



Stefanov, Kristian Ivanov (2024) *Harnessing brain imaging data to personalise management of fatigue in inflammatory arthritis*. PhD thesis.

<https://theses.gla.ac.uk/84203/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University of Glasgow

Harnessing brain imaging data to personalise management of fatigue in inflammatory arthritis.

By

Kristian Ivanov Stefanov

BSc (Neuroscience), University of Glasgow

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

November 2023

School of Infection and Immunity

College of Medical, Veterinary and Life Sciences

University of Glasgow

Abstract

Rheumatoid arthritis and psoriatic arthritis are chronic inflammatory conditions in which chronic fatigue persists in the majority of patients despite successful management of disease activity. This multidimensional, disabling fatigue correlates with various brain characteristics. Current treatments inadequately address fatigue, emphasising the importance of exploring its neural underpinnings and what potential imaging the brain has to inform the management of fatigue in these inflammatory arthritis conditions. To do so, I applied brain measures to stratify inflammatory arthritis patients into fatigue-related subgroups with potentially amendable biological differences, identify correlates of different subdimensions of fatigue, and predict fatigue follow-up after fatigue-specific or pharmacological treatments in different inflammatory arthritis cohorts of rheumatoid and psoriatic arthritis. I hypothesised that there are (1) subtypes of fatigue in patients with rheumatoid arthritis, illustrated by distinct subgroups stratified by a relationship between neuroimaging brain characteristics and fatigue; (2) statistically significant correlates of subcomponents of fatigue; (3) statistically significant predictors of fatigue scores after non-pharmacological treatments in rheumatoid arthritis; (4) statistically significant predictors of fatigue scores after pharmacological treatments in rheumatoid and psoriatic arthritis; (5) models that can predict individual fatigue outcomes above chance in a trial of non-pharmacological treatments in rheumatoid arthritis using machine learning to combine multiple neuroimaging and clinical variables.

I found a link between neuroimaging brain connectivity and distinct subgroups in rheumatoid arthritis related to fatigue subdimensions, albeit only within a specific cohort. Associations emerged between brain imaging metrics and baseline fatigue subcomponents, showing varied correlations with different metrics. In rheumatoid arthritis patients undergoing exercise or cognitive-behavioural interventions, baseline brain imaging predictors of fatigue centred on structural connectivity from the precuneus to the anterior cingulate cortex. In contrast, I did not find significant neuroimaging predictors of fatigue in rheumatoid arthritis patients who started a new disease-modifying antirheumatic drug. However, I did find such predictors in psoriatic arthritis patients, encompassing cortical thickness of the visual pericalcarine cortex and functional connectivity within the default mode and salience networks, involving the inferior parietal lobule and anterior cingulate cortex. Finally, models using diverse neuroimaging and clinical modalities along with different machine learning algorithms outperformed models using solely the baseline median

fatigue. Significantly, these models did not surpass chance level or replicate their utility in usual care patients in an independent rheumatoid arthritis cohort. Overall, despite not finding a model that can predict individual fatigue outcomes, this research advanced our understanding by pinpointing different fatigue-related brain circuits, delineating associations with subcomponents, and identifying group-level predictors of fatigue. If such findings are utilised by future studies using molecular and brain stimulation techniques, neuroimaging can offer innovative solutions to patients to significantly improve their quality of life.

Dedication

This work is dedicated to

My beloved parents

My big brother

My best friend

And my grandad

Посвещавам този труд на

Родителите ми Диана и Иван

Големя ми брат Стефан

Най-добрия ми приятел Добри

И на дядо ми Кръстю

Почивай в мир

Acknowledgements

I would like to express my gratitude to the many people that motivated and shaped this PhD journey.

I am deeply grateful to my supervisors: Neil, Gordon, and Jonathan. Neil, you've been the driving force behind my academic journey. I greatly appreciate how you consistently give me sound advice and enable me to play to my strengths but always push me to see the bigger picture and shape me as a more complete person in science and life. Most of all, I am grateful for all the times you intuitively recognised when I needed help, offering assistance both professionally and personally, and providing a reassuring presence at every step I took. Gordon, thank you for always being the patient and down-to-earth person you are whenever I've reached out for support. You've taught me how to navigate between focusing on details and understanding the clinical context of each problem I encountered. Jonathan, thank you for guiding me in approaching each question scientifically, fulfilling the role of the always approachable supervisor, and reassuring me of my contributions to the team.

I would like to highlight Salim as the person without whom this PhD would never have been possible. You are the person who introduced me to the world of MRI, but you've done so much more as a friend of seven years. Working with you has honed my analytical skills, engaging in debates with you has sharpened my critical thinking, and our conversations have deepened my understanding of what it means to have a true friend in this academic environment. Thank you for all you have done for me.

I also must thank the team I have been so happy to work with, which includes Flavia, Norah, Joel, Maxine, James, Andrew and more recently, Tyron. Thank you all for your support throughout this PhD. Flavia, you've been my PhD counterpart, reminding me I am not alone in this endeavour and always surprising me at how quickly one can learn and adapt to a totally distant field to their own. Norah, I appreciate your wholehearted efforts to help me at every possible moment and take inspiration from you, who never gives up when out of your comfort zone. Joel, thank you for introducing me to the world of TMS and patient interaction and always having my back when needed. Finally, Maxine can always put a smile on my face with her sense of humour but also with her ability to look after and protect both me and the team.

I would also like to thank our collaborators at the University Michigan for their support in all aspects of neuroimaging and warm welcome during my short visit. This of course includes Chelsea, Steve, Dan, Andrew and Noah and Tony as well.

(To my family and friends from back home) Трудно е да си далеч от дома. Още по-трудно е, когато си в различна държава. Това, което обаче прави този опит почти невъзможен, е чувството, че не можеш да бъдеш до най-близките си хора през най-хубавите и най-тежките моменти. Мамо, тате, брат ми, вашата обич, отдаденост и вечна подкрепа са източник на моята увереност, стремеж да градя и решителност да остана верен на себе си. Благодаря ви. На дядо, адаш, пиша това с надеждата винаги да следвам твоя пример, първи и несравним. Без Добри, нямаше да мога да съхраня здравия си разум, нито усмивката на лицето си; тази докторантура е благодарение и на теб. Ламбрина, още един човек, без когото докторантурата нямаше да бъде възможна. Благодаря ти, че ми помогна да стъпя на краката си тук. На приятелите ми от България, които станаха едни от най-близките ми, когато неочаквано се озоваха в Шотландия: Ники, винаги ме зареждаш с позитивна енергия и хляб, и Йони, винаги способна да ме развеселиш. Искам още да благодаря на семейство Бошеви - чувствах се като у дома си с вас. И благодаря на другите ми по-близки приятели, Свилена и Юлия, за подкрепата ви през добрите и лошите моменти, изразена във всеки разговор през годините.

Table of Contents

Dedication	iv
Acknowledgements	v
List of tables	xiv
List of figures	xvi
Author's Declaration	xx
Definitions/abbreviations	xxi
Chapter 1 Introduction	1
1.1 Overview	1
1.2 Inflammatory Arthritis	1
1.2.1 Description	1
1.2.2 Epidemiology	2
1.2.3 Pathophysiology	2
1.2.4 Symptoms.....	4
1.2.5 Classification criteria	4
1.2.6 Differences	4
1.2.7 Treatments & outcomes	5
1.3 Fatigue.....	7
1.3.1 Role in inflammatory arthritis.....	7

1.3.2 Patient perspective	8
1.3.3 Sources and types of fatigue	8
1.3.4 Interventions, correlates and predictors of fatigue	9
1.3.5 The brain and fatigue	13
1.4 Brain Anatomy & Function.....	16
1.4.1 Cerebrum.....	16
1.4.2 Cerebellum	18
1.4.3 Brain Stem.....	18
1.5 Basis of Magnetic Resonance Imaging	19
1.5.1 Structural MRI	20
1.5.2 Functional MRI.....	20
1.5.3 Diffusion Tensor Imaging (DTI).....	23
1.5.4 Brain atlases	25
1.6 Neurobiological markers of fatigue in inflammatory arthritis	28
1.7 Machine learning & prognosis	32
1.8 Aims, rationales, and hypotheses	37
Chapter 2 General Methods	45
2.1 Datasets	45
2.1.1 RA study dataset 1	45

2.1.2 RA trial dataset (LIFT).....	47
2.1.3 PsA study dataset	49
2.1.4 RA study dataset 2	50
2.2 Preprocessing methods.....	52
2.2.1 Realignment	52
2.2.2 Slice-timing correction.....	52
2.2.3 Outlier-detection	52
2.2.4 Coregistration.....	53
2.2.5 Segmentation and normalization (MNI-space)	53
2.2.6 Smoothing	53
2.2.7 Subject-space preprocessing	53
2.2.8 Denoising	56
2.2.9 Diffusion Tensor Imaging (DTI).....	58
2.3 General Linear Model (GLM).....	58
2.4 Prognostic model building and evaluation.....	61
2.4.1 Bias-variance trade-off.....	61
2.4.2 Feature selection.....	62
2.4.3 Machine learning algorithms.....	64
2.4.4 Model evaluation.....	75

2.4.5 Performance metrics.....	77
2.5 Model interpretation.....	79
2.6 Summary	82
Chapter 3 Identifying subgroups of fatigue	83
3.1 Introduction.....	83
3.2 Methods.....	85
3.2.1 Patient recruitment and MRI preprocessing.....	85
3.2.2 Dynamic fMRI connectivity	85
3.2.3 Canonical correlation and clustering analyses	88
3.2.4 Statistical analyses	89
3.3 Results.....	90
3.3.1 Patient characteristics.....	90
3.3.2 Dynamic connectivity is associated with dimensions of fatigue in the RA study	91
3.3.3 Reliable clusters from canonical connectivity scores	93
3.3.4 Baseline clusters did not differentiate on clinical traits	95
3.3.5 Failed connectivity-fatigue association and cluster replication in LIFT.....	96
3.3.6 Integral brain regions for clustering patients	97
3.4 Discussion	99
Chapter 4 Neuroimaging markers of subdimensions of fatigue	103

4.1 Introduction	103
4.2 Methods.....	106
4.3 Results	108
4.4 Discussion	114
Chapter 5 Predictors of fatigue in RA after non-pharmacological interventions	118
5.1 Introduction	118
5.2 Methods.....	120
5.2.1 Pre-selected approach.....	120
5.2.2 Agnostic approach.....	120
5.3 Results	122
5.3.1 Patient characteristics and treatment effects	122
5.3.2 Previous fatigue correlates were not predictive of post-treatment fatigue.....	123
5.3.3 Structural and PASAT connectivity fatigue predictors.....	124
5.3.4 Results across treatment groups did not overlap between groups.....	127
5.4 Discussion	128
Chapter 6 Predictors of fatigue in RA and PsA after pharmacological interventions ..	131
6.1 Introduction	131
6.2 Methods.....	133
6.3 Results	135

6.4 Discussion	138
Chapter 7 Prognostic models of fatigue of non-pharmacological interventions in RA	141
7.1 Introduction	141
7.2 Methods.....	144
7.2.1 Datasets and modalities.....	144
7.2.2 Feature selection.....	145
7.2.3 Algorithm and parameter tuning	148
7.2.4 Imputation of missing values	150
7.2.5 Model evaluation, comparisons, and significance	153
7.3 Results	155
7.3.1 Patient characteristics.....	155
7.3.2 All model descriptions	156
7.3.3 Best model descriptions	159
7.3.4 Best model significance and validation	162
7.3.5 Model interpretation.....	163
7.4 Discussion	168
Chapter 8 Conclusions	174
8.1 New knowledge.....	174
8.2 Limitations	178

8.3 Alternative approaches.....	179
8.4 Future directions.....	181
Contributions.....	183
Appendices	185
Appendix A CONN toolbox default atlas regions	185
Appendix B BRAF-MDQ	187
Appendix C Chalder Fatigue Severity	188
Appendix D Non-pharmacological predictors	189
Appendix E PROMIS Fatigue-FM Profile.....	194
Appendix F Pharmacological predictors.....	195
Appendix G Machine learning configuration	198
Appendix H Best performing features	199
Appendix I Feature effect on fatigue	207
List of references.....	210

List of tables

Table 1-1: Comparison table for rheumatoid and psoriatic arthritis.	6
Table 1-2: Summary table of correlates and predictors of fatigue.	13
Table 1-3: Brain networks affiliated to regions in the Desikan-Killiany atlas.	27
Table 1-4: Summary table of MRI correlates of fatigue in inflammatory arthritis.	31
Table 1-5: Advantages and disadvantages of machine learning.	37
Table 2-1: MRI sequences of the RA trial (LIFT) dataset.	48
Table 2-2: General linear model design in the functional connectivity CONN toolbox.	60
Table 2-3: Description of prediction algorithm types.	65
Table 2-4: Overview of advantages and disadvantages of selected machine learning algorithms.	73
Table 3-1: Clinical characteristics of patients in the RA study and LIFT with PASAT fMRI as Median (IQR).	90
Table 3-2: Differences between the included and excluded patients from RA study 1 as Median (IQR).	91
Table 3-3: Cluster index criteria values and p-values after comparisons with simulated data.	93
Table 3-4: Clinical characteristics of RA study clusters Median (IQR).	95
Table 4-1: Morphometric correlates of fatigue subdimensions.	109
Table 4-2: Structural connectivity correlates of fatigue subdimensions.	110

Table 4-3: Resting-state functional connectivity correlates of fatigue subdimensions.....	111
Table 4-4: PASAT functional connectivity correlates of fatigue subdimensions.....	112
Table 5-1: Clinical characteristics of patients in the LIFT trial as Median (IQR).	122
Table 5-2: Comparison between patients with completed and missing follow-up data as Median (IQR).	122
Table 5-3: Agnostic structural connectivity predictors of future fatigue.	125
Table 5-4: Agnostic PASAT functional connectivity predictors of future fatigue.	125
Table 5-5: Analysis in single treatment groups.....	127
Table 6-1: Clinical characteristics of patients with rheumatoid and psoriatic arthritis as Median (IQR).	135
Table 6-2: Comparison between patients with completed and missing follow-up data in cohorts starting a new DMARD treatment as Median (IQR).	135
Table 6-3: Neuroimaging predictors of fatigue after pharmacological interventions in psoriatic arthritis.....	136
Table 7-1: Alternative runs of cross-validation.	147
Table 7-2: Machine learning algorithms and hyperparameter search.....	150
Table 7-3: Configuration of the best model for each modality.....	156
Table 7-4: Best-performing multimodal models using resting-state and PASAT functional connectivity.....	163
Table 8-1: Summary of brain regions found across thesis results and their affiliated brain networks.....	177

List of figures

Figure 1-1: Pathophysiology of RA and PsA.....	3
Figure 1-2: Theoretical model of fatigue as a multifaceted phenomenon and its management.	15
Figure 1-3: The four lobes of the cerebral cortex.	16
Figure 1-4: Structures of the subcortex.....	17
Figure 1-5: Layer organisation of neuronal columns.....	18
Figure 1-6: Cerebellum and Brainstem.....	19
Figure 1-7: Structural MRI of tissue types.....	20
Figure 1-8: Estimation of functional connectivity.....	22
Figure 1-9: Diffusion Tensor Imaging (DTI).....	24
Figure 1-10: Clinical application of diffusion tensor imaging (DTI).	24
Figure 1-11: Desikan-Killiany atlas.....	26
Figure 1-12: : CONN toolbox default anatomical atlas.....	28
Figure 1-13: Machine learning methods to detect non-linear relationships.....	34
Figure 2-1: Flow chart of RA study 1.....	46
Figure 2-2: Flow chart of RA trial.....	49
Figure 2-3: Flow chart of PsA study.....	50
Figure 2-4: Flow chart of RA study 2.....	51

Figure 2-5: Subject-space preprocessing pipelines.....	55
Figure 2-6: MNI-and subject-space processing of functional imaging data.....	57
Figure 2-7: Elements of linear regression.....	58
Figure 2-8: Bias-variance trade-off based on model fit.....	62
Figure 2-9: Relief feature selection.....	64
Figure 2-10: Regularisation techniques.....	66
Figure 2-11: Support vector methods.....	68
Figure 2-12: Tree-based methods.....	70
Figure 2-13: Samples from a Gaussian process prior and posterior.....	72
Figure 2-14: Strategies to evaluate models on predictive performance.....	75
Figure 2-15: Repeated nested ten-fold cross-validation.....	77
Figure 2-16: Example of ALE plots.....	81
Figure 2-17: Summary of general methods and chapter specific additions.....	82
Figure 3-1: Analysis pipeline for identifying patient subgroups.....	86
Figure 3-2: Summary illustration of a canonical correlation analysis (CCA) and clustering.....	88
Figure 3-3: Patients in the RA study are different from patients in the RA trial (LIFT).....	91
Figure 3-4: Significant canonical correlations at baseline and trend level at follow-up.....	92
Figure 3-5: Null distributions and real values of the ptbiserial clustering index for baseline and follow-up data.....	94

Figure 3-6: The obtained 4-cluster solution for the RA study.	95
Figure 3-7: Failed replication in the LIFT (RA trial).....	96
Figure 3-8: Brain connections that contribute to patient clustering at baseline.	97
Figure 3-9: Brain connections that contribute to patient clustering at follow-up.	98
Figure 4-1: Correlation matrix and scatterplots of fatigue subdimensions.	108
Figure 4-2: Brain connectivity correlates with subdimensions of fatigue.	113
Figure 5-1: Treatment by time interaction in LIFT.....	123
Figure 5-2: Putamen volume analyses based on the RA study.	123
Figure 5-3: Dorsal attention network (DAN) connectivity with medial prefrontal cortex (mPFC) and fatigue.	124
Figure 5-4: Fatigue associations within and across treatment groups.	126
Figure 6-1: Neuroimaging predictors of fatigue after pharmacological interventions in psoriatic arthritis.....	137
Figure 7-1: Framework of multiple imputation within a repeated 10-fold cross-validation.	152
Figure 7-2: Fatigue characteristics in the trial and study RA cohorts.....	155
Figure 7-3: Cross-validation performance across modalities.....	157
Figure 7-4: Cross-validation performance across time.	158
Figure 7-5: Best model against baseline median performance.	159
Figure 7-6: Best model comparison across all modalities.	160
Figure 7-7: Best model outcome comparison.	161

Figure 7-8: Functional connectivity modality effect.....	161
Figure 7-9: Best model significance.	162
Figure 7-10: Feature importance.	164
Figure 7-11: Accumulated local effects (ALE) plots of the four most influential features and treatment allocation.	164
Figure 7-12: Visualisation of top imaging features.....	165
Figure 7-13: Model behaviour over the case with the greatest fatigue change.....	166
Figure 7-14: Model behaviour over the five cases with the greatest fatigue change.....	167

Author's Declaration

I, Kristian Ivanov Stefanov, declare that the work described in this thesis is original and was generated as a result of my own work. No part of this thesis has been submitted for any other degree, either at the University of Glasgow, or at any other institution.

