidis, *et al.*, “A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary ct angiography”, *European heart journal*, vol. 40, no. 43, pp. 3529–3543, 2019.

- [253] S. Mendis, P. Puska, B. Norrving, W. H. Organization, *et al.*, *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.
- [254] R. Cubbon, A. Woolston, B. Adams, *et al.*, “Prospective development and validation of a model to predict heart failure hospitalisation”, *Heart*, vol. 100, no. 12, pp. 923–929, 2014.
- [255] M. Soret, S. L. Bacharach, and I. Buvat, “Partial-volume effect in pet tumor imaging”, *Journal of nuclear medicine*, vol. 48, no. 6, pp. 932–945, 2007.
- [256] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine learning and data mining methods in diabetes research”, *Computational and structural biotechnology journal*, vol. 15, pp. 104–116, 2017.
- [257] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in python”, *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [258] S. Shetty and Y. Rao, “Svm based machine learning approach to identify parkinson’s disease using gait analysis”, in *2016 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, vol. 2, 2016, pp. 1–5.
- [259] H. Gaßner, D. Jensen, F. Marxreiter, *et al.*, “Gait variability as digital biomarker of disease severity in huntington’s disease”, *Journal of neurology*, vol. 267, no. 6, pp. 1594–1601, 2020.
- [260] H. Park, S. Hong, I. Hussain, D. Kim, Y. Seo, and S. J. Park, “Gait monitoring system for stroke prediction of aging adults”, in *International Conference on Applied Human Factors and Ergonomics*, Springer, 2019, pp. 93–97.
- [261] T. Matsuura, K. Sakashita, A. Grushnikov, F. Okura, I. Mitsugami, and Y. Yagi, “Statistical analysis of dual-task gait characteristics for cognitive score estimation”, *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.
- [262] H.-J. An, J.-I. Kim, Y.-R. Kim, *et al.*, “The effect of various dual task training methods with gait on the balance and gait of patients with chronic stroke”, *Journal of physical therapy science*, vol. 26, no. 8, pp. 1287–1291, 2014.
- [263] N. E. Fritz, F. M. Cheek, and D. S. Nichols-Larsen, “Motor-cognitive dual-task training in neurologic disorders: A systematic review”, *Journal of neurologic physical therapy: JNPT*, vol. 39, no. 3, p. 142, 2015.
- [264] Y. S. Park, K. Ji-Won, K. Yuri, and K. Moon-Seok, “Effect of age and sex on gait characteristics in the korean elderly people”, *Iranian journal of public health*, vol. 47, no. 5, pp. 666–673, 2018.
- [265] W.-S. Im, H.-S. Choe, M.-G. Jeong, T.-B. Ryu, and H.-U. Choe, “A comparison of gait characteristics between korean and western young people”, *Journal of the Ergonomics Society of Korea*, vol. 25, no. 2, pp. 33–42, 2006.
- [266] T. Ryu, H. S. Choi, H. Choi, and M. K. Chung, “A comparison of gait characteristics between korean and western people for establishing korean gait reference data”, *International journal of industrial ergonomics*, vol. 36, no. 12, pp. 1023–1030, 2006.
- [267] F. K. Ciliberti, L. Guerrini, A. E. Gunnarsson, *et al.*, “Ct-and mri-based 3d reconstruction of knee joint to assess cartilage and bone”, *Diagnostics*, vol. 12, no. 2, p. 279, 2022.