Signature:

Printed Name: KRISTIAN IVANOV STEFANOV

Definitions/abbreviations

Cases

ALE: accumulated local effects	79
BDNF: brain derived neurotrophic factor	13
BOLD: blood oxygenation level-dependent	20
BRAF MDQ: Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire	88
CBA: cognitive-behavioural approach.....	41
CBT: cognitive-behavioural therapy	8
CCA: canonical correlation analysis.....	88
CNS: Central Nervous System.....	8
CONN toolbox: functional connectivity toolbox.....	59
CRP: C reactive protein	4
CSF: cerebrospinal fluid	16
DAN: dorsal attention network	29
DMARDS: disease modifying anti-rheumatic drugs	5
DMN: default mode network	100
DTI: diffusion tensor imaging.....	23
DWI: diffusion-weighted imaging	23
EPI: echo planar imaging	45

ESR: erythrocyte sedimentation rate.....	4
FA: fractional anisotropy	23
FC: functional connectivity.....	21
FDR: false discovery rate.....	60
fMRI: functional magnetic resonance imaging.....	20
FOV: field of view	45
GLM: general linear model.....	41
IA: inflammatory arthritis	1
IL: interleukin.....	4
JAK: Janus kinase	5
LIFT: Lessening the Impact of Fatigue in Inflammatory Rheumatic Diseases: A Randomised Trial	47
LIME: Local Interpretable Model-agnostic Explanations	81
MD: mean diffusivity	23
ML: machine learning	32
MNI: Montreal Neurological Institute	53
MRI: magnetic resonance imaging	19
MS: multiple sclerosis.....	9
NMDA: N methylD-aspartate	14
PASAT: Paced Auditory Serial Addition Test.....	46

PEP: personalised exercise programme	41
PNS: Peripheral Nervous System	8
PsA: Psoriatic Arthritis	1
RA: Rheumatoid Arthritis	1
RF: random forest	69
RMSE: root mean squared error	78
ROI: region-of-interest.....	21
SC: structural connectivity	58
SGCCA: sparse generalized canonical correlation analysis	145
SMN: somatosensory network	100
SVM: support vector machine	67
SVR: support vector regression	67
TE: echo time	45
TIV: total intracranial volume.....	107
TMS: transcranial magnetic stimulation	83
TR: repetition time	45
UC: usual care	41

Chapter 1 Introduction

1.1 Overview

Patients with inflammatory arthritis (IA) report chronic fatigue as multidimensional and disabling (Sumpton et al., 2020). Epidemiological studies reinforce fatigue as multifactorial with correlates in clinical and psychosocial domains (Doumen et al., 2022), but current treatments fail to alleviate fatigue (Walter et al., 2018). Neuroimaging has associated fatigue with brain characteristics of chronic fatigue populations, thus establishing the key role of the brain in fatigue processing (Goni et al., 2018). To extend these findings, this thesis will address the question of how brain imaging can deconstruct the heterogeneity of fatigue to provide pathophysiological insight and inform treatment stratification in IA. This will include applying brain measures to stratify IA patients into fatigue-related subgroups with potentially amendable biological differences, identify correlates of different subdimensions of fatigue, and predict fatigue follow-up after various treatments known to alleviate fatigue, in different IA cohorts of rheumatoid and psoriatic arthritis. In this thesis, prediction is defined as using some baseline characteristics to infer a future event or outcome. This chapter will describe the role of fatigue in IA, review neuroimaging findings in chronic fatigue, and discuss the current gaps in knowledge.

1.2 Inflammatory Arthritis

1.2.1 Description

There are multiple forms of arthritis, which vary in origin and phenotype. Osteoarthritis is the most prevalent kind of arthritis. Here joint cartilage is degraded, resulting in pain, stiffness, and immobility. Mechanical stress induces such changes either through injury or accumulated strain. In contrast, joint inflammation drives other common forms of IA, for example, rheumatoid arthritis (RA) and psoriatic arthritis (PsA). Patients with PsA also have psoriasis, itchy or painful patches of abnormal skin, in addition to their joint symptoms. Both are chronic autoimmune diseases in which the immune system incorrectly recognises joint constituents as harmful. This leads to inflammation not only in the joints but throughout the body (Merola et al., 2018).

1.2.2 Epidemiology

Epidemiological studies estimate the global prevalence of RA and PsA to be 0.46% (Almutairi et al., 2021) and 0.13% (Scotti et al., 2018), respectively, although the rates were recently recorded to be higher (0.779% and 0.287%) in England (Scott et al., 2022). Also, women are twofold to threefold more likely to develop RA than men, but men and women are equally likely to develop PsA (Scott et al., 2022). The risk of developing RA is influenced by genetics (60% heritability), smoking, and prior infections (Venetsanopoulou et al., 2023). People are more likely to develop PsA if they have relatives with the disease (80–100% heritability), more severe concomitant psoriasis, or obesity, but data for other factors, like smoking, is conflicting (Schett et al., 2022). In RA, 30% to 40% of people experience work disability five year after diagnosis and a third of them withdrawal from employment (Galloway et al., 2020). The impact of PsA has similar rates on work disability (16–39%) and unemployment (20–50%) (Cooksey et al., 2021). Finally, patients with RA and PsA are more likely to suffer and die from respiratory and cardiovascular comorbidities than the general population since they can worsen their disease activity and treatment response, as well as incur additional health costs (Kerola et al., 2022).

1.2.3 Pathophysiology

In RA (**Figure 1-1**), the immune system initially attacks the joint synovium, which disrupts the section of fluid from the synovium that nourishes the cartilage and lubricates the joints. Inflammation then spreads across the surrounding tissue, deforms the joint, and breaks down the bone and cartilage. One notable distinction of the many ways PsA differs from RA is that immune cells also infiltrate the entheses where the tendons or ligaments insert into the bone, as well as more commonly being asymmetrical and axial in distribution. Local inflammation degrades the joint in both disorders but can also spread systemically via pro-inflammatory signals (cytokines) throughout the body and manifest in symptoms beyond the joints, such as in the heart, blood vessels, and lungs.

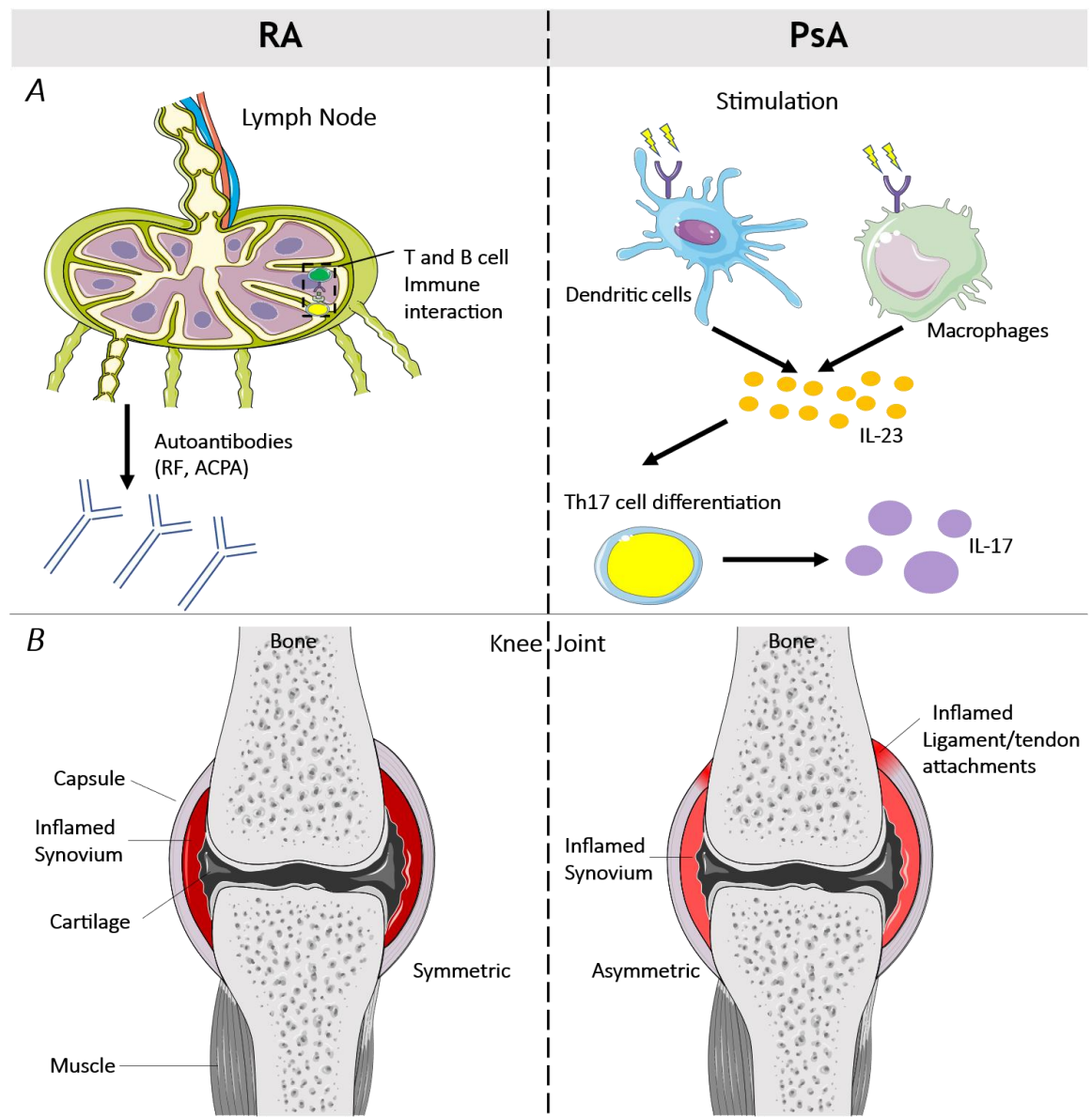


Figure 1-1: Pathophysiology of RA and PsA **A:** In early RA, autoantibodies can be detected as a result of interactions between T and B immune cells. These antibodies target either modified endogenous IgG antibodies (rheumatoid factor) or endogenous proteins that have been modified via citrullination (anti-citrullinated protein antibodies). These antibodies are not commonly detected in PsA but instead have characteristic T-helper cell 17 responses like secreting IL-17 due to stimulation from antigen-presenting cells such as dendritic cells and macrophages. **B:** The initial target of these immune responses in RA is the synovium, a membrane that produces synovial fluid for joint lubrication and cartilage nutrition. Although inflammation of the synovium can occur in PsA, the primary immune targets are the entheses (ligament/tendon attachment sites) on one side of the body, compared with joints on both sides in RA (symmetrical distribution). Despite these differences, the two processes similarly lead to inflammation at those sites and erode the bone and cartilage of the joints. Illustrations are based on Schett et al. (2017), while the figure was partly generated using Servier Medical Art, licensed under a Creative Commons Attribution 3.0 (unported license) and edited using Inkscape (2020). Abbreviations: ACPA, anti-citrullinated peptides antibodies; IgG, immunoglobulin G; IL-17/23, interleukin-17/23; PsA, psoriatic arthritis; RA, rheumatoid arthritis; RF, rheumatoid factor; Th17, T helper cell 17.

1.2.4 Symptoms

Both RA and PsA are progressive diseases characterised by swelling and stiffness of the joints, pain during motion and throughout the body, and a sense of debilitating fatigue. Women with RA exhibit symptoms around middle age, but they appear later in men (Alamanos et al., 2006). In PsA, psoriasis is present for typically 10 years before any symptoms in the joints occur, typically between 35 and 45 years of age, with PsA affecting 11–30% of psoriatic patients (Greb et al., 2016). However, the severity of skin symptoms is unrelated to the severity of joint symptoms.

1.2.5 Classification criteria

To classify RA, trials implement the American College of Rheumatology/ European League Against Rheumatism classification criteria (Studenic et al., 2023). Initially, chronic synovitis in at least one joint needs to be present in a patient. Afterwards, practitioners evaluate the number of affected joints and duration of symptoms, along with evidence of inflammatory markers in serum. These include disease-specific autoantibodies such as rheumatoid factor and antibodies against cyclic citrullinated peptide. Similarly, the classification criteria for PsA categorise patients according to their personal or family history of psoriasis, psoriatic nail dystrophy, and inflammation of the digits.

1.2.6 Differences

The two types of IA differ and overlap in symptoms, biomarkers, comorbidities, and potential treatments (**Table 1-1**). Inflammation of the spinal joints and nail dystrophy can occur in PsA but not in RA. PsA patients typically test negative for rheumatoid factor and cyclic citrullinated peptide antibodies in the blood, compared to approximately 80% of positive cases in RA patients (Verheul et al., 2015). Patients with RA also present with higher values of acute-phase inflammatory markers like C reactive protein (CRP) and erythrocyte sedimentation rate (ESR) while patients with PsA are more likely to be diabetic and obese (Labitigan et al., 2014). Although inflammation occurs in both RA and PsA, the amount of regulatory immune proteins (interleukins) that are increased differs (Mc Ardlle et al., 2015). Therapies exploit these differences, such as interleukin (IL)-12 and 23 inhibitors for PsA and IL-6 receptor inhibitors for RA (Veale and Fearon, 2015).

1.2.7 Treatments & outcomes

Care for IA involves both pharmacological and supportive treatments (Smolen et al., 2020, Gossec et al., 2020). Corticosteroids and non-steroidal anti-inflammatory drugs relieve symptoms of pain, stiffness, and inflammation. Disease-modifying anti-rheumatic drugs (DMARDs) control disease activity and the rate of joint damage in the early stages and may achieve remission (Glintborg et al., 2011). Although clinicians typically first prescribe synthetic DMARDs, newly developed biological DMARDs have drastically improved remission chances in patients who fail to respond to these therapies. They are composed of injectable “proteins” that target specific pathological immune cells or signals (Scott et al., 2010), similar to the most recent Janus kinase (JAK) inhibitors, which can be administered orally and have been shown to be just as effective (Bechman et al., 2019). Supportive interventions aim to improve joint and psychological health through exercise and weight loss programmes (van Zanten et al., 2015, Almodovar et al., 2018) and help patients adapt to their environment through occupational therapy (Macedo et al., 2009). Nevertheless, both types of care do not cure the disease, and for a large cohort of these patients, they are not optimal in managing symptoms such as widespread pain and chronic fatigue.

Characteristic	RA	PsA
Clinical		
Psoriasis	-	+++
Enthesitis	-	+++
Synovitis	+++	++
Serological		
Rheumatoid factor	+++	-
Cyclic citrullinated peptide antibodies	+++	-*
C-reactive protein	+++	++
Erythrocyte sedimentation rate	+++	++
TNF- α -driven	+++	+++
IL-17A-driven	-	+++
IL-12/23-driven	-	+++
IL-6-driven	+++	-
Comorbidities		
Infections	✓	✓
Osteoporosis	✓	✓
Overweight/obese	+	+++
Anxiety and depression	+	+++
Synthetic DMARDS		
Methotrexate	✓	✓
Hydroxychloroquine	✓	-
Corticosteroids	✓	-
Biological DMARDS		
TNF- α inhibitors (e.g., Adalimumab)	✓	✓
IL-17A inhibitors (e.g., Secukinumab)	-	✓
IL-12/23 inhibitors	-	✓
IL-6 receptor inhibitors	✓	-
Targeted synthetic oral small-molecule DMARDS		
Janus kinase 1/3 inhibitor (Tofacitinib)	✓	✓
Janus kinase 1/2 inhibitor (Baricitinib)	✓	-

Table 1-1: Comparison table for rheumatoid and psoriatic arthritis. The two diseases share or differ in characteristics of clinical representation, serological testing, and therapeutic strategies, which include synthetic, biologic, and small molecule DMARDS. Modified from Merola et al. (2018). Abbreviations: DMARD, disease-modifying antirheumatic drug; HLA, human leukocyte antigen; IL, interleukin; PsA, psoriatic arthritis; RA, rheumatoid arthritis; Th17, T helper cell 17; TNF, tumour necrosis factor. The number of plus signs (+, ++, +++) indicates the frequency of occurrence of the category, whereas the dash (-) indicates that the category is uncommon for that disease or that the therapy is not approved. *A small percentage of PsA patients can be positive (Punzi et al., 2007).

1.3 Fatigue

1.3.1 Role in inflammatory arthritis

Fatigue varies in chronic illnesses, but it lasts and is more intense than in healthy people. It does not require prior exertion, persists even after sleep or rest, and also overwhelms patients both physically and mentally. Fatigue is a prevalent and burdensome symptom of IA, but it is subjective and difficult to measure. Considered chronic when it lasts for three or more months (David et al., 1990), it affects approximately 80% and 49.5% of RA and PsA patients, respectively (Pollard et al., 2006, Husted et al., 2009). Furthermore, 50% of people with RA and 29% of people with PsA experience severe fatigue. However, due to different definitions and measures, these rates vary considerably (Seifert and Baerwald, 2019); for instance, an international survey study in rheumatic diseases (N=6120) found 57% of patients with PsA and 41% of patients with RA to be severely fatigued (Overman et al., 2016). Despite differences in how studies measure fatigue, they unequivocally find that fatigue, along with pain, debilitates patients with RA and PsA (Feldthusen et al., 2013, Gudu et al., 2016).

Fatigue not only negatively affects physical functioning but also psychological wellbeing and work outputs in IA. In RA (n=120), fatigue and emotional representation (an aspect of illness perception) explained 58% of the mental quality of life scores (SF-36), after adjusting for pain, disease activity, and social support (Berner et al., 2018). Therefore, treating fatigue is pivotal for improving the quality of life of these patients, irrespective of assessing other clinical and psychosocial characteristics. Fatigue was also independently associated with all aspects of work disability in 1747 patients with RA commencing the biologic DMARD etanercept (Druce et al., 2018). The disability aspects included presenteeism, activity impairment, productivity loss, and absenteeism, supporting the case for targeting fatigue in conjunction with disease-modifying treatments to optimise work-related outcomes and so provide greater societal returns. Another large cohort study found that PsA had similar effects to RA on quality of life and work disability when compared to healthy populations (Salaffi et al., 2009). Fatigue plays a key role in debilitating patients with PsA and is consequently part of the PsA Core Domain Set of mandatory measures, required in all PsA clinical trials and longitudinal observational studies (Orbai et al., 2017).

1.3.2 Patient perspective

IA patients contrast their fatigue with the “tiredness” that they had before the disease or with what healthy people experience. They differentiate their fatigue by its intensity, duration, and frequency, define it as “unearned” and “unresolving”, and outline the physical, cognitive, emotional, and social effects on their lives. Both people with RA and PsA describe it as extreme weariness and energy-draining, which could last from minutes to days and be very frequent (several times a week, daily) or even constant (Hewlett et al., 2005, Sumpton et al., 2020). Physically, fatigue leaves patients feeling heavy and weak, limiting their daily activities. Cognitively, they experience difficulties in learning, focusing, solving problems, and interacting with others. These effects leave patients with an emotional toll of feeling irritable, frustrated, and unmotivated, which restricts their roles in the family and wider social circle. Overall, a meta-analysis of qualitative studies in RA summarised fatigue in these patients as “A vicious circle of an unpredictable symptom” (Primdahl et al., 2019).

IA patients express that fatigue leaves them lonely, misunderstood, and without much professional support on how to handle these symptoms. Primdahl et al. (2019) also depicted that patients use variable language to explain their fatigue to others. However, patients felt that their fatigue was not understood and was dismissed by professionals. Furthermore, some people with RA thought their efforts at managing fatigue were unsuccessful (Hewlett et al., 2005), explaining findings of lower self-efficacy for coping with fatigue compared to other RA symptoms (Riemsma et al., 1998). Conversely, RA patients who had cognitive-behavioural therapy (CBT) increased their self-efficacy in managing their fatigue and improved their daily lives (Dures et al., 2012). Overall, the perspective of patients is by nature subjective and defines their fatigue as qualitatively different from tiredness and far-reaching into many domains that may not be picked up by standard questionnaires and clinical assessments.

1.3.3 Sources and types of fatigue

Fatigue due to the peripheral nervous system (PNS) involves alterations at the muscle or joint level that make movements difficult. Central fatigue refers to both physically and mentally exerting oneself but not being able to sustain movements or perform cognitive tasks due to changes at the level of the central nervous system (CNS) (Zwarts et al., 2008). People with IA suffer from both types of fatigue (Staud, 2012). As IA is primarily a disease of the

peripheral immune system, predicting the basis of fatigue related to these disorders has also focused on inflammation. Fatigue can increase with inflammatory disease flares (Pope, 2020), but serum concentrations of pro-inflammatory cytokines do not predict the intensity of fatigue, nor do they explain the similar intensity of fatigue in people with non-inflammatory diseases (Jaime-Lara et al., 2020).

Subjective perceptions of fatigue do not consistently correlate with performance on demanding tasks in both RA and other chronic diseases, including multiple sclerosis (MS) and chronic fatigue syndrome (Jennekensschinkel et al., 1988, Johnson et al., 1997, Beatty et al., 2003, do Espirito Santo et al., 2018). To address this discrepancy, some researchers have distinguished between trait and state fatigue (Genova et al., 2013). Trait fatigue is a stable form that expresses the global status of patients, which does not significantly change over time and is usually assessed with fatigue scales. State fatigue is a transient condition characterised by decreased performance during an acute but sustained effort, typically during a task. Overall, to accommodate both the perspective of the patient and the need for objective measures by researchers, both task performance and subjective measures of fatigue should be investigated (Marrelli et al., 2018).

1.3.4 Interventions, correlates and predictors of fatigue

Before assigning fatigue-specific treatments in the context of IA, physicians typically first address fatigue-related comorbidities such as anaemia and dysfunctional thyroid functioning. The next step to improving chronic fatigue would be to alleviate disease activity through pharmacological immune therapies (Almeida et al., 2016). In early RA, patients are relieved of fatigue if they achieve remission within three months (Holdren et al., 2019), but studies find that most who report fatigue at baseline and improve their disease state will remain chronically fatigued at follow-up (Walter et al., 2018). In established RA, up to 70% of patients with high disease activity and severe fatigue report a clinically significant reduction in fatigue following targeted immune therapy (Druce et al., 2015), and yet 62% of those patients who attain full disease remission following these advanced therapies still report significant fatigue (Druce et al., 2016).

The usual care of patients would also include educating them via booklets (e.g., Arthritis Research UK (2014)) or other means on how to handle fatigue, such as how to plan and prioritise their activities, pace themselves, and set achievable goals. Epidemiological studies have reported significant fatigue associations with cardiorespiratory fitness and psychosocial

factors, such as coping strategies and illness perception (Treharne et al., 2008), justifying the trialling of active interventions that lessen fatigue in some patients. This includes graded physical exercise tailored to the physical condition and motivation of the patient (van Zanten et al., 2015) and CBT. The latter targets negative beliefs (e.g., all or nothing) and behaviours (e.g., activity avoidance) and suggests alternatives to better manage their symptoms via written materials and professional consultations (Hewlett et al., 2015).

Both active types of interventions exhibit positive evidence for lowering fatigue, but further improvements are needed. Despite the small effect sizes, a meta-analysis discovered fatigue reduction in six studies on physical activity and 13 studies on psychosocial interventions in RA (Cramp et al., 2013). To test the applicability of these treatments in routine clinical practice, a recent trial of accessible fatigue-specific therapies was conducted in a range of inflammatory rheumatic diseases (Bachmair et al., 2022). Patients with IA received telephone-delivered physical activity support or cognitive behavioural approaches and achieved clinically and statistically significant improvements compared with an educational booklet. However, the overall effect sizes were moderate. The trial also followed patients for six months and one year after these interventions and found that patients retained their improvements. The treatments additionally improved mental health-related quality of life, sleep, work disability and depression, which may cumulatively raise general well-being. Such factors may maintain the persistence and impact of fatigue and often correlate with this symptom.

Disease activity and fatigue are weakly correlated in RA compared to stronger associations in PsA (Pollard et al., 2006, Gudu et al., 2016). However, biological DMARDS in both disorders only moderately decrease fatigue (Almeida et al., 2016, Reygaerts et al., 2018), which relates to multiple other factors. Fatigue has been associated with a lack of motivation, loss of appetite, and bodily pain in PsA (Sumpton et al., 2020). In RA, a study associated fatigue severity with pain and poor sleep quality but also with psychosocial factors such as role functioning, depressive mood, self-efficacy on fatigue, and helplessness (van Hoogmoed et al., 2010). Disease activity and tender joint count were moderately associated with fatigue, whereas inflammatory markers, swollen joint count, and anaemia were not. Psychosocial factors are amendable to change, such as through CBT, and were the focus of a systematic review in RA (Matcham et al., 2015). Mood was most consistently associated with fatigue among psychosocial variables, but at the multivariate level, several studies found non-significant associations between fatigue and pain, disability, and depression. Mood may dilute its influence on fatigue due to its collinearity with other variables such as

pain perception, inflammation, and socioeconomic variables. Disease-related and psychosocial measures may also interact with contextual factors, which range from managing co-morbidities to work and caring responsibilities, with similar conclusions drawn in PsA (Rosen et al., 2016). Overall, fatigue is associated with physical and psychosocial factors rather than inflammation alone, but longitudinal studies are required to track fatigue progression and how its associated factors interact with one another.

Factors that contribute to fatigue can be both time-dependent associates like inflammation that are modifiable with interventions, but also predisposing factors that could support early identification of at-risk patients who may not benefit from traditional biotherapy for fatigue (Goertz et al., 2021). Predisposing factors are identified through longitudinal studies like one trial in patients with early RA during their initial 24 months of treat-to-target therapy (Holten et al., 2022). The trial found that early disease remission decreased the risk of fatigue at follow-up. They also identified baseline predictors of fatigue at 24 months in low objective disease activity, including swollen joints, ultrasound power Doppler score, and a higher patient global assessment. Another study in early RA followed patients in a 2-year treat-to-target trial and its 3-year extension (Doumen et al., 2022). Only one in four patients made lasting improvements and 20% even experienced worsening multidimensional fatigue, and these patients also reported more pain and impaired mental health over time, irrespective of disease activity. Higher pain, patient global assessment, disability, lower mental components (SF-36), and fewer swollen joints at baseline predicted higher fatigue over 5 years, while early disease remission strongly improved 5-year fatigue, even if relapses occurred. Positive association between fatigue and disease activity as well as between early remission and resolved fatigue at follow-up may at first contradict the other finding that little inflammation at baseline predicted fatigue at follow-up. This contradiction could be explained by the existence of two subsets of fatigue: patients where high inflammatory disease activity was the cause of fatigue and for whom early, intensive treatment improved fatigue, and patients with fatigue at baseline for whom fatigue was triggered by different factors and captured by patient global assessment. The association between disease activity and fatigue in the second study was mediated by patient global assessment, pain, mental health, and sleep quality, implying an indirect relation between fatigue and inflammation. The presence of these predictors could thus prompt additional nonpharmacological approaches early on to improve fatigue. Overall, longitudinal studies demonstrate that IA fatigue may be a composite of inflammation-driven fatigue and fatigue with a stronger psychosocial background with the potential for different strategies at the beginning of their disease treatment.

Identifying both time-dependent and predisposing factors can inform clinical decisions on fatigue. However, individuals differ in which factors contribute to their fatigue, as well as the prominence and interplay of these factors. For instance, chronic fatigue occurs in conditions without any underlying inflammatory activity (Jaime-Lara et al., 2020), despite early disease remission being a strong predictor of fatigue improvement in RA. Furthermore, psychosocial factors that may predispose individuals to fatigue are measured through subjective self-reported responses to questionnaires. Treatment decisions for fatigue therefore require biomarkers—biological factors that are measured at baseline and predict future outcomes. One potential biomarker is central sensitization, or the amplification of sensory input across multiple sensory systems that can be characterised by amplified pain responses, unpleasant sensations to physical stimuli, and heightened sensitivity to environmental stimuli, including light and sound. Central sensitization describes enhanced sensitivity of the CNS to pain, which may also partially explain the high rates of fatigue across inflammatory and non-inflammatory disorders. Chronic fatigue and pain characterise fibromyalgia, the prototypical disorder of central sensitization (Boomershine, 2015), while IA patients with comorbid fibromyalgia are more likely to experience severe fatigue (Overman et al., 2016). In an unselected general population, a study predicted fatigue using baseline central sensitization, measured by a wind-up ratio test that compares the pain intensity after a single pinprick to that of ten consecutive pin pricks conducted at single-second intervals within a 1 cm² area (Druce and McBeth, 2019). The study showed that the central sensitization wind-up ratio at the hands and feet predicted fatigue a year later in 290 volunteers, independently of baseline musculoskeletal pain, fatigue, depression, anxiety, physical activity, body fat, and number of medications, thus suggesting the potential of central sensitization as a biomarker of fatigue. A summary table (**Table 1-2**) describes how the measure fits with other variables related to fatigue.

Type	Variables	Description	References
Associations	disease activity, tender joint count, lack of motivation, loss of appetite, bodily pain, sleep quality, role functioning, mood, self-efficacy on fatigue, helplessness, managing co-morbidities, caring responsibilities	variables that correlate with fatigue, identified through cross-sectional designs	(Sumpton et al., 2020), (Rosen et al., 2016), (Matcham et al., 2015), (van Hoogmoed et al., 2010)
Predictors	early disease remission, low objective disease activity, ultrasound power Doppler score, patient global assessment, pain, disability, mental components, wind-up ratio (central sensitization)	variables that predict fatigue at a later time point, identified through longitudinal designs	(Doumen et al., 2022), (Holten et al., 2022), (Druce and McBeth, 2019)
Mediators	patient global assessment, pain, mental health, sleep quality	variables that mediated an association between disease activity and fatigue, identified through a longitudinal design	(Doumen et al., 2022)

Table 1-2: Summary table of correlates and predictors of fatigue.

1.3.5 The brain and fatigue

Central sensitization may coincide with fatigue due to disrupted brain processing and plasticity, which is the ability of focal and widespread neuronal systems to reorganise their structure or function to better adapt to the internal or external environment (Poldrack, 2000). The role of the brain in central sensitization and potentially fatigue is demonstrated by altered brain derived neurotrophic factor (BDNF) levels in fibromyalgia that underline brain plasticity (Nugraha et al., 2012) as well as treatments that act on the CNS and improve central sensitization, such as antidepressants and gabapentinoids. Furthermore, non-pharmacological approaches that improve fatigue such as physical activity and behavioural-cognitive approaches are also applied in fibromyalgia. Central sensitization also links with changes in the brain's chemistry because CNS-acting drugs directly affect neurotransmitters like pregabalin, which antagonises excitatory glutamatergic inputs (Becker and Schweinhardt, 2012). Studies that measure neurotransmitters or their metabolites in plasma or cerebrospinal fluid have demonstrated lower noradrenaline and serotonin levels in chronic pain patients compared to healthy populations (Brummett and Clauw, 2011). Similarly, fatigue is associated with decreased monoaminergic neurotransmission (Korte and Straub, 2019), such as diminished dopamine in the cerebrospinal fluid of interferon- α treated

patients (Felger et al., 2013). Glutamate changes have also been assessed; studies have evaluated changes in components of glutamatergic neurotransmission and used drugs that antagonise this pathway to improve depression and fatigue. Ketamine, an antagonist of the glutamatergic N-methyl-D-aspartate (NMDA) receptor, has been shown to reduce fatigue in patients with bipolar depressive disorder (Saligan et al., 2016) and MS (Fitzgerald et al., 2021). Serum antibodies against a subunit of the NMDA receptor have been shown to correlate with fatigue in patients with systemic lupus erythematosus (Schwartz et al., 2019).

Central fatigue itself has been subdivided into physical, motivational, and cognitive subcomponents that are localised in different brain regions and pathways and summarised by Korte and Straub (2019). These pathways include six inflammatory-related disruptions that can interfere with brain communication and provoke neurodegeneration, but other non-immune mechanisms have also been explored (Matura et al., 2018). The link with inflammation stems from fatigue consistently correlating with low mood, anxiety, increased sleep, and hyperalgesia, collectively called sickness behaviours, which also emerge after immune challenges like interferon- α treatment, but patients continue to report them months after their peripheral inflammation had subsided (Russell et al., 2019). Central inflammation appears in individuals with sickness behaviours that may arise from peripheral cytokines passing the blood-brain barrier or signalling via the vagus nerve and affecting the brain (**Figure 1-2**), based on animals presented with peripheral immune challenges like lipopolysaccharide injections (Dantzer et al., 2008). The resulting changes have the potential to disrupt brain networks, resulting in various sickness behaviours (Kraynak et al., 2018). Given the inherent relationship between fatigue and these behaviours, biological insights from these more maturely studied symptoms could be transferred to accelerate our understanding of fatigue. Although these findings stem from animal and post-mortem work that may be insufficient to describe fatigue mechanisms in live patients, modern neuroimaging methods have already assessed sickness behaviours and have the power to safely assay neurobiology.

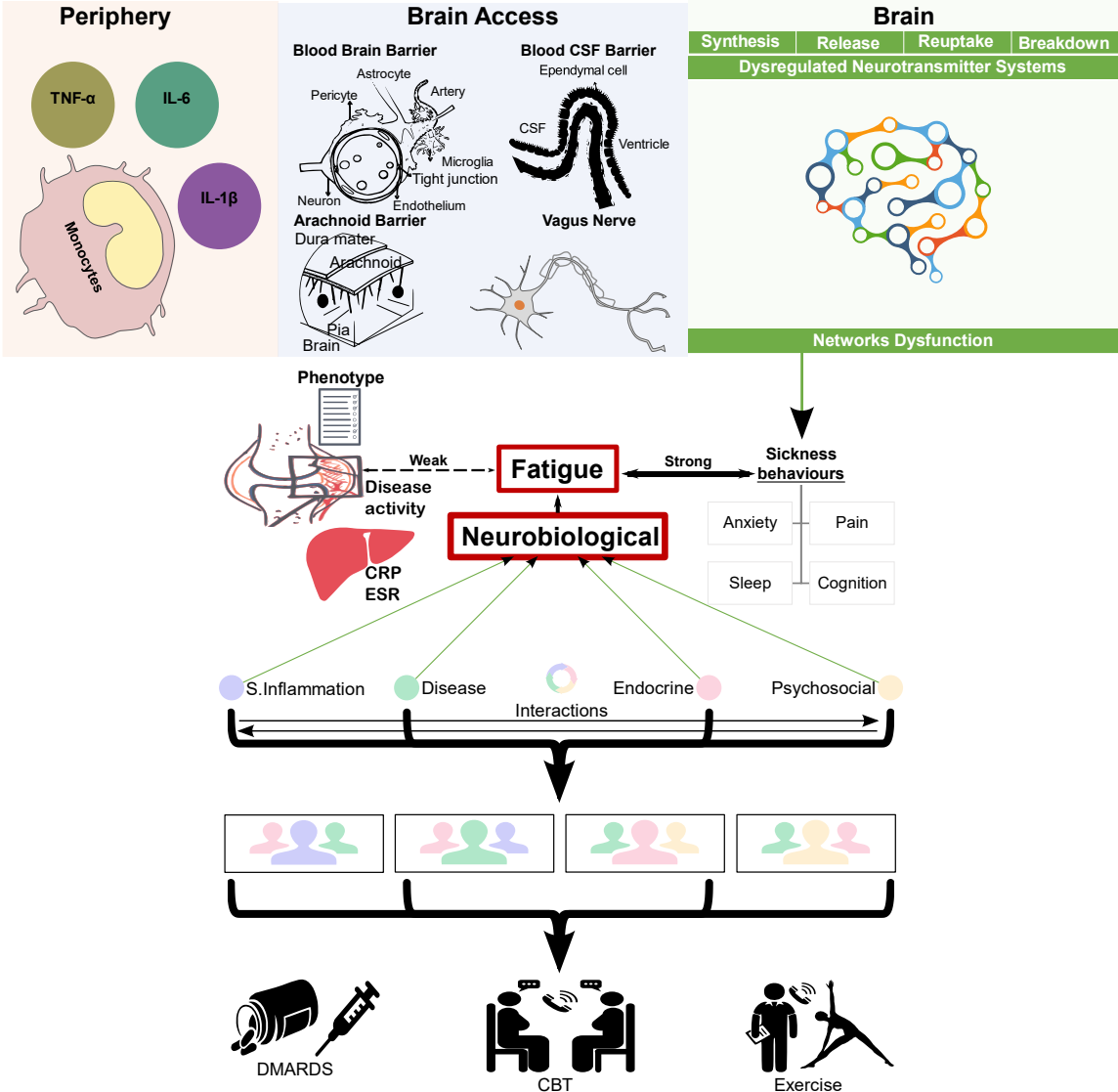


Figure 1-2: Theoretical model of fatigue as a multifaceted phenomenon and its management. Pre-clinical experiments suggest cytokines such as TNF- α and IL-6 pass the neurovascular barriers through receptor-mediated transport, leaky tight junctions, or fenestrated vessels at certain locations (e.g., circumventricular organs) while the vagus nerve samples inflammatory mediators and indirectly passes these signals to the brain. Monocytes may also gain access to the brain by expressing signals like interleukin-1 β and interacting with endothelial cells. The trafficking of these cells and signals likely interferes with the synthesis, release, reuptake, and breakdown of multiple neurotransmitters either directly or through mechanisms such as oxidative stress. Fatigue cannot be explained by disease severity, peripheral inflammation (e.g., C-reactive protein, erythrocyte sedimentation rate), or phenotypic traits on their own. Immune cells and signals affecting the brain and its neurochemistry directly could fill that gap. Multiple interacting factors can generate and maintain the neurobiological drive of fatigue in each patient with some effects being more dominant in certain groups. Consistently stratifying such groups could inform selection in clinical trials and the appropriate treatment, which can be delivered remotely. Illustrations were made using Inscape (2020). Abbreviations: CBT, cognitive-behavioural therapy; CRP, C-reactive protein; DMARDs, disease-modifying antirheumatic drugs; ESR, erythrocyte sedimentation rate; IL, interleukin; S. Inflammation, Systemic Inflammation; TNF- α , tumour-necrosis factor α .