- [268] D. Shepherd and B. Seedhom, “Thickness of human articular cartilage in joints of the lower limb”, *Annals of the rheumatic diseases*, vol. 58, no. 1, pp. 27–34, 1999.
- [269] G. Cai, P. Otahal, F. Cicuttini, *et al.*, “The association of subchondral and systemic bone mineral density with osteoarthritis-related joint replacements in older adults”, *Osteoarthritis and Cartilage*, vol. 28, no. 4, pp. 438–445, 2020.
- [270] H. Bonakdari, J.-P. Pelletier, F. Abram, and J. Martel-Pelletier, “A machine learning model to predict knee osteoarthritis cartilage volume changes over time using baseline bone curvature”, *Biomedicines*, vol. 10, no. 6, p. 1247, 2022.
- [271] S. A. Stehle, R. Aubonnet, M. Hassan, *et al.*, “Predicting postural control adaptation measuring eeg, emg, and center of pressure changes: Biovrsea paradigm”, *Frontiers in Human Neuroscience*, vol. 16, 2022.
- [272] R. Aubonnet, A. Shoykhet, D. Jacob, G. Di Lorenzo, H. Petersen, and P. Gargiulo, “Postural control paradigm (biovrsea): Towards a neurophysiological signature”, *Physiological Measurement*, vol. 43, no. 11, p. 115 002, 2022.
- [273] D. Jacob, R. Aubonnet, M. Recenti, *et al.*, “Assessing early-stage parkinson’s disease using biovrsea”, in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroX-RAINE)*, IEEE, 2022, pp. 271–276.
- [274] A. Petel, D. Jacob, R. Aubonnet, *et al.*, “Motion sickness susceptibility and visually induced motion sickness as diagnostic signs in parkinson’s disease”, *European Journal of Translational Myology*, vol. 32, no. 4, 2022.
- [275] J. F. Golding, “Predicting individual differences in motion sickness susceptibility by questionnaire”, *Personality and Individual Differences*, vol. 41, no. 2, pp. 237–248, 2006.
- [276] C. J. Dakin and D. A. Bolton, “Forecast or fall: Prediction’s importance to postural control”, *Frontiers in neurology*, vol. 9, p. 924, 2018.
- [277] S. A. Balk, D. B. Bertola, and V. W. Inman, “Simulator sickness questionnaire: Twenty years later”, in *Driving Assesment Conference*, University of Iowa, vol. 7, 2013.
- [278] F. Barollo, R. Friðriksdóttir, K. J. Edmunds, *et al.*, “Postural control adaptation and habituation during vibratory proprioceptive stimulation: An hd-eeg investigation of cortical recruitment and kinematics”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1381–1388, 2020.
- [279] K. J. Edmunds, H. Petersen, M. Hassan, *et al.*, “Cortical recruitment and functional dynamics in postural control adaptation and habituation during vibratory proprioceptive stimulation”, *Journal of neural engineering*, vol. 16, no. 2, p. 026 037, 2019.
- [280] R. Maire, A. Mallinson, H. Ceyte, *et al.*, “Discussion about visual dependence in balance control: European society for clinical evaluation of balance disorders”, 2017.
- [281] K.-M. Jang, Y. Shin Woo, and H. Kyoon Lim, “Electrophysiological changes in the virtual reality sickness: Eeg in the vr sickness”, in *The 25th International Conference on 3D Web Technology*, 2020, pp. 1–4.