1.4 Brain Anatomy & Function

The brain and spinal cord's primary function is to control and integrate functions. They make up the CNS and communicate with the rest of the body via the cranial and spinal nerves of the PNS. Glial cells support the neurons in the brain, which are connected to one another by their axons and dendrites. The grey matter in the brain consists of glial cells as well as the cell bodies, dendrites, and axonal endings of neurons, whereas the white matter is comprised of the long myelinated axon tracts of neurons. The brain receives nourishment, mechanical support, and immune protection from its vascular and ventricular systems, which circulate blood and cerebrospinal fluid (CSF). The human brain itself is anatomically divided into the cerebrum, cerebellum, and brain stem.

1.4.1 Cerebrum

The cerebrum is separated into a left and a right hemisphere by a fissure and connected by a white matter bundle called the corpus callosum. The outer structures of each hemisphere constitute the cerebral cortex, while the structures below are referred to as the subcortex. The cerebral cortex underlies perception and many conscious processes (Jawabri and Sharma, 2019), which are allocated to four lobes (**Figure 1-3**).

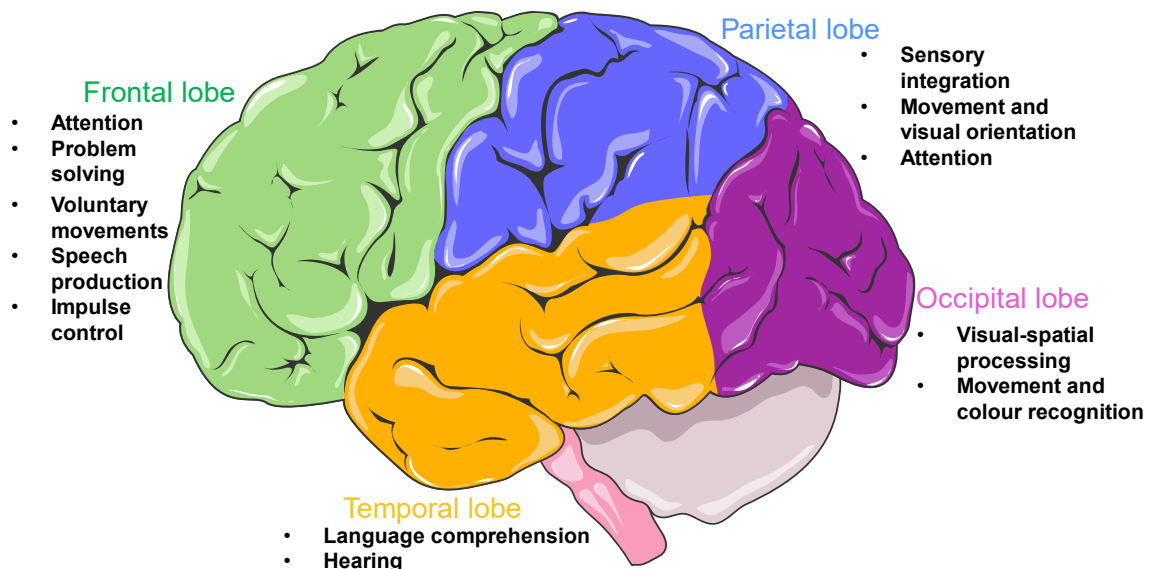


Figure 1-3: The four lobes of the cerebral cortex. Each hemisphere of the cerebral cortex divides into four lobes. The frontal lobe is separated from the parietal lobe by a groove called the central sulcus, and the lateral sulcus separates the frontal and parietal lobes from the temporal lobe. The parieto-occipital sulcus separates the parietal lobe from the occipital lobe. Each lobe has dedicated brain regions associated with specific functions. The figure was created using images from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 (unported license).

The subcortex (**Figure 1-4**) consists of the basal ganglia and limbic system brain regions, as well as the thalamus and hypothalamus. The cells of the cerebrum are organised into columns subdivided into layers of distinct numbers and types of cells (**Figure 1-5**). The most common form is six-layer columns (neocortex), but structures of the limbic system have a three-layer (allocortex) organisation (Strominger et al., 2012).

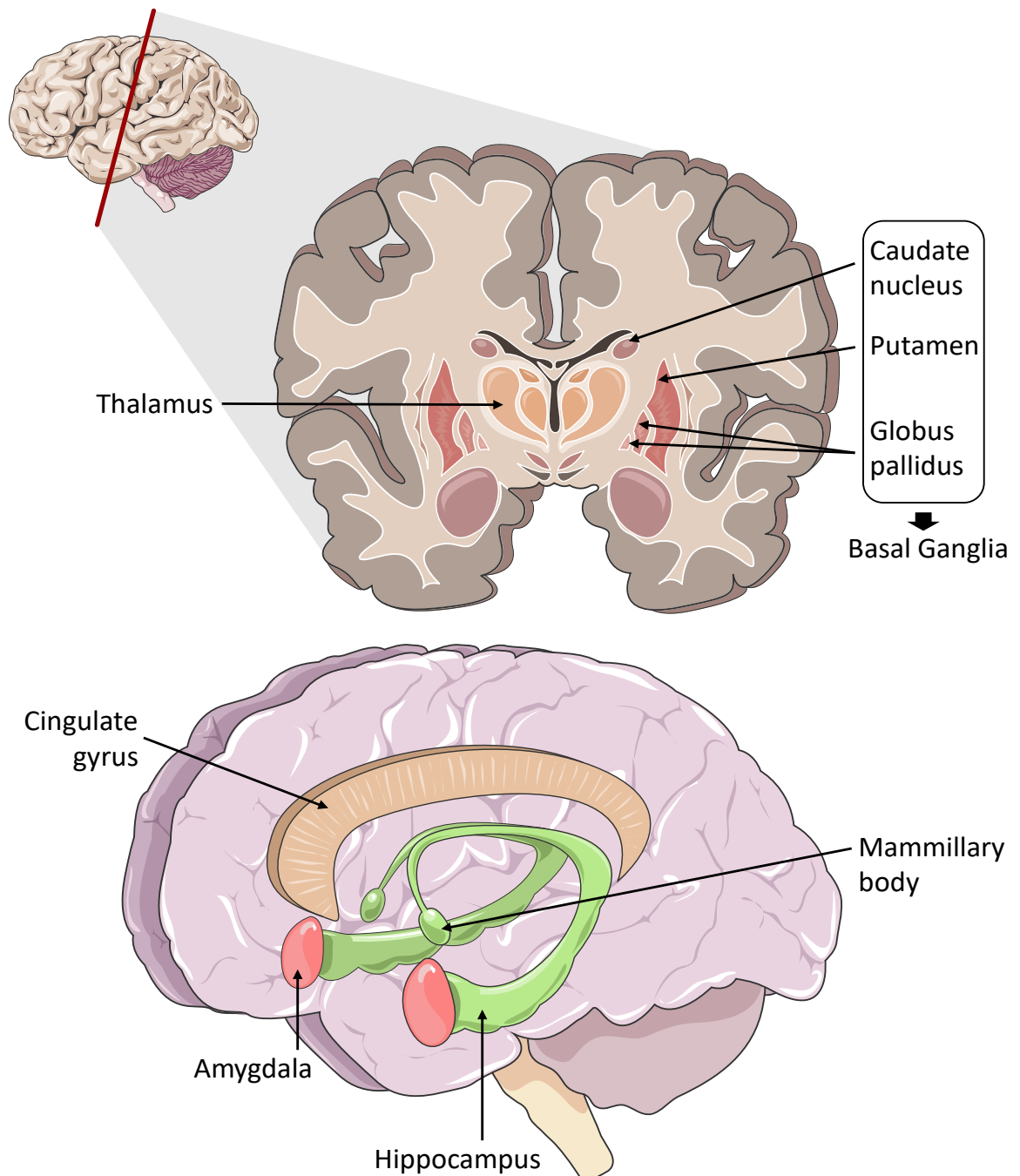


Figure 1-4: Structures of the subcortex. The structures of the Basal Ganglia and Thalamus are visualised on the top and those of the Limbic System on the bottom. The figure was created using images from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 (unported license).

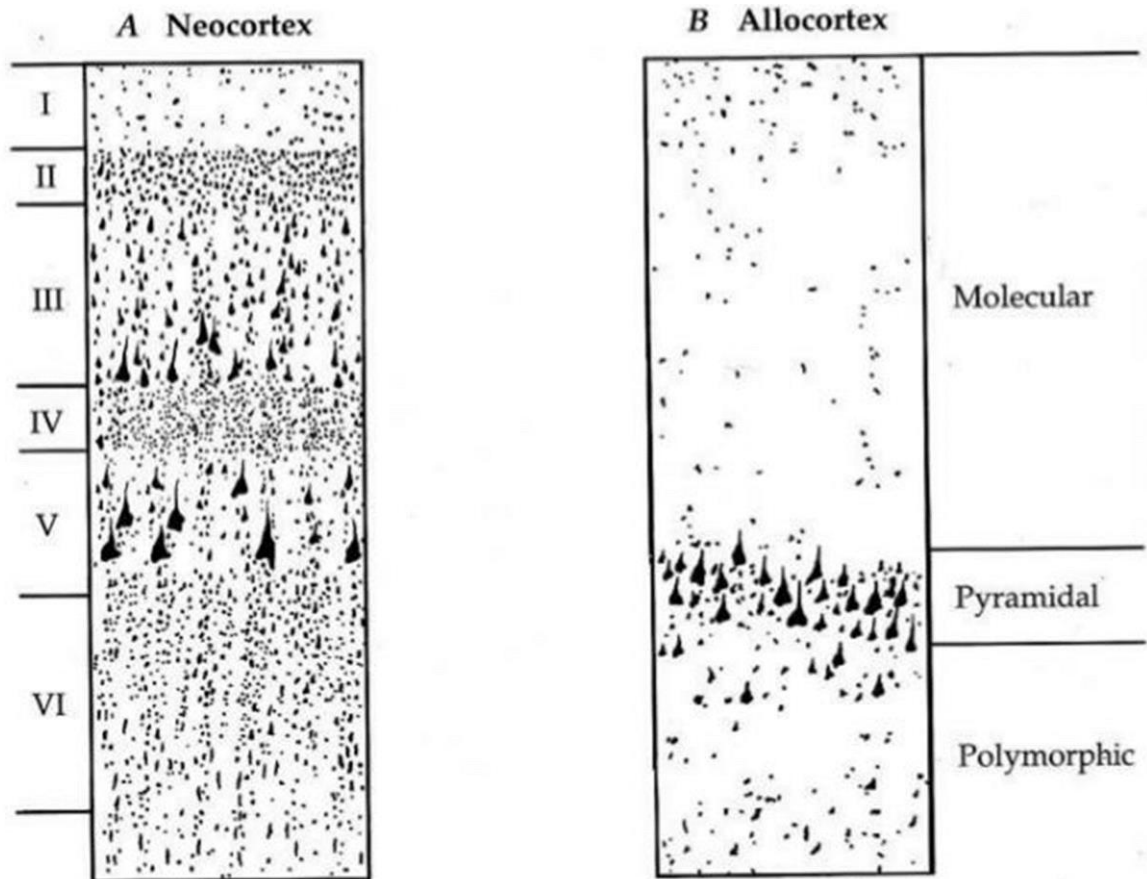


Figure 1-5: Layer organisation of neuronal columns. The cerebrum divides into a phylogenetically older allocortex consisting of three-layered neuronal columns and a more recent six-layered neocortex. Each cortical layer contains different neuronal shapes, sizes, and densities as well as different organisations of nerve fibers. Images are produced through histological Nissl staining (Martin, 2003).

1.4.2 Cerebellum

The cerebellum (**Figure 1-6**) monitors and fine-tunes motor behaviour. This translates to learning and conducting precise, coordinated movements rather than their actual initiation (Thach et al., 1992). Additionally, the cerebellum contributes to implementing language, attention, and the experiences of fear and pleasure (Strick et al., 2009).

1.4.3 Brain Stem

The brain stem (**Figure 1-6**) is a group of structures that lie deep within the brain. Anatomically, it divides into the medulla oblongata as a continuation of the spinal cord, the pons, and the midbrain. These regions control autonomic functions such as breathing, heart rate, blood pressure, and involuntary reflexes such as swallowing, but also interact with the motor and associated cortices to manage fine movements of the limbs and face (Angeles Fernandez-Gil et al., 2010).

The current section discussed general brain anatomy and function to lay the foundation for comprehending the role of the brain in fatigue. Before delving into this role, understanding magnetic resonance imaging is crucial, as it can non-invasively measure brain structure and function.

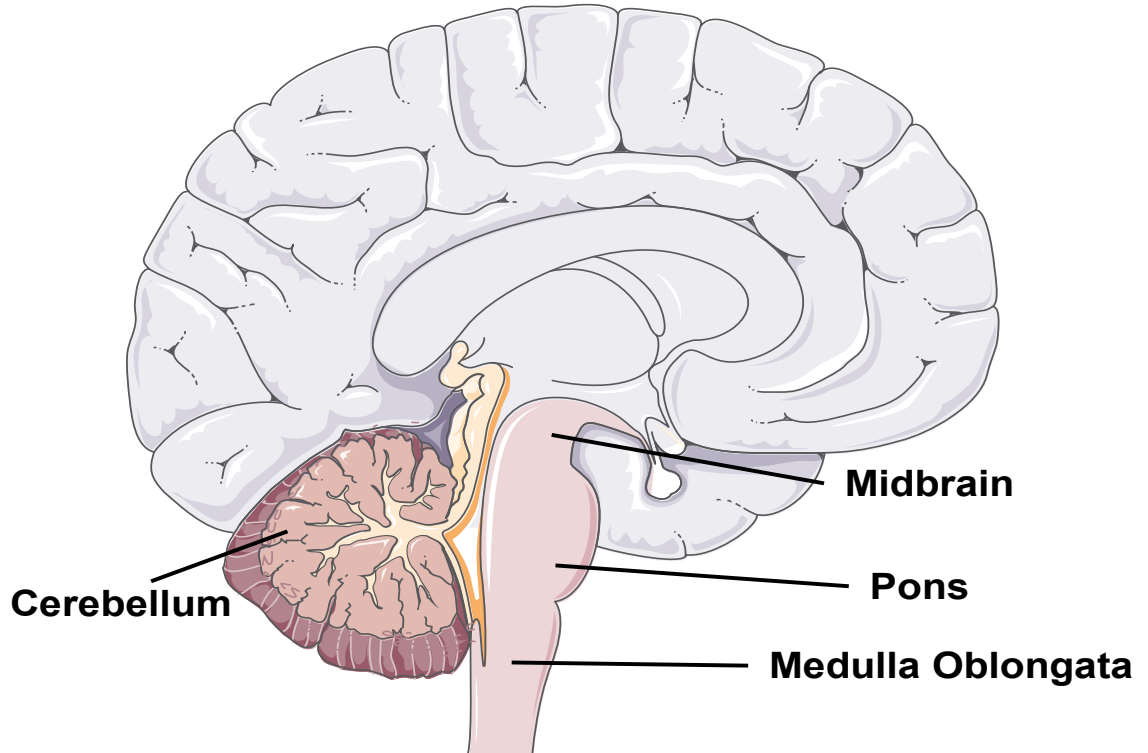


Figure 1-6: Cerebellum and Brainstem. The brainstem connects the rest of the brain to the spinal cord, with the medulla oblongata as its lower extremity, followed by the pons and midbrain. The cerebellum sits behind the brainstem. The figure was created using images from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 (unported license).

1.5 Basis of Magnetic Resonance Imaging

Before the advent of magnetic resonance imaging (MRI), the tools to investigate the brain suffered from being indirect (behavioural assessments), invasive (biopsies), or damaging (ionizing X-rays). MRI visualises the brain without these limitations by exploiting the magnetic properties of water in different tissues (Huettel et al., 2014). The MRI scanner first aligns the water molecules' protons to its constant magnetic field, then misaligns those protons with a weaker radio frequency pulse. After the pulse ends, the protons align again and subsequently emit low-energy photons. The photons are the signal that the scanner records as it passes through slices of the brain, creating a set of 2D images that can then be used to construct a 3D image of the brain. The intensity of each voxel (3D version of a pixel) of the image reflects the intensity of the signal. As thickness and hardness determine the time and amount of realignment changes, the produced 3D image can distinguish the location and shapes of different tissue types like grey matter, white matter, and CSF. To evaluate both the structure and function of these tissues, different modalities of MRI can be used.

1.5.1 Structural MRI

Structural MRI distinguishes between grey matter, white matter, and CSF as the protons in white matter realign more quickly than those in grey matter and consequently white matter appear brighter, while CSF appears darker because the protons in CSF take longer to reach alignment (**Figure 1-7**). The images produced then enable us to measure the volume and thickness of both grey matter and white matter. These properties can inform us of any disease-related or developmental processes by investigating any changes in time, differences between groups, or associations with other variables.

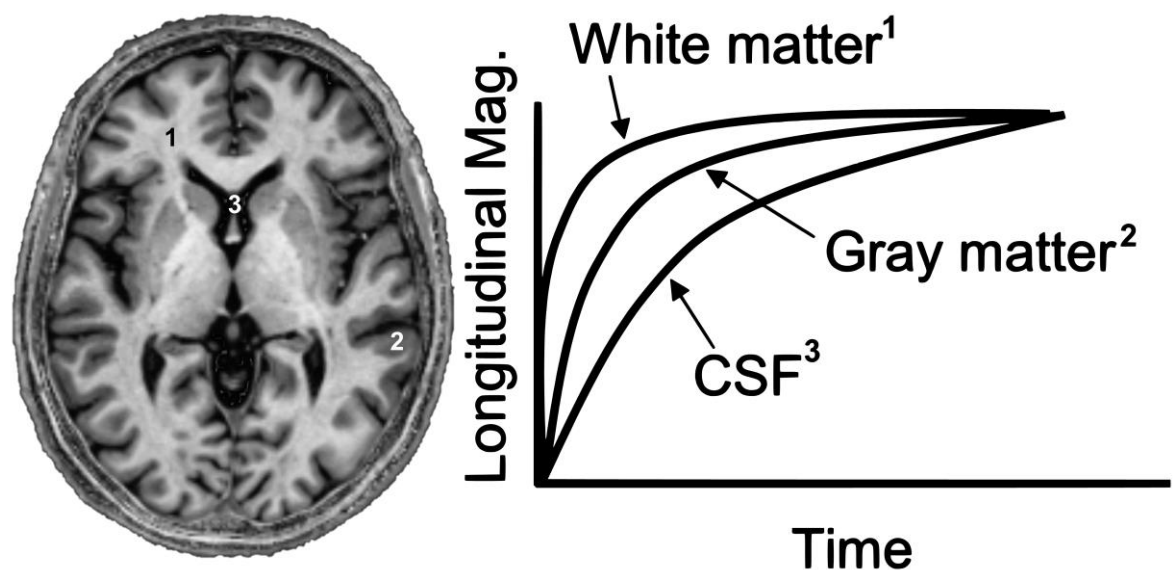


Figure 1-7: Structural MRI of tissue types. A standard (left) T1-weighted image depicts the difference in brightness of the three tissue types. This contrast is due to their different T1-relaxation (realignment) times, as shown in the plot on the right, where longitudinal magnetisation is equivalent to signal intensity. Adapted from Pooley et al. (2005).

1.5.2 Functional MRI

A structural MRI scan results in one high spatial resolution 3D image. In contrast, functional MRI (fMRI) creates a timeseries of 3D images that follow changes in brain function within a session. The most common signal that fMRI uses as a proxy for brain function is the blood oxygenation level-dependent (BOLD) contrast. A signal difference (contrast) arises from changes in its magnetic properties as oxygenated blood is deoxygenated to sustain neuronal activity (Ogawa et al., 1990). Because the BOLD signal correlates with neuronal activity, BOLD images can help locate which brain regions are active (increased oxygen consumption) or inactive (baseline oxygen consumption) during tasks.

The ability to acquire sequences related to neuronal activity led to many task-based designs to map function to defined brain structures. However, it was discovered that even in the absence of a task or stimulus (resting-state), synchronous activity between brain regions is present termed functional connectivity (Biswal et al., 1995). The brain regions that exhibit such similarity can be grouped within resting-state networks, and consequently, many such networks have been described and reproduced (Beckmann et al., 2005, Fox et al., 2005, Damoiseaux et al., 2006). These include the default mode, dorsal attention, and salience networks that are the basis for generating internal thoughts, executing external tasks, and filtering sensory cues, respectively. While the default mode network is more active at rest, all other networks are still active and can be extracted using independent component analysis (**Figure 1-8A**). The technique maps networks by decomposing the timeseries of BOLD signal changes into multiple statistically independent spatial patterns that are then compared to templates of networks from previous studies. The classical approach to estimating functional connectivity (FC) is to calculate Pearson correlation coefficients between two timeseries of BOLD signal changes measured over the duration of the sequence. The timeseries are the low-frequency changes in BOLD from a source (seed) and a target location. The seed is typically a brain region-of-interest (ROI), which is a set of voxels that encompass an anatomically distinct brain region. The target can be the time-course of a single voxel, or a time-course averaged across all the voxels of another ROI (**Figure 1-8C**). The correlations are then repeated for every pair-wise connection based on the selected ROIs and become part of a connectivity matrix (**Figure 1-8D**).

In terms of clinical applications, FC has been proposed as a biomarker for distinguishing different disease states. However, this application has been hindered by the high variability of FC (Wang et al., 2011), which may require longer scanning times (Anderson et al., 2011). Task blocks can also be used to estimate less variable FC as they better focus the attention of a participant and instil more homogenous cognitive states (Shah et al., 2016). Conversely, studies have used dynamic FC to improve FC biomarker detection, which uses fluctuations of FC in direction and strength within a timeseries (Chang and Glover, 2010). These fluctuations are detected when the fMRI timeseries are divided into consecutive windows of a certain length. Typically, they range from 30 to 60 seconds, with the rule of thumb (Leonardi and Van De Ville, 2015) that the minimum should be the inverse of the lowest frequency of interest. For example, 0.01 Hz would equate to a 100-second window while frequencies lower than 0.01 Hz correspond to scanner noise, such as thermal drift in shims, gradients, and radiofrequency components (Yan et al., 2009). One can then extract a series of connectivity matrices that span the fMRI sequence (**Figure 1-8D**). The additional

variability of dynamic FC allows for studies to better explain differences in individual behavioural and cognitive traits and classify neurological diseases, potentially because dynamic FC captures latent functional boundaries, especially in regions with high flexibility and adaptability (Peng et al., 2022).

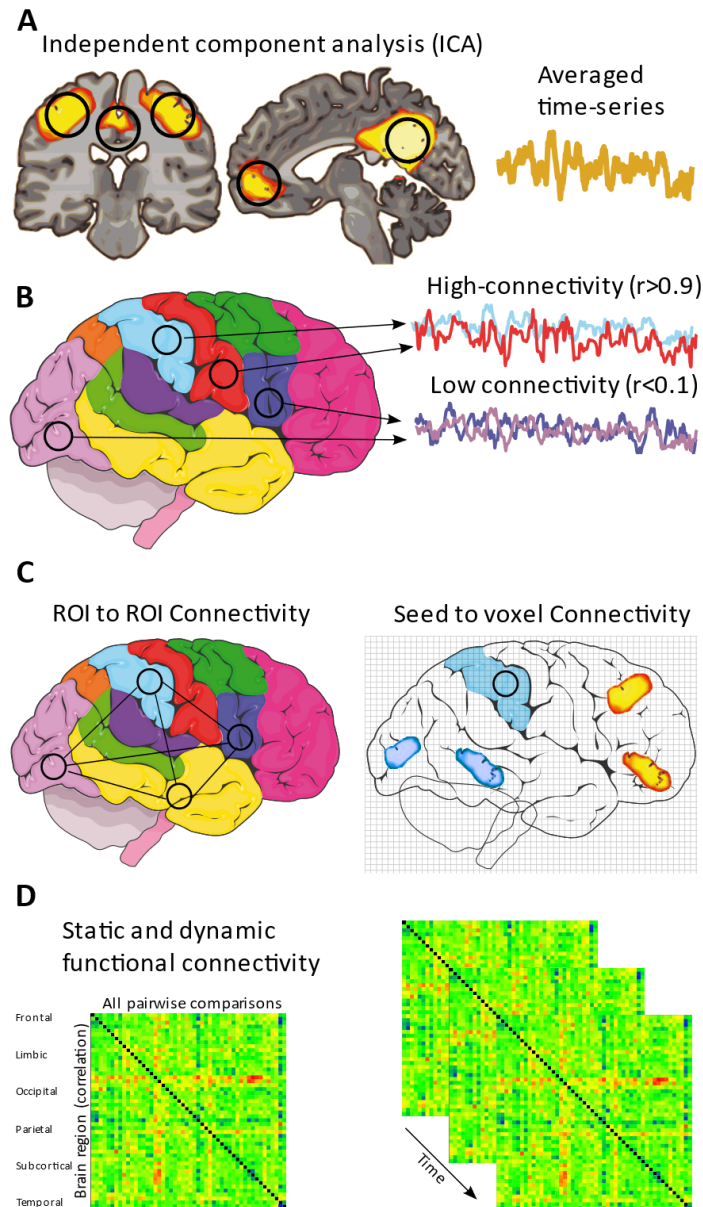


Figure 1-8: Estimation of functional connectivity. **Panel A** displays a component from independent component analysis that can be a network of brain regions. The timeseries of brain regions and networks can have high and low connectivity between them (**Panel B**). Brain networks can be seeds in seed to region-of-interest (ROI) or seed to voxel analyses (**Panel C**). **Panel D** comprises a static functional connectivity matrix that spans the correlation between the same brain regions on the x and y axes. The upper triangle after the diagonal (**black**) is then used to derive a single vector of values for each subject. The second image displays dynamic connectivity, where a series of connectivity matrices are extracted throughout time, typically under a set window of time. Adapted from Scheinost et al. (2017), while the figure was partly generated using Servier Medical Art, licensed under a Creative Commons Attribution 3.0 (unported license) and edited using Inkscape (2020).

1.5.3 Diffusion Tensor Imaging (DTI)

Water diffuses in different directions in the brain, and diffusion-weighted imaging (DWI) can be used to capture the magnitude of that diffusion (**Figure 1-9A**). When you measure diffusion in at least six non-collinear directions, you can compute a diffusion tensor in each voxel of the brain: a 3*3 matrix that can be used to calculate the preferred direction of the diffusion, often visualised by an ellipsoid with a direction (**Figure 1-9C**). Compared to grey and CSF, white matter fibre bundles constrain the diffusion of water along specific directions (anisotropy), which are identified as the tensor. The tensor can be colour-coded, yielding a cartography of the tracts' position, direction, and anisotropy (**Figure 1-9B**). Fractional anisotropy (FA) is a tensor quantification method calculated from the eigenvalue of the diffusion tensor that compares the relative strength of the principal component of the tensor to the remaining 2 directions, which quantifies the fraction of the tensor that is anisotropic. Therefore, FA captures the preferred direction of water movement along the direction of axons. Mean diffusivity (MD) measures the average water motion, independent of the directionality of tissues. MD is therefore a measure of average restriction to motion or cell density. Therefore, DWI studies can be interpreted such that reduced FA or increased MD indicate disrupted white matter integrity, with both metrics mapping well onto established pathologies like MS lesions. Both FA and MD are sensitive to more subtle differences in axonal count, myelination, and organisation, which can be affected by subclinical events like atherosclerosis and neuroinflammation but also by adaptive processes of synaptic reorganisation in response to long-term physiological or external challenges.

While DWI measures the magnitude or distance of diffusion inside the brain, diffusion tensor imaging (DTI) adds information on the direction of the diffusion by applying multiple diffusion gradients. Diffusion tractography (**Figure 1-9D**) creates an estimation of how white matter pathways run through the brain. Starting from a ROI, mathematical algorithms look for adjacent voxels whose diffusion direction is continuous with the previous one. This is repeated for all surrounding voxels till a path through the brain of highly similar voxels is created. These paths are called streamlines. In the same way as a connectivity matrix is created in functionality connectivity, the number of paths, or streamlines, connecting two ROIs can be determined. This can be converted to a probability score for each pair of ROIs that populate a matrix, similar to FC, called structural connectivity (SC). Clinically, DTI can also localise track-specific white matter lesions (**Figure 1-10**) and tumours, aid in neurosurgical planning, and assess white matter maturation during development. However,

there are limits to such mappings. As diffusion effects are averaged over a voxel, it complicates the biophysical interpretation of the diffusion tensor. FA is frequently interpreted as "white matter integrity," but many factors (e.g., cell death, changes in myelination, an increase in extracellular or intracellular water) can cause changes in FA. Overall, this difficulty in the interpretation of DTI is because the scale at which diffusion is measured with DTI is very different from the size scale of individual axons. Additionally, the tensor model is only able to represent one major fiber direction in a voxel; thus, DTI tractography can be confounded by regions of crossing fibers (regions in which fibres are interdigitating, brushing past each other, curving, bending, or diverging). There are also risks of partial volume effects, where two or more types of tissue are present in a voxel and can produce a tensor that represents neither tissue well. Finally, in standard streamline tractography, all decisions are made locally, so errors can accumulate. Those limitations result in the chance of false positive and false negative connections.

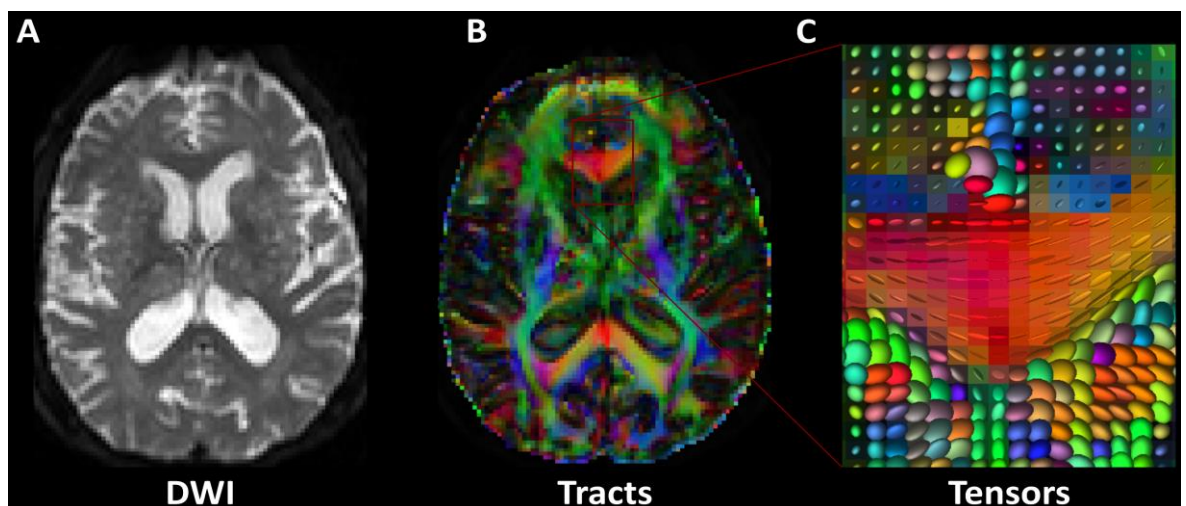


Figure 1-9: Diffusion Tensor Imaging (DTI). Based on diffusion-weighted imaging (DWI), an initial DWI image (A) can be used to create tracts (B) by estimating tensors (C). Colour-coding of the tensors can reveal the tracts' position, direction (red for right-left, blue for superior-inferior, green for anterior-posterior), and anisotropy (as indicated by the tract's brightness). Tractography methods use tensors to track fibres along their whole length and reveal the gross anatomy of cortical tracts. Images were created using FSLeyes in FSL 5.0.

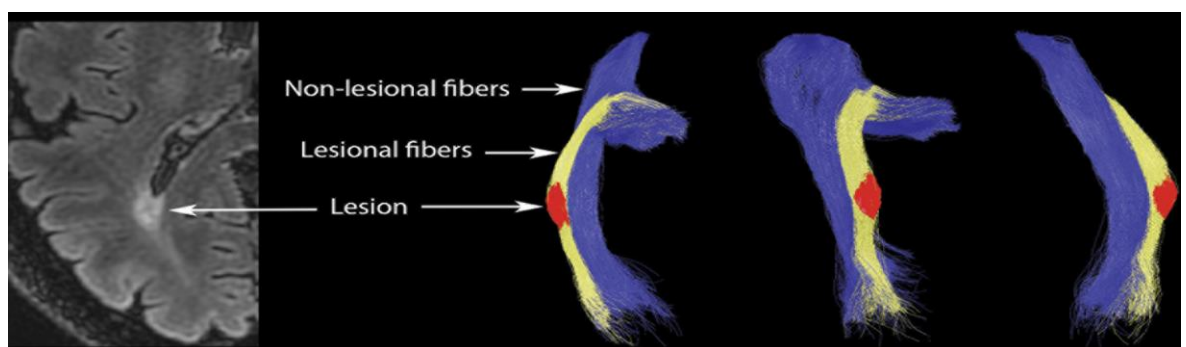


Figure 1-10: Clinical application of diffusion tensor imaging (DTI). In demyelinating diseases like multiple sclerosis, DTI can localise both the lesions and affected white matter tracks, as shown by Klistorner (2016).

1.5.4 Brain atlases

Beyond delineating grey matter from white matter and CSF, structural scans can be used to segment the brain into specific brain regions. Regions on the surface of the brain are termed cortical regions, while regions that are deep inside the brain in the form of grey matter nuclei are termed subcortical regions. These regions may be defined on the basis of brain atlases that are in a common standard space using neuroanatomical boundaries. Brain atlases can differ based on parcellation size, neuroanatomical coverage, and complexity of brain region shapes while the choosing which atlas to use is dependent on the research question (Revell et al., 2022). A commonly used atlas is the Desikan-Killiany anatomical atlas (Desikan et al., 2006), which defines 68 cortical regions based on probabilistic information from surface landmarks and 16 subcortical regions based on volumetric landmarks (**Figure 1-11**). These regions can also be related to specific brain networks (**Table 1-3**). The Desikan-Killiany atlas is a more coarse parcellation that addresses regions that can be theoretically targeted in potential interventions, and consequently used in these analyses. The other atlas used in this thesis was the more fine-grained default atlas used by the CONN toolbox, a software used to analyse functional connectivity data. The atlas defines regions that include the Harvard-Oxford Atlas and cerebellar areas from the AAL atlas, consisting of 132 regions, shown in **Figure 1-12** and listed in **Appendix A**.

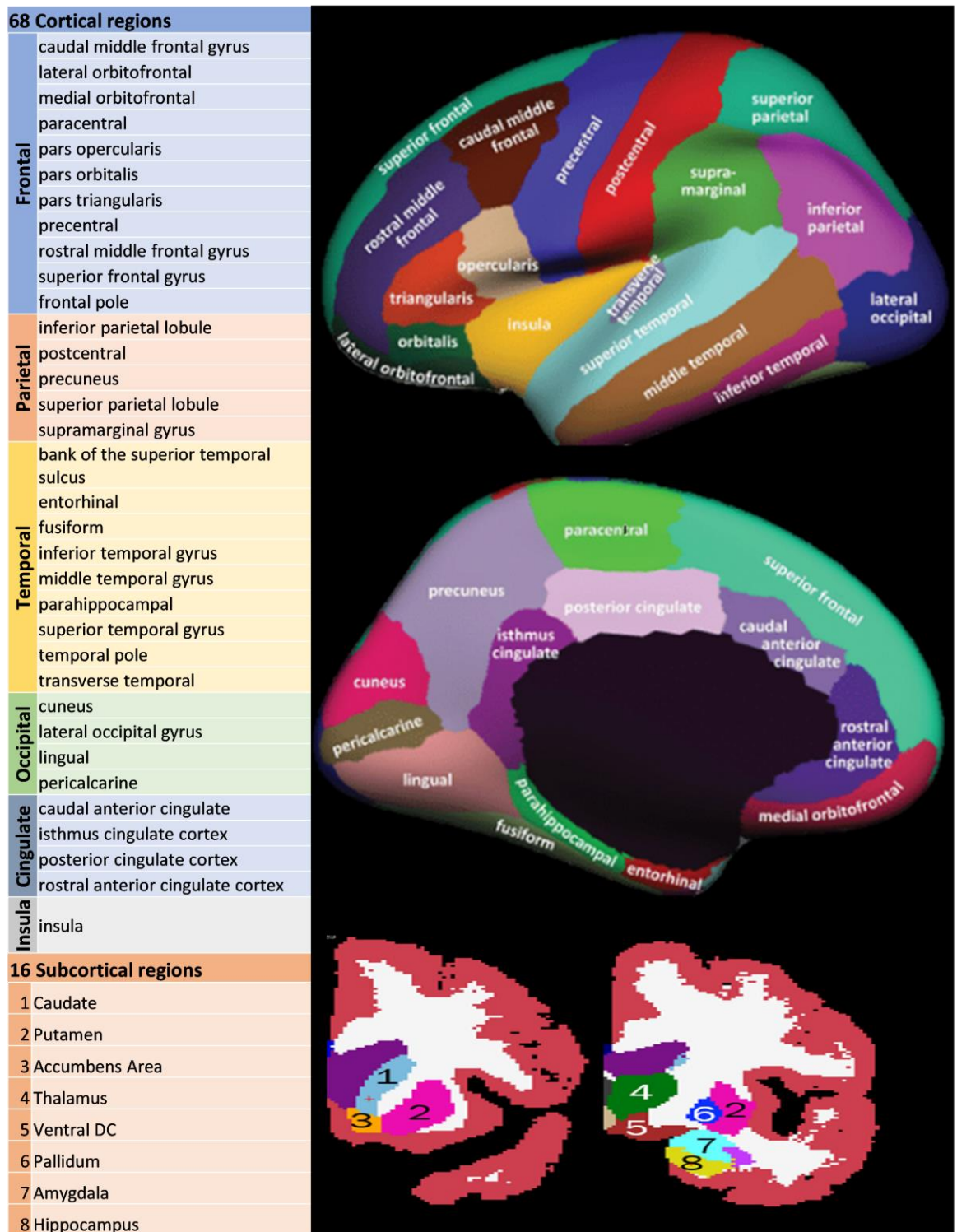


Figure 1-11: Desikan-Killiany atlas. The atlas used for these series of analyses segments the brain into 68 cortical and 16 subcortical regions in Free Surfer (<https://surfer.nmr.mgh.harvard.edu/>).