- [282] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim, “Characteristic changes in the physiological components of cybersickness”, *Psychophysiology*, vol. 42, no. 5, pp. 616–625, 2005.
- [283] P. Perrin, A. Mallinson, C. Van Nechel, *et al.*, “Defining clinical-posturographic and intra-posturographic discordances: What do these two concepts mean?”, *The journal of international advanced otology*, vol. 14, no. 1, p. 127, 2018.
- [284] M. Shafeie, N. Zolfaghari, and K. M. V. McConville, “Abdominal muscle behavior and motion sickness during paired visual input with roll motion”, in *2013 6th International Conference on Human System Interactions (HSI)*, IEEE, 2013, pp. 247–251.
- [285] D. Tal, S. Shemy, G. Kaminski-Graif, G. Wiener, and D. Hershkovitz, “Vestibular evoked myogenic potentials and motion sickness medications”, *Clinical Neurophysiology*, vol. 127, no. 6, pp. 2350–2354, 2016.
- [286] B. J. Yates, P. S. Bolton, and V. G. Macefield, “Vestibulo-sympathetic responses”, *Comprehensive Physiology*, vol. 4, no. 2, p. 851, 2014.
- [287] R. Boshra, K. Dhindsa, O. Boursalie, *et al.*, “From group-level statistics to single-subject prediction: Machine learning detection of concussion in retired athletes”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1492–1501, 2019.
- [288] B. E. Gavett, R. A. Stern, and A. C. McKee, “Chronic traumatic encephalopathy: A potential late effect of sport-related concussive and subconcussive head trauma”, *Clinics in sports medicine*, vol. 30, no. 1, pp. 179–188, 2011.
- [289] C. M. Baugh, J. M. Stamm, D. O. Riley, *et al.*, “Chronic traumatic encephalopathy: Neurodegeneration following repetitive concussive and subconcussive brain trauma”, *Brain imaging and behavior*, vol. 6, no. 2, pp. 244–254, 2012.
- [290] M. Recenti, C. Ricciardi, K. Edmunds, D. Jacob, M. Gambacorta, and P. Gargiulo, “Testing soft tissue radiodensity parameters interplay with age and self-reported physical activity”, *European Journal of Translational Myology*, vol. 31, no. 3, 2021.
- [291] K. J. Edmunds, O. C. Okonkwo, S. Sigurdsson, *et al.*, “Soft tissue radiodensity parameters mediate the relationship between self-reported physical activity and lower extremity function in ages-reykjavik participants”, *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [292] F. Doná, C. Aquino, J. Gazzola, *et al.*, “Changes in postural control in patients with parkinson’s disease: A posturographic study”, *Physiotherapy*, vol. 102, no. 3, pp. 272–279, 2016.
- [293] C. C. Harro, A. Kelch, C. Hargis, and A. DeWitt, “Comparing balance performance on force platform measures in individuals with parkinson’s disease and healthy adults”, *Parkinson’s Disease*, vol. 2018, 2018.
- [294] K. R. Foster, R. Koprowski, and J. D. Skufca, “Machine learning, medical diagnosis, and biomedical engineering research-commentary”, *Biomedical engineering online*, vol. 13, pp. 1–9, 2014.
- [295] W. Luo, D. Phung, T. Tran, *et al.*, “Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view”, *Journal of medical Internet research*, vol. 18, no. 12, e323, 2016.



- [296] A. Sakai, Y. Minoda, and K. Morikawa, “Data augmentation methods for machine-learning-based classification of bio-signals”, in *2017 10th Biomedical Engineering International Conference (BMEiCON)*, 2017, pp. 1–4. DOI: 10.1109/BMEiCON.2017.8229109.
- [297] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, “Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary”, *Journal of artificial intelligence research*, vol. 61, pp. 863–905, 2018.
- [298] P. Figueiredo, E. A. Marques, V. Gudnason, *et al.*, “Computed tomography-based skeletal muscle and adipose tissue attenuation: Variations by age, sex, and muscle”, *Experimental gerontology*, vol. 149, p. 111306, 2021.
- [299] P. Drid, M. Drapsin, T. Trivic, D. Lukač, S. Obadov, and Z. Milosevic, “Asymmetry of muscle strength in elite athletes”, *Biomedical Human Kinetics*, vol. 1, no. 2009, pp. 3–5, 2009.
- [300] G. T. Mangine, D. H. Fukuda, M. B. LaMonica, *et al.*, “Influence of gender and muscle architecture asymmetry on jump and sprint performance”, *Journal of sports science & medicine*, vol. 13, no. 4, p. 904, 2014.
- [301] B. Keshavarz and J. F. Golding, “Motion sickness: Current concepts and management”, *Current opinion in neurology*, vol. 35, no. 1, pp. 107–112, 2022.
- [302] T. C. Hain, P. A. Hanna, and M. A. Rheinberger, “Mal de débarquement”, *Archives of Otolaryngology–Head & Neck Surgery*, vol. 125, no. 6, pp. 615–620, 1999.







Department of Engineering, School of Technol-  
ogy

Reykjavík University

Menntavegur 1

102 Reykjavík, Iceland

Tel. +354 599 6200

Fax +354 599 6201

[www.ru.is](http://www.ru.is)

ISBN: 978-9935-539-07-6 Electronic version

ISBN: 978-9935-539-06-9 Print version

ORCID Marco Recenti 0000-0001-9440-8434

<https://orcid.org/0000-0001-9440-8434>