Brain Network	Regions
Default Mode Network (DMN)	Frontal Pole Superior Frontal Gyrus Posterior Cingulate Inferior Parietal Lobule (IPL), Angular Gyrus (AG) And Supramarginal Gyrus (SMG) Precuneus Middle Temporal Gyrus Parahippocampal Cortex Pars Orbitalis/ Triangularis/ Opercularis Bank of the Superior Temporal Sulcus
Dorsal Attention Network (DAN)	Superior Parietal Lobule Lateral Occipital Gyrus Caudal Middle Frontal Gyrus Lingual Gyrus
Saliience Network (SN)	Insula Rostral/Caudal Anterior Cingulate
Sensorimotor Network (SMN)	Postcentral Gyrus Precentral Gyrus Paracentral Lobule
Lateral Frontoparietal Network (FPN)	Rostral Middle Frontal Gyrus Inferior Temporal Gyrus Posterior Cingulate (Midcingulate) Precuneus Thalamus Caudate
Visual Network	Cuneus Pericalcarine Lateral Occipital Gyrus Lingual Gyrus

Table 1-3: Brain networks affiliated to regions in the Desikan-Killiany atlas. The affiliations are based from Uddin et al. (Uddin et al., 2019) using a common visual network for better understanding of results. Regions in grey are identified as core regions while some brain regions can be part of more than one network. Some regions in the atlas are not referenced in the paper and so are not listed here.

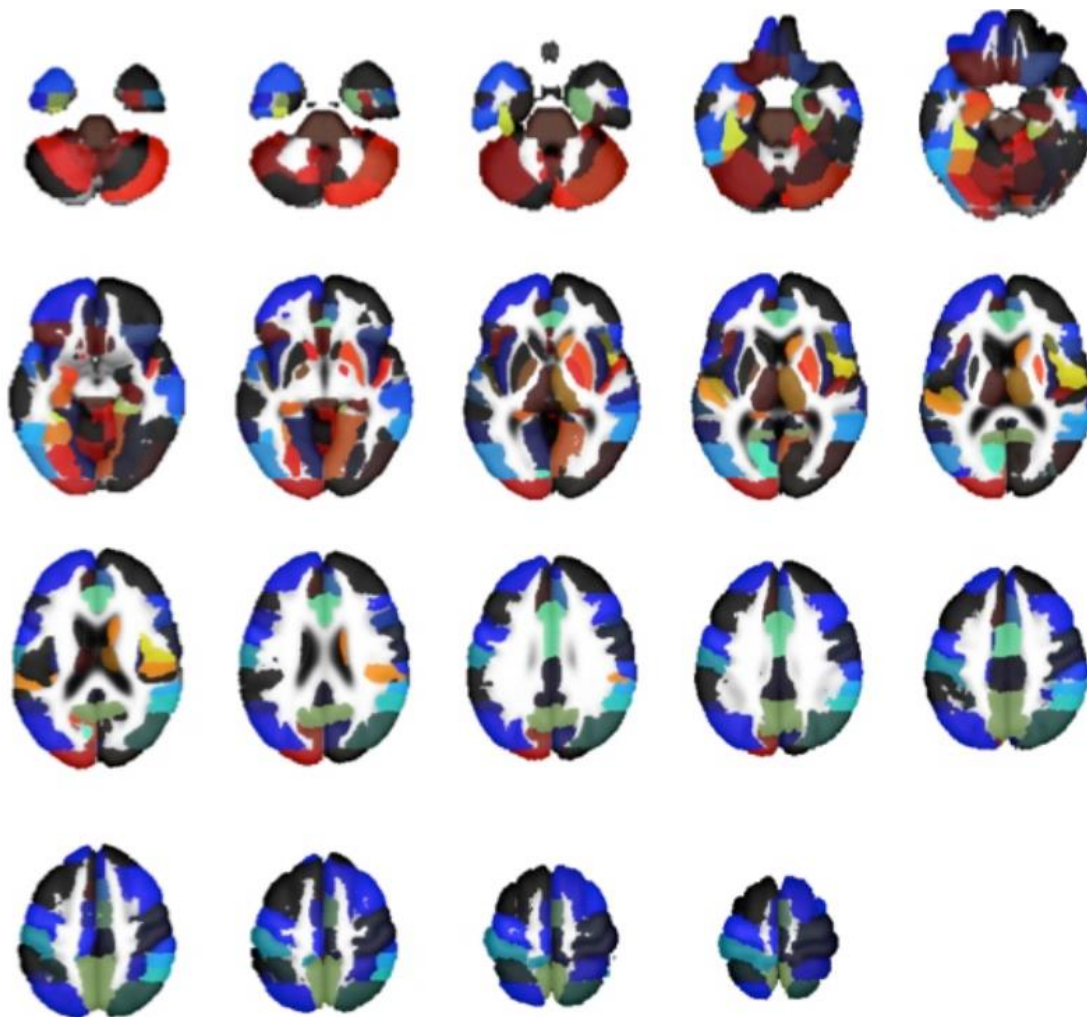


Figure 1-12: : CONN toolbox default anatomical atlas. Different views of the CONN default atlas incorporating the Harvard-Oxford Atlas, a probabilistic atlas covering 91 cortical and 15 subcortical structural areas as well as 26 cerebellar areas from the AAL atlas. Taken from Nieto-Castanon and Whitfield-Gabrieli (2021).

1.6 Neurobiological markers of fatigue in inflammatory arthritis

After describing the principles of MRI, using neuroimaging to study fatigue presents several challenges when compared to studying cognitive neuroscience in healthy populations or other clinical behaviours such as pain. Such challenges include: (1) heterogeneity of groups with chronic fatigue in the context of both fatigue expression and primary disease; (2) differences between physical and mental fatigue; (3) a limited number of interventions frequently not specific to fatigue; (4) a lack of fatigue reporting in studies of relevant cohorts that have other primary research objectives; and (5) tasks that induce fatigue during fMRI are less established and more difficult to interpret compared with ones for pain and cognition. However, clinical studies may have higher statistical power to identify neurobiological effects because disease likely affects the brain more than experimental manipulations in

healthy participants (Szucs and Ioannidis, 2020). Studies can exploit the heterogeneity in IA and compare groups with the same disease but with contrasting levels of fatigue. Finally, research on IA fatigue can also benefit from findings and tools from studies in conditions with more established neuroimaging literature, like MS, another inflammatory disorder where chronic fatigue is a patient priority.

A study recorded BOLD during a task that involved working memory and emotional processing in 23 patients with an IA condition called systemic lupus erythematosus (Barraclough et al., 2019). Patients' BOLD signal strength in the left caudate positively correlated with cognitive fatigue, as fMRI frequently associates IA-related fatigue with subcortical regions like the caudate. It, along with the putamen, are deep-lying grey matter nuclei that receive dopaminergic inputs and are jointly called the dorsal striatum. The striatum is part of the basal ganglia, which interacts with other cortical and subcortical regions such as the thalamus to execute precise movements but also affects working memory, decision-making, and emotional behaviour. FC can integrate such individual regions and provide more holistic insights, which is important for complex behaviours like fatigue. In 54 patients with RA, current fatigue was positively correlated with FC between the dorsal attention network (DAN) and the medial prefrontal cortex, extracted while performing a fatigue-inducing task (Basu et al., 2019). This result was reproduced in the same cohort six months later. The medial prefrontal cortex is part of the DMN, so these results demonstrate that patients with higher current fatigue show stronger synchrony between two networks that would typically work in opposition to one another as their activity is typically anticorrelated and associated with opposing tasks (Fox et al., 2005). If there is chronic excess activation and communication of the basal ganglia and regions of the default mode and dorsal attention networks, such functional demands may hypothetically deplete the functional reverse of the brain and surface as fatigue in patients, as suggested in MS (Capone et al., 2020).

Consistent with functional findings, structural changes in sub-cortical and frontal brain regions have been reported in IA cohorts. In the same 54 study participants with RA, greater grey matter volume in the putamen correlated with higher fatigue (Basu et al., 2019), with similar results in 20 patients with another IA condition called ankylosing spondylitis (Wu et al., 2014). Generally, fatigue is associated with larger volumes of grey matter in specific subcortical regions that underlie stronger internal communication within the same region. The opposite trend of negative associations with fatigue is observed for cortical regions, such as those of the dorsal and default mode networks, with both patterns observed in ankylosing

spondylitis. How these alterations arise and generate fatigue is difficult to disentangle, as the discussed structural and functional imaging markers are not specific to a single disruptive process. The study in people living with ankylosing spondylitis further documented differences in white matter integrity between patients and 20 matched controls in the form of decreased FA in tracts that connect these regions, such as the superior longitudinal fasciculus that links posterior parietal regions (DAN) with frontal areas. The use of DTI can thus add further detail by focusing on processes that affect the microarchitecture of white matter. Overall, both structural and functional changes of the brain are linked to fatigue in IA, while changes in white matter microstructure point towards potential disruptive processes like neuroinflammation and the necessity to assess the microenvironment of the brain.

Neuroimaging findings of fatigue still have gaps in knowledge, including more comprehensive neurobiological descriptors of fatigue and the potential for brain measures to predict treatment responses (Davies et al., 2021). For instance, it is unclear if different expressions of fatigue, like physical and cognitive fatigue, have distinct neuroimaging correlates. Such explorations could determine if phenotypical differences in fatigue involve different brain regions and identify a potential common hub for fatigue in the brain. It is also unclear if baseline brain characteristics can predict fatigue levels after pharmacological or psychological/physical treatments in IA patients. Additionally, it raises the question of whether predictors of fatigue after treatments are common or distinct between different diseases, such as RA and PsA. Identifying such predictors could potentially offer insights into how patient fatigue should be managed. Neuroimaging investigations could also determine if there are neurobiological subgroups of patients with potentially different fatigue mechanisms, aiming to guide mechanistic studies and subsequently enhance our understanding of fatigue.

Despite the findings using MRI modalities in IA (**Table 1-4**), inconsistencies were previously observed in brain imaging of chronic fatigue disorders, both when studies examined different conditions or the same disease (Goni et al., 2018). Studies used diverse methods and lacked both statistical power and stratification. Only 7 of the 26 studies reviewed used a longitudinal design, which enables replication of findings. Brain-behaviour associations are difficult to reproduce due to an imbalance of both small effect and sample sizes (Marek et al., 2022) compared with replicable mappings of brain functions like face perception, which have been shown to require only 15 participants (Yovel and Kanwisher, 2004). Studies in psychiatric and neurological conditions suffer less from low power as they

have double or larger mean effect sizes (Cohen's $d = 0.32$) than ones in the general population (Libedinsky et al., 2022). Although measures like structural MRI strongly correlate within repeated sessions of the same individuals ($r > 0.8$), reliable neuroimaging requires a greater number of participants in studies and improvements in the quality of the data acquired (Zuo et al., 2019). To identify generalizable brain-behaviour associations, steps can be taken to: (1) internally and/or externally validate findings; (2) use within-subject longitudinal designs; (3) use both rest and task states tailored to the behaviour of interest in functional imaging; (4) use multiple modalities in the same cohort; (5) experimentally manipulate behaviour through pharmacological and/or psychological interventions; and (6) experimentally manipulate brain signatures via neuromodulatory techniques relevant to the behaviour.

Research area	Cohort	Key findings	References
Functional activity and connectivity	Systemic lupus erythematosus (n = 23)	Cognitive fatigue positively correlated with activity (BOLD signal strength) in the left caudate, extracted during a working memory task	(Barracough et al., 2019)
	Rheumatoid arthritis (n = 54)	Current fatigue was positively correlated with functional connectivity between the dorsal attention network and the medial prefrontal cortex, extracted while performing a fatigue-inducing task The result was reproduced in the same cohort six months later	(Basu et al., 2019)
Grey/white matter characteristics	Rheumatoid arthritis (n = 54)	Fatigue in the last seven days positively correlated with grey matter volume in the putamen	(Basu et al., 2019)
	Ankylosing spondylitis (n = 20, healthy controls n = 20)	Fatigue impact (Fatigue Severity Scale) positively correlated with grey matter volume of the putamen and negatively with the caudate Fatigue impact was negatively associated with grey matter thickness of regions of the dorsal and default mode networks Lower white matter integrity (fractional anisotropy) in the superior longitudinal fasciculus compared to healthy controls that links posterior parietal regions (dorsal attention network) with frontal areas	(Wu et al., 2014)

Table 1-4: Summary table of MRI correlates of fatigue in inflammatory arthritis.

Some of the already mentioned studies in IA have implemented these strategies, such as using fatigue-inducing tasks, combining functional with DTI data, and analysing longitudinal data to replicate baseline findings. Longitudinal data also allows for the mapping of fatigue correlates via prediction methods that use baseline brain metrics to infer future changes in clinical variables. As fatigue correlates are spread throughout the brain in IA rather than focusing on a small number of regions, a recent study sought to predict fatigue using an agnostic multivariate approach in 54 patients with RA who had both structural MRI and DTI (Goni et al., 2022). Specifically, the approach considered 900 neuroimaging variables at baseline to classify patients who improved their fatigue levels from those who did not after six months. Both structural MRI (67.9%) and DTI (63.8%) performed better than chance, unlike when clinical variables were used to make the same predictions. Although applying prediction methods does not offer mechanistic insight into fatigue, they do provide clinicians with useful biomarker tools to stratify patients and aid in their decision-making. A significant advancement in creating such tools involves applying machine learning to develop high-performing models that make individual predictions, exemplified by Goni et al. (2022) in predicting fatigue outcomes in RA. Machine learning can uniquely use numerous predictors in a multivariate way, offering a potential method to improve fatigue management in IA by accurately predicting treatment response and general fatigue outcomes. This approach is the focus of the next section.

1.7 Machine learning & prognosis

Computer programs that perform complex tasks such as decision-making, visual perception, and translation between languages normally requiring human intelligence, are defined as artificial intelligence. If artificial intelligence models use real data to learn rules and modify decision thresholds, such models demonstrate machine learning (ML). To improve the quality of healthcare, artificial intelligence applications have been tasked with organising medical records, improving diagnostics, and discovering new drugs and disease subtypes. Prognos Health compiles prescriptions, medical claims, lab results, and other sources to create patient profiles (<https://prognoshealth.com/products/life-sciences/patient-journey>). Subtle Medical helps radiologists by improving the quality of images from the head, abdomen and breast by increasing the resolution and removing signal noise (<https://subtlemedical.com/subtlemr/>). Insitro builds predictive models from biological data to reveal crucial information like disease subtypes that is only available when looking at large datasets (<https://insitro.com/approach>). ML can also improve cost-effectiveness by

avoiding medical errors and unnecessary hospital visits, reducing drug discovery and treatment costs, and automating remote patient support and healthcare record writing. For example, Goldman Sacks estimated that if remote prevention and care tools are applied, it could save \$200 billion annually in the US, while early diagnosis can cut treatment costs by more than 50% in many cancer cases (Aboshiha et al., 2019), with such advantages applicable in settings with limited resources (Wahl et al., 2018). Before clinical research implements ML in healthcare, it needs to delineate the advantages, disadvantages, and interpretability of ML.

ML develops algorithms that learn from data to predict outcomes of interest and then apply the learnt patterns to new data. Patterns are made of features which can be raw images, questionnaire scores, or other data attributes that are related to the outcome. ML predicts traits of unseen individuals whereas classical statistics would infer traits of a population based on a sub-sample from that population. Classical statistics requires choosing an a priori model that incorporates our knowledge of the biological system, while ML chooses a predictive algorithm from all of those possible from the given data. Therefore, ML and traditional statistics differ in purpose as ML attempts to make the prediction as accurate as possible with the given data, while classical statistics infer relationships between variables. In practice, both goals of traditional statistics and ML can use the same model, but ML allows using data with more variables than individuals (high dimensionality) as well as the possibility of more complicated, non-linear, relationships.

High-dimensional data poses a multicollinearity problem where any variable could be expressed as a linear combination of other variables. Many possible combinations would produce similar results, making the solutions unstable, and even a small change in data would create large changes in how a model performs. ML methods overcome high dimensionality by forcing models to be simpler. ML methods simplify their final algorithms by either penalizing the importance of features used in the model based on their relevance to the outcome or by combining multiple models, developed from subsamples of features. To use non-linear relationships, ML can employ methods such as kernels, decision trees, or deep learning that employ neural networks (**Figure 1-13**). Kernels transform non-linear observations into a higher-dimensional space in which they become separable. Decision trees split data multiple times to learn complicated non-linear rules. Random forest algorithms form multiple trees from resamples and then average those outputs to improve the stability of the final model. The deep learning neural networks consist of an input layer that reads the selected features, several hidden layers that perform multiple functions at the

same time such as data transformation and feature selection, and an output layer of artificial neurons like simple linear regression models. The combinations of millions of such neurons learn highly non-linear dependencies between the data. Deep learning frequently performs best but requires much more data than available in neuroimaging of most clinical cohorts like chronic fatigue diseases and has therefore not been as successful (Koppe et al., 2021).

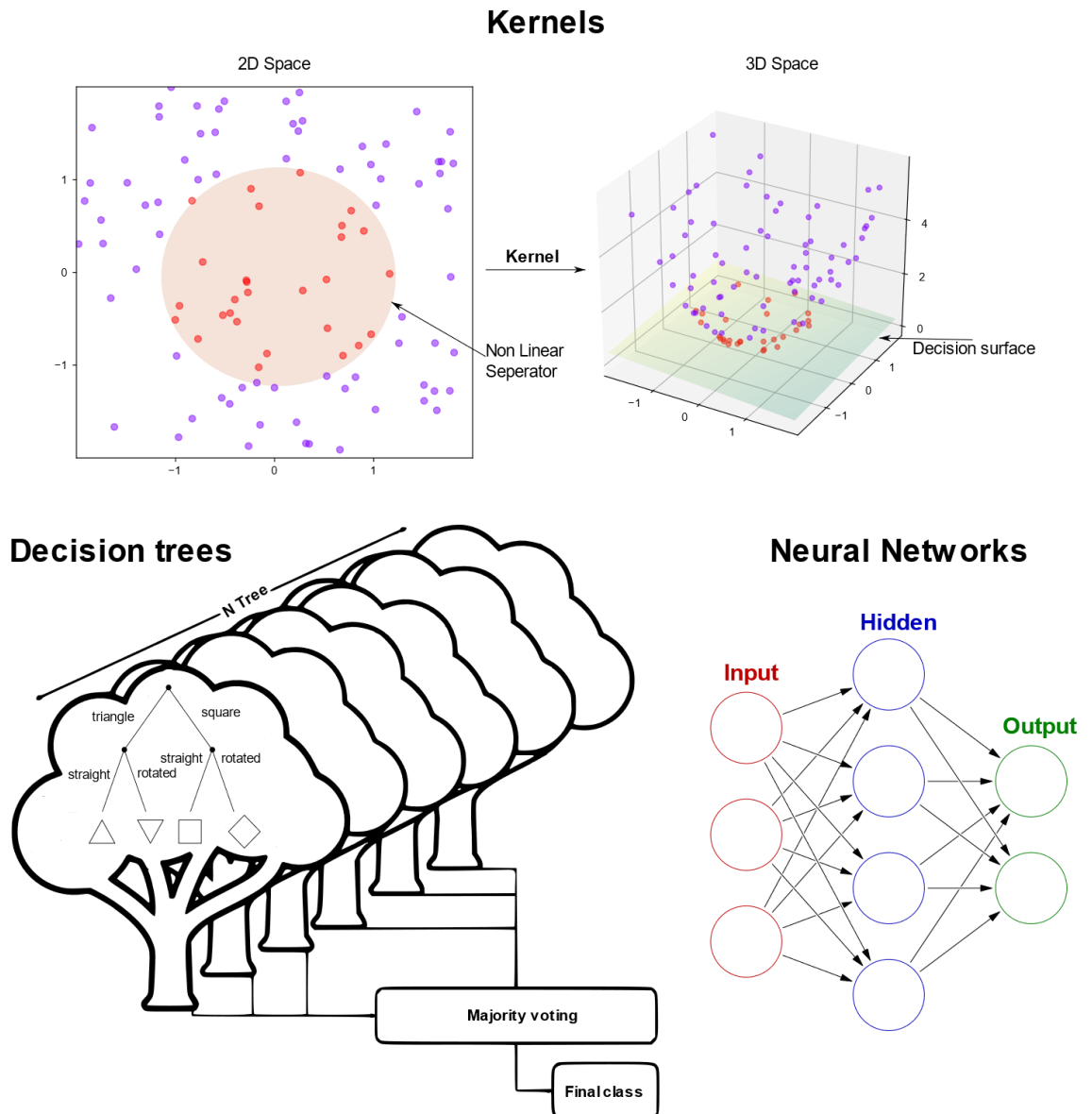


Figure 1-13: Machine learning methods to detect non-linear relationships. The top panel displays the potential kernels for identifying different classes that can be non-linearly separated by projecting the data from 2D to 3D space. The left bottom panel illustrates a simple decision tree to distinguish different classes of objects, based on non-linear rules. Agglomerating multiple trees to stabilise the final model is performed by the random forest methods. The right bottom panel shows the architecture of neural networks that detect non-linear trends by different data transformations, performed by hidden layers of artificial neurons. The images were created using Inkscape (2020).

ML predicts target variables that can be either a category (e.g., healthy or patient) or a continuous measurement (e.g., life expectancy). The final algorithms split into classifier models for categories and regression models for continuous variables, but both approaches make up supervised learning, where the correct answer is used to increase accuracy. Another branch of ML is unsupervised learning, which finds structure without defining a target variable. For example, clustering methods assign individuals to groups so that data points within groups are closer to each other than data points in other groups using some measure of distance. Analysts can then statistically describe the groups, compare them with other characteristics, or use them for predictions. Both supervised and unsupervised ML fit certain data and research questions better than traditional statistics. Unlike traditional statistics, ML methods are free from a priori assumptions such as the type of error distribution or additivity of parameters that are often not met in clinical data but are overlooked in the literature. ML can analyse various data types (imaging, demographic, laboratory) and integrate them into predictions without any prior data transformations. Finally, ML can identify novel targets from extensive lists of potential variables without prior knowledge of which variables to select. The advantages of ML over conventional statistics have translated into better predictions of outcomes like deterioration in the ward (Churpek et al., 2016), mortality in acute coronary syndrome (Hernesniemi et al., 2019), survival in patients with ovarian cancer (Paik et al., 2019), complications of bariatric surgery (Nudel et al., 2021), and risk of metabolic syndrome (Shimoda et al., 2018).

ML in imaging detects differences that are difficult to see with the naked eye, such as predicting an individual's gender or age using brain images (Schulz et al., 2020). In medical fields, such as oncology, ML helps combine biomarkers with mechanistic understanding. For instance, a study not only developed a highly accurate ML model to predict the stages of non-small-cell lung cancer, but also identified which genes regulate disease progression, differentiated genes that predicted survival rates, and suggested potential drug candidates that affect these genes (Jin et al., 2021). However, ML is more challenging to inform on biomarkers and mechanistic insight in illnesses based on symptoms only rather than quantifiable biological parameters. This is because the ML patterns discovered might not necessarily be connected to the pathophysiological pathways or to modifiable risk factors that facilitate the development of new therapies. These biomarkers, however, can be used to track treatment efficacy, choose the best treatment plan, direct care to those who need it the most, and develop specific treatments by identifying clinically meaningful subtypes.

Medical ML applications have some restrictions. Supervised ML lacks consensus on how to tailor performance measures to specific research questions. Regression models, for example, can be compared based on the average of the absolute difference between predicted and actual values, but this metric only provides an idea of the magnitude of errors, not their direction. In the absence of such standardisation, models are less comparable across studies. Supervised ML can overfit a model to the underlying dataset, causing it to fail to generalise to new patients. In addition, developing such models takes significantly more time and computational resources than developing traditional statistical models. Unsupervised ML methods lack statistical control because they always generate subgroups without first determining whether they exist at all. Unsupervised ML also lacks standardisation because the criteria used to determine the clustering method and the number of clusters measure various characteristics, such as the consistency or separation of the clusters. Finally, the purpose of clustering can vary depending on whether it is to identify biologically unique subgroups that can be linked to other metrics or clinically significant groups that differ on parameters like treatment response.

Traditional statistics can complement the advantages and disadvantages of ML (**Table 1-5**) in multiple ways. It can compute confidence intervals, serve as feature selection criteria, or assess whether the features chosen are connected to the construct of interest (i.e., identify false positives used in prediction but not informative). Statistical tools can also evaluate the stability and error rates of complex analysis pipelines. For example, Dinga et al. (2019) tested the stability of combining steps of feature selection and canonical correlation by removing one subject and rerunning the analysis multiple times. They found that doing so significantly changed the importance of the features and revealed that their model was unstable. Another potential problem with ML is that confounds can conceal models predicting something other than the outcome. Recently, Esteva et al. (2017) showed that deep learning can be used in dermatology to treat both common skin diseases and particular tumours. They did note, however, that if a biopsy image included a ruler, the algorithm was more likely to classify the lesion as malignant because a dermatologist would use a ruler when suspecting a malignancy. Due to unintended associations with the clinical outcome, other factors may influence ML models like scanner type or motion during scanning. Traditional statistics can aid ML models by regressing the influence of confounds before predicting the outcome. Finally, statistics can assess the cost-benefit of ML to cheaper alternatives such as using current symptom severity and compare ML-aided versus present decision-making protocols.

Because ML makes decisions at the individual level, medical applications are beginning to incorporate ML to improve healthcare quality and cost-effectiveness by learning patterns that go beyond what conventional statistics can deduce. The shortcomings of ML lie in the potential instability and confounding of its outputs. Traditional statistics can compensate for ML's limitations, which is especially important when investigating symptom-based traits like fatigue that do not associate with biological measures. Although it is still in its infancy, ML has the potential to address novel research questions about fatigue by discovering biomarkers and how they integrate into mechanistic pathways.

Term	Advantages	Disadvantages
Machine Learning: use real data to learn rules and modify decision thresholds	<ul style="list-style-type: none"> fit certain questions better than traditional statistics more variables than individuals (high dimensionality) non-linear relationships analyse various data types 	<ul style="list-style-type: none"> patterns not necessarily connected to mechanistic pathways less comparable results across studies compared to classical statistics
Supervised Learning: predict variables for unseen observations	<ul style="list-style-type: none"> more accurate predictions of outcomes no prior knowledge needed integrate different data types without prior data transformation 	<ul style="list-style-type: none"> can overfit and fail to generalise can conceal confounding effects requires more time and computational resources
Unsupervised Learning: finds structure without defining a target variable	<ul style="list-style-type: none"> identify informative subgroups does not require labels can be fast to implement 	<ul style="list-style-type: none"> no statistical control no standardisation (lacks gold standard criteria)

Table 1-5: Advantages and disadvantages of machine learning.

1.8 Aims, rationales, and hypotheses

Despite the revolution in immunological treatments for IA, the majority of patients still experience chronic fatigue. Neuroimaging has found correlates of fatigue, but the potential of brain imaging to tailor the management of fatigue in IA is currently unknown. To assess this potential, the following aims were set:

Aim 1: Identify subtypes of fatigue in patients with RA using brain connectivity to determine if there are neurobiological subgroups of patients with potentially different fatigue mechanisms.

Fatigue in IA varies in progression and expression even among patients with the same diagnosis. Patient comorbidities, like psychiatric illnesses, exacerbate this heterogeneity, making it difficult to explore fatigue mechanisms, match patients to current treatments, and develop new treatments. Clustering based on clinical data provides insights into trends in

fatigue expressions but may not yield subtypes that reflect biological differences. Clustering based on biological data reflects variability that is frequently unrelated to any pathology and instead reflects nuisance variance, such as groups with similar brain size, body type, or common ancestry. Instead of clustering clinical or biological data separately, a study in depression constructed variables that best represent the clinically relevant resting-state FC features and then clustered those variables to identify subtypes (Drysdale et al., 2017). The aim of this approach is to create clusters based on shared signatures of brain dysfunction. In pain, such signatures are used to assess pain as well as explain how different mechanisms generate overlapping clinical presentations (Tracey et al., 2019), with similar applications possible in fatigue. Each patient subgroup would have homogeneous connectivity patterns, the neurophysiological and genetic basis of which could then be investigated in humans and animals. Molecular techniques such as optogenetics can assert how the dysfunction of brain circuits that comprise the patterns of connectivity affects animal behaviour, whereas human studies can modulate connectivity via non-invasive neurostimulation and develop patient-specific therapeutics. In summary, identifying neurobiological subgroups of patients with different fatigue mechanisms can guide mechanistic studies and subsequently enhance our understanding of fatigue.

The use of ML has limits since cluster analysis would always produce subgroups regardless of the data structure—even if there were no clusters at all (Liu et al., 2008). Also, using subtypes to predict individual treatment responses may be inefficient since it assumes that all individuals in a given "biotype" are equally likely to respond. For instance, Drysdale et al. (2017) used biotype information to predict response to transcranial magnetic stimulation; however, if a patient shifted from biotype 1 to adjacent biotype 2, their projected response probability would plummet from 80% to 20%. Even if the biotypes are distinct, the probability of a response would vary smoothly with the clustering variables rather than discretely with respect to subtypes; thus, models with continuous variables, such as subject-specific connectivity values, would outperform models that only use subtype information. Treatment decisions would therefore depend on the response probabilities of patients rather than the average probability of their biotype. For this reason, neurobiological groups of patients that reflect different mechanisms of clinical symptoms like fatigue are primarily used to better deconstruct the heterogeneity of these symptoms.

To identify neurobiological subtypes of fatigue in RA using brain imaging, I plan to use fMRI connectivity and fatigue scores at baseline from RA patients who received usual care.

To validate the results, I will also apply these methods to data from the same patients after six months as well as to a different RA cohort.

Hypotheses: Neurobiological subtypes of fatigue exist in patients with RA, illustrated by a relationship between brain connectivity and fatigue, and distinct subgroups when RA patients are stratified based on brain connectivity. There will be a significant canonical correlation between fMRI connectivity and subdimensions of fatigue in RA patients and a cluster analysis on fatigue-related connectivity will reveal distinct subgroups of patients. I hypothesise that the individuals will come from multiple continuous distributions, which will confirm the presence of subgroups in terms of their brain-fatigue relationship. The subgroups will also be statistically different in terms of disease activity, systemic inflammation, pain, sleep, and depression since most of these variables link with central mechanisms of sickness behaviours. There will be a significant canonical correlation between brain connectivity and fatigue and distinct subgroups after cluster analysis in a different RA cohort that will externally validate the previous findings.

Aim 2: Identify associations between brain imaging metrics and baseline scores of subdimensions of fatigue in RA patients.

This aim approaches the challenge of deconstructing the heterogeneity of fatigue, akin to the previous aim, but from a different angle. Instead of exploring potential neurobiological subgroups and evaluating their phenotypic variations, this approach will begin with different fatigue phenotypes that can potentially be neurobiologically distinct. Neuroimaging provides a window into the neurobiological foundations of central fatigue, which encompasses subtypes like 'motivational', 'physical', and 'cognitive' fatigue (Karshikoff et al., 2017). This begs the issue of whether different subtypes of central fatigue have different neurobiological associations. At present, there exists no formal categorisation for fatigue subtypes, and questionnaires gauging fatigue facets do not distinctly segregate patients into these groups (Chorus et al., 2003). It is plausible that subtypes of fatigue may reflect various facets of the same symptom rather than being the basis of clinically separate subgroups. Each facet of fatigue might be linked to the perturbation of specific neural circuits, which may include the mesolimbic (motivational), nigrostriatal (physical), and mesocortical (cognitive) dopamine pathways (Korte and Straub, 2019). While these pathways are distinct, a common stressor can lead the brain into a fatigued state, impacting multiple neural circuits. Within IA, inflammatory cytokines may disrupt neurotransmitter release in various brain regions (Soliven and Albert, 1992), or sustained psychosocial stress could alter noradrenaline

metabolism (Matsumoto et al., 2021). Nevertheless, no studies have yet compared brain imaging correlates of fatigue subdimensions in IA. Here, I will use neuroimaging data to identify descriptors of fatigue subdimensions and then discuss their differences. In practical terms, while aim 1 would delineate potential biological subgroups informing the selection strategy in mechanistic studies, this approach would provide neurobiological descriptors of fatigue phenotypes for subsequent exploration of their mechanisms.

To identify associations between brain imaging metrics and subdimensions of fatigue, I plan to use baseline scores of a multidimensional fatigue questionnaire. Furthermore, I will use not only FC, but also SC and morphometric metrics of the brain as potential correlates.

Hypotheses: There are significant predictors of subcomponents of fatigue (physical, living, cognitive, emotional) that mostly differ from one another as they are based on different brain circuits but also share some common brain regions due to pathological mechanisms affecting multiple brain circuits.

Aim 3: Identify baseline brain imaging predictors of fatigue in RA patients after six months of receiving fatigue-specific interventions, including a personalised exercise programme and a cognitive-behavioural approach.

Building predictive models often begins by selecting candidate factors that have previously been linked with the outcome of interest. For example, lower cortisol awakening response and vitamin D deficiency have been associated with the chronicity of depression (Vreeburg et al., 2013, Milaneschi et al., 2014), while BDNF, inflammatory markers, and metabolic syndrome were significantly different in depressed patients compared to healthy controls (Bus et al., 2014, Vogelzangs et al., 2014). However, predictive models should not solely rely on such associations since they do not necessarily underlie the ability of variables to predict the outcome. A study demonstrated that CRP, IL-6, cortisone, metabolic measures, BDNF, and vitamin D were unable to predict the chronicity of depression, despite using the same sample of patients in which the previous findings were made (Dinga et al., 2018). Therefore, associations or differences with healthy populations at the group level do not imply the ability to make predictions for new cases in individual subjects. This may be because the effect sizes are too small to possess sufficient prognostic ability for long-term outcomes in individual patients, and biological markers implicated in the aetiology of the disorder are not necessarily good prognostic markers.

To explore baseline brain imaging predictors of fatigue in RA, data will be used from RA patients who underwent brain MRI before randomisation into usual care (UC) or fatigue-specific interventions (personalised exercise programme (PEP) or a cognitive-behavioural approach (CBA)). The T1, DTI, and fMRI modalities will provide morphometric metrics as well as structural and functional connectivity. These metrics will predict fatigue scores after six months in the fatigue-specific intervention groups, employing univariate general linear models (GLMs). The models will assess as predictors of fatigue both previously identified fatigue correlates and agnostically each neuroimaging variable. Also, I will investigate whether different subcomponents of fatigue (physical, living, cognitive, and emotional) yield different neuroimaging predictors.

Hypotheses: There will be statistically significant predictors of fatigue scores using the agnostic approach. However, previous correlates will not predict fatigue as these are markers of current fatigue in patients who did not receive fatigue-specific treatments.

Aim 4: identify baseline brain imaging predictors of fatigue after starting a new pharmacological treatment in RA and PsA cohorts.

Transdiagnostic approaches aim to address clinically challenging questions by tackling the heterogeneity of symptoms within and across different disorders. In neuroimaging, this approach can involve identifying brain features that associate or predict symptoms by applying the same methodology to different disorders. Focusing on continuous behavioural measures rather than diagnostic categories, transdiagnostic results may capture individuals along behavioural and biological spectrums, ranging from subclinical to severe symptoms. Patients, instead of adhering to a single distinct pattern, exhibit diverse patterns of symptoms and comorbidities, posing challenges in treating spectrum behaviours like fatigue due to the absence of predefined exemplar patterns. By leveraging brain imaging in diverse samples, the goal is to identify a 'transdiagnostic' network that generalises across a range of clinical categories. For example, autism spectrum disorder and attention deficit hyperactivity disorder share symptoms like attention deficits, the neurofunctional basis of which is undefined. A study used whole-brain FC in those clinical conditions to determine disorder-specific and shared (cross-diagnostic) neurofunctional pathology (Lake et al., 2019). They identified patterns that predict disease-specific clinical scores but also patterns that crossover between these scores. Although different disease scores implicated different brain circuitry (sharing only 2% of connections), components predictive of attention deficits translated across disorders. Other studies have combined data not only from different psychiatric

disease cohorts but also from healthy participants (Geller et al., 2021, Parkes et al., 2019). These studies focused on impulsivity and anhedonia, the reduced ability to experience pleasure, and found that these behaviours converged on similar brain connectivity traits across participants, which were also important for clinical severity in the subjects with diagnosed disease. In turn, focusing on dimensional phenotypes rather than diagnostic comparisons can lead to valuable findings in IA fatigue as well.

Previous research has often avoided transdiagnostic approaches due to uncertainties surrounding the specificity of an effect: how much patients drive outcomes relative to healthy control participants. Furthermore, mixing patients with different conditions can obscure whether transdiagnostic markers reflect the desired phenotype or are driven by unmeasured confounds. Nevertheless, the transdiagnostic approach gives unique insights into symptoms that can range from average to subclinical to clinical, exhibit dynamic fluctuations as disease states change, and rely on the same brain circuits. Furthermore, transdiagnostic biomarkers have inspired initiatives to improve treatments such as deep brain stimulation by targeting neural circuits linked to functional domains that cut across diagnoses (Widge et al., 2017). Although more complex, such approaches have shown early evidence of successful stimulation in the context of cognitive control (Basu et al., 2021). The ultimate objective of transdiagnostic biomarkers is therefore to provide a translational avenue by demonstrating that the underlying brain features can be modulated by current or novel therapeutics.

To identify baseline brain imaging predictors of fatigue scores after starting a new pharmacological treatment in PsA and RA, I will reapply the analysis from Aim 3 to a different RA cohort and PsA cohort that had brain MRI before starting a new DMARD. The brain metrics will include both morphometric features as well as resting-state FC. Predictors of fatigue will be compared between the RA and PsA cohorts who started new pharmacological treatments.

Hypotheses: There will be statistically-significant predictors of fatigue scores after starting a new DMARD in both the RA and PsA cohorts. Some brain metrics will be shared between results in RA and PsA due to a common transdiagnostic network of fatigue, but others will be distinct due to differences between diseases in which brain metrics have the most influence on fatigue outcomes.

Aim 5: To predict individual fatigue scores in RA patients six months and one year after receiving usual care, a personalised exercise programme, or a cognitive-behavioural approach by constructing machine learning regression models that use baseline clinical and multimodal neuroimaging data.

ML enables the use of many different types of factors to predict individual outcomes. For fatigue, prognostic factors could include genetic, immune, endocrine, neuroimaging, socio-demographic, and symptom-based information (**Figure 1-2**). Prognostic models can determine not only whether an individual should receive treatment but also what type of treatment is most likely to improve their well-being. Examples of ML models that use neuroimaging biomarkers to provide differential predictions include selecting CBT or an antidepressant (McGrath et al., 2013) or the optimal combination of two antidepressants (Williams et al., 2015). Quantitative rather than qualitative measures allow for cut-off points to assess levels of severity, which can then be fine-tuned with data from future studies, as has previously been shown in the early prediction of sepsis (Amrollahi et al., 2022). Severity cut-off points reflect the dimensionality of variables and move away from limiting categorical predictions such as whether a person has chronic fatigue or whether fatigue will improve. Instead, a dimensional approach can capture the full spectrum of alterations and provide a framework to accommodate comorbidity and sub-threshold conditions or changes. Overall, correlates of outcomes might not necessarily have strong prognostic value but can contribute to ML models. If properly applied, ML can incorporate multiple types of predictors, determine the optimal treatment choice for each individual, and estimate the magnitude of change, which is more informative for multidimensional outcomes like fatigue.

To predict individual fatigue scores in RA patients six months and one year after receiving UC, PEP, or CBA, I plan to construct multivariate models. Single modality models will use brain morphometric, structural or functional connectivity data separately as well as clinical data. A multimodal model will also integrate all data modalities, including treatment allocation. I will then externally validate the best-performing model in the RA cohort that had only UC to test the reproducibility of the model.

Hypotheses: I hypothesise that the multivariate model will perform better than chance through permutation testing; the neuroimaging modalities will perform differently between each other shown by Wilcoxon signed-rank tests; the treatment allocation will have high discriminative power on performance based on feature importance; and the performance will be similar in the different RA cohorts.

1.9 Chapter contents

The above introduction described IA, the effect fatigue has on patients with such diseases, the role of the brain in fatigue, and the pros and cons of using ML in medicine. In the rest of the thesis, I will examine the aims in the previous section to widen the scope of tools and knowledge on fatigue in IA as a clinically pertinent topic.

Chapter 2 describes the datasets from all patient cohorts and outlines the ML and MRI preprocessing methods used in subsequent analyses. In Chapter 3 (aim 1), I identify neurobiological subtypes of fatigue in RA in patients who received UC and internally validate the subtypes using data collected in the same patients after six months. Brain associations of subdimensions of fatigue are explored in Chapter 4 (aim 2) by using baseline brain imaging and clinical data from a larger RA cohort. Chapter 5 (aim 3) focuses on baseline brain imaging predictors of fatigue in RA following fatigue-specific interventions using the cohort data from the previous chapter. Specifically, this second RA cohort underwent brain MRI before being randomised to PEP and CBA groups. Chapter 6 (aim 4) explores a third RA cohort as well as a PsA cohort that underwent brain MRI before they started a new DMARD, aiming to identify brain imaging predictors of fatigue after the new pharmacological treatment and potentially find a transdiagnostic fatigue network. In Chapter 7 (aim 5) I employ ML techniques to construct regression models in the UC, PEP, and CBA RA patient groups, aiming to predict individual continuous fatigue outcomes using brain imaging and clinical data. Chapter 8 integrates the findings from previous chapters, discussing their contributions to the current understanding of fatigue. Additionally, I describe potential novel treatments based on neuroimaging biomarkers and targeted modulation of altered brain networks to improve fatigue. Finally, I reflect on the future work that needs to be done before neuroimaging can be implemented in the clinical management of fatigue.

Chapter 2 General Methods

2.1 Datasets

2.1.1 RA study dataset 1

The dataset was based on an observational cohort study, which approached RA patients who attended a UK regional rheumatology service (Schrepf et al., 2018). Inclusion criteria required patients to meet the 2010 American College of Rheumatology/ European League Against Rheumatism classification and experience clinically significant fatigue (>3 on the Chalder scale) for over three months. Patients with MRI contradictions such as metal implants were not included. The exclusion criteria also encompassed alternative medical explanations for fatigue (symptomatic cardiorespiratory disease, a history of cancer in the previous five years, unstable thyroid disease, beta-blocker prescription) and left-handedness. The patients in the study had multimodal imaging (T1-weighted, fMRI BOLD, DTI) and a clinical battery of tests. The full procedure was repeated at a second session, six months later with recruitment shown in **Figure 2-1**. The clinical phenotyping included ESR (Wolfe and Pincus, 2001) and CRP (Pepys, 1981) measures of inflammation, fatigue (Chalder Fatigue Scale, Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire), pain severity (measured from a rating scale of 1 to 10), and fibromyalgia (Wolfe et al., 2011). Venous blood was drawn by a trained phlebotomist during routine clinical hours (approximately 9 a.m.–5 p.m.), immediately processed, and analysed for ESR (Westergren method). CRP was also analysed promptly from serum using an ADVIA® XPT immunoassay System (Siemens). The study adhered to ethical guidelines with approval from the North of Scotland Research Ethics Committee. All participants provided informed written consent following the Declaration of Helsinki, which is applicable to all the datasets used in this thesis.

All neuroimaging was conducted in Aberdeen (Scotland, UK) using a 3T Philips Achieva system. The study acquired T2*-weighted (fMRI) single-shot echo planar imaging (EPI) sequence [repetition time (TR) = 3000 ms, echo time (TE) = 30 ms, flip angle = 90°, in-plane SENSE acceleration 2, $1.88 \times 1.88 \times 5 \text{ mm}^3$ voxels, matrix size 128×128 with 30 slices, field of view (FOV) = 240 mm, and 226 volumes] followed by a T1-weighted fast-field echo 3D structural scan for normalization (TR = 8.2 ms, TE = 3.8 ms, inversion time (TI) = 1018 ms flip angle = 8°, $0.94 \times 0.94 \times 1 \text{ mm}^3$ voxels, matrix size 240×240 with 160 slices and FOV = 240 mm). The images during diffusion-weighted tensor sequences were acquired along 16 gradient directions ($b = 800 \text{ s/mm}^2$, number of excitations = 2) along with

an unweighted ($b = 0$) image with an overall of 17 volumes. DTI images were recorded as a series of 66 axial slices, using a single-shot spin EPI sequence with the following parameters: TR = 7151 ms, TE = 55 ms, flip angle = 90° , voxel size = $2 \times 2 \times 2 \text{ mm}^3$, matrix size = $224 \times 224 \times 132$ and FOV = 224.

The fMRI sequence scan was conducted as a block design of 3×3 min 'on' periods, interspersed by 4×30 s rest or 'off' periods. The 'on' periods consisted of the Paced Auditory Serial Addition Test (PASAT) task (Cook et al., 2007), employed to temporarily induce mental fatigue. During the task, patients are instructed to listen to a sequence of numbers from 1 to 9 and sum consecutive numbers such as third to the fourth, and fourth to the fifth. Afterwards, participants respond with a press of a button whenever the sum of two numbers is 10. Simultaneously, three alternating numbers are used as a distractor on the screen which the participants are asked to focus upon. To assess the task, the number of correct responses to the calculation task was recorded.

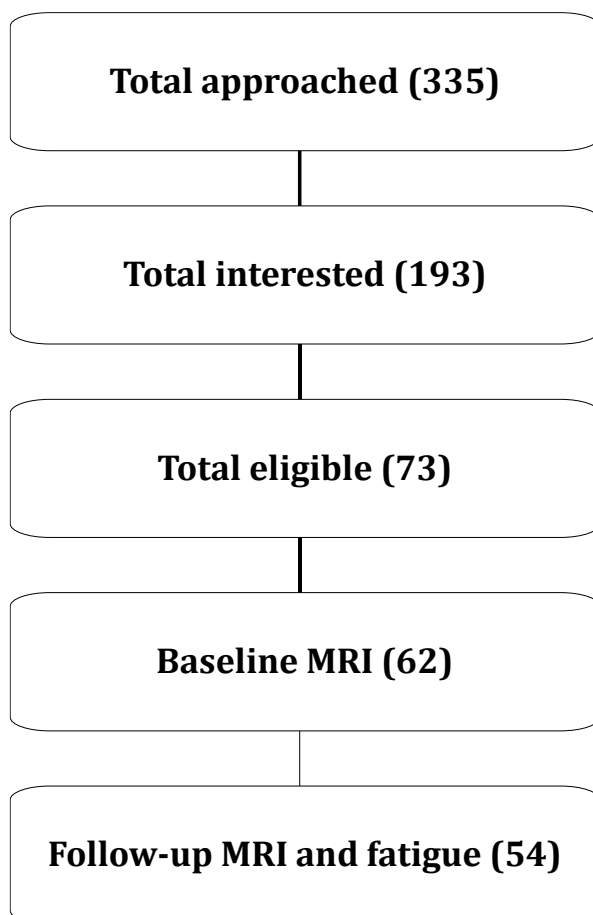


Figure 2-1: Flow chart of RA study 1. RA patients attending a UK regional rheumatology service were approached. Patients who showed interest were then screened for eligibility such as fatigue levels and contra-indications to MRI. The patients that were considered in the analysis needed to have baseline and follow-up MRI as well as fatigue assessments.

2.1.2 RA trial dataset (LIFT)

The dataset consisted of a subset of participants in a randomised control trial (NCT03248518) to test the effectiveness of remotely delivered CBA and PEP compared to UC (Martin et al., 2019), which was called “Lessening the Impact of Fatigue in Inflammatory Rheumatic Diseases: A Randomised Trial” (LIFT). Participants were randomly allocated to receive either of the two treatments or UC alone (1:1:1 ratio) using a computer-generated sequence. The recruitment consisted of RA patients attending major secondary care rheumatology services in the UK and consent to take part in the trial and MRI sub-study. Recruitment for the study began in August 2017 and concluded in September 2019 and all scanning took place in Aberdeen, Edinburgh, and Glasgow. To participate in the study, patients needed to be at least 18 years old, have provided consent for the parent trial randomisation, meet the classification criteria for RA per the 2010 American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) guidelines, report significant fatigue persisting for over three months and rated ≥ 6 on a 1-10 Visual Analogue Scale during screening, and be considered to have stable RA, as defined by unchanged immunomodulatory therapy in the preceding three months. Exclusion criteria included alternative medical explanations for their fatigue and MRI contradictions. All participants had a similar battery of tests to the RA study dataset, along with T1, T2* (resting-state and PASAT task) and diffusion MRI sequences (**Table 2-1**) at baseline and six months follow-up after the interventions. Recruitment is displayed in **Figure 2-2**. Research personnel undertaking outcome assessments and scans were blind to treatment allocation.

Modality	<i>sMRI</i>		<i>fMRI</i>		<i>Diffusion MRI</i>	
Vender	<i>Philips</i>	<i>Siemens</i>	<i>Philips</i>	<i>Siemens</i>	<i>Philips</i>	<i>Siemens</i>
Sequence	<i>Achieva</i> 3D SPACE	<i>PRISMA</i> 3D MP- RAGE	<i>Achieva</i> T2* FFE- EPI	<i>PRISMA</i> T2* GRE- EPI	<i>Achieva</i> SE-EPI	<i>PRISMA</i> SE-EPI
TR (ms)	8.2	2500	1950	1950	7010	7600
TE (ms)	3.8	4.37	26	26	90	74
TI (ms)	1025.7	1100				
FA (deg)	8	7	70	70	90	
FOV (mm)	256 x 256	256 x 256	240 x 240	240 x 240	220 x 220	240 x 240
Matrix size	256 x 256	256 x 256	128 x 128	128 x 128	96 x 96	120 x 120
Slices	160	160	30	30	60	69
Voxel size (mm)	0.94 x 0.94 x 1	1 x 1 x 1	1.88 x 1.88 x 3.5	1.9 x 1.9 x 3.5	2.29 x 2.29 x 2.3	2 x 2 x 2
Slice gap (mm)	0	0	1.5	1.5	0	0
Volumes			308	308		
Gradient directions					64 (b=1200 s/mm ²)	64 (b=2000 s/mm ²)
Volumes (b=0 s/mm²)					8	
Scan time	5:53	3:45	10:12	10:08	09:27	08:38

Table 2-1: MRI sequences of the RA trial (LIFT) dataset. Abbreviations: FA, flip angle; fMRI, functional magnetic resonance imaging; FOV, field of view; sMRI, structural magnetic resonance imaging; TE, echo time; TI, inversion time; TR, repetition time.

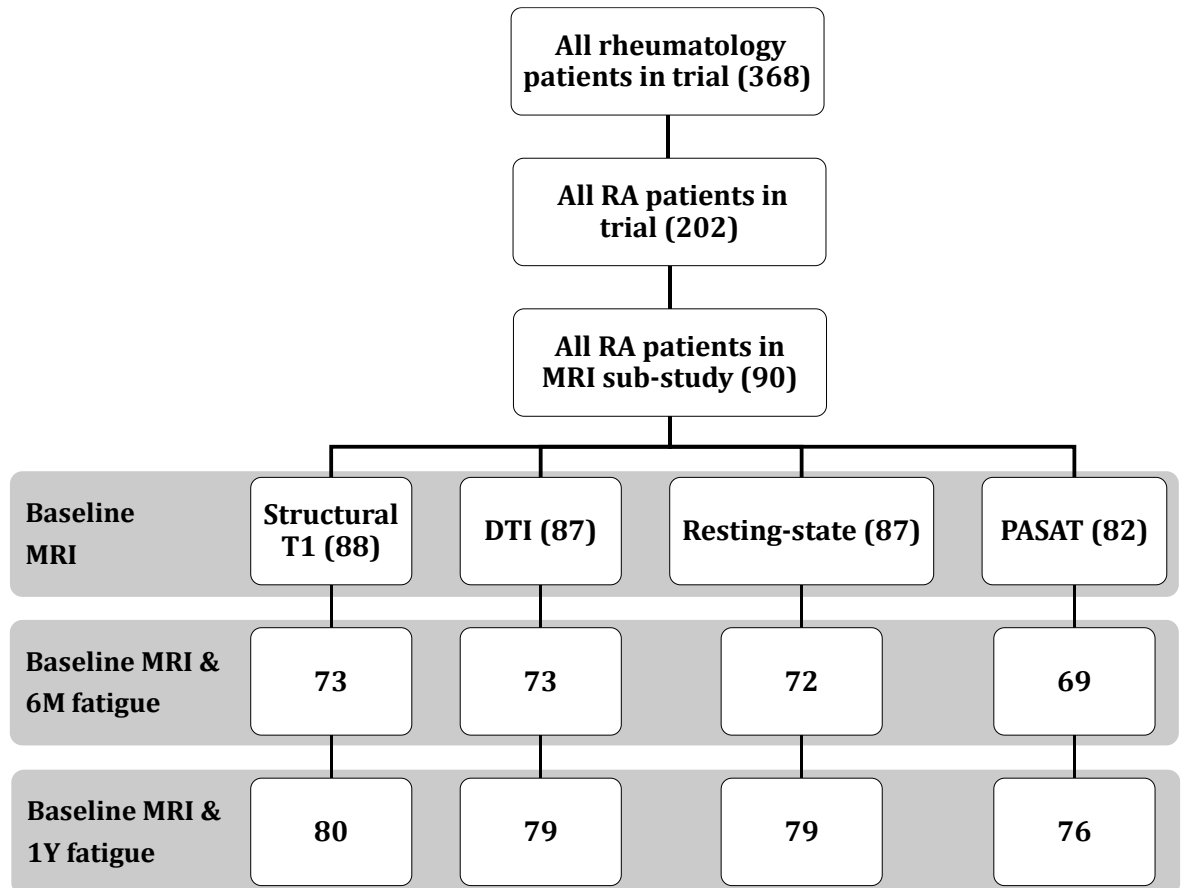


Figure 2-2: Flow chart of RA trial. The randomised control trial approached patients with inflammatory rheumatic disease, including RA. The patients with RA could then also participate in the MRI sub-study of this trial. Not all MRI sequences were conducted at baseline. Two patients who were enrolled in the sub-study, eligible and consented to MRI but subsequently withdrew from the study. One patient completed a T1 sequence but did not complete other sequences due to scanner technical issues. The PASAT task sequence was not available for patients scanned in Glasgow while other patients did not complete the sequence in Aberdeen and Edinburgh. There were drop-outs in follow-up fatigue assessments due to unavailability for clinical visits at 6M (six months) and 1Y (one year).

2.1.3 PsA study dataset

The study recruited participants with PsA with active disease who fulfilled CASPAR criteria (Taylor et al., 2006) and were due to start a new DMARD as part of their standard clinical care. Exclusion criteria included any contraindication to MRI or evidence of peripheral neuropathy. The study was designed to be part of a series of studies in multiple rheumatic diseases to characterise centralised pain. Eligible patients who agreed to participate had to attend a clinical research facility in Glasgow at baseline, three and six months. At the baseline visit, the patients undertook a clinical assessment, blood samples, questionnaires, and an MRI brain scan. Fatigue assessments included the Patient-Reported Outcomes Measurement Information System (PROMIS), specifically the Fatigue Fibromyalgia (FM) Profile (Kratz et al., 2016). At the three- and six-month visits, only the clinical assessments, blood sampling and questionnaire assessments were performed. Recruitment for the study is

shown in **Figure 2-3**. This was a longitudinal study that aimed to investigate centralised pain in PsA patients and how it changes with pharmacological treatments.

The patients undertook scans on a 3 Tesla Siemens PRISMA (Siemens, Erlangen, Germany) in Glasgow UK using a 32 channels phased-array head coil. These included a T1-weighted fast-field echo 3D structural images for normalization (TR = 2500 ms, TE = 2.88 ms, inversion time (TI) = 1070 ms, flip angle = 8°, FOV = 256 mm, with 176 slices, 1 mm iso-voxel) and a functional images at rest using a T2*-weighted multiband EPI sequence (TR = 800 ms, TE = 30 ms, flip angle = 52°, FOV = 216 mm, acceleration factor 6, 60 slices, 440 volumes at 2.4 mm iso-voxel).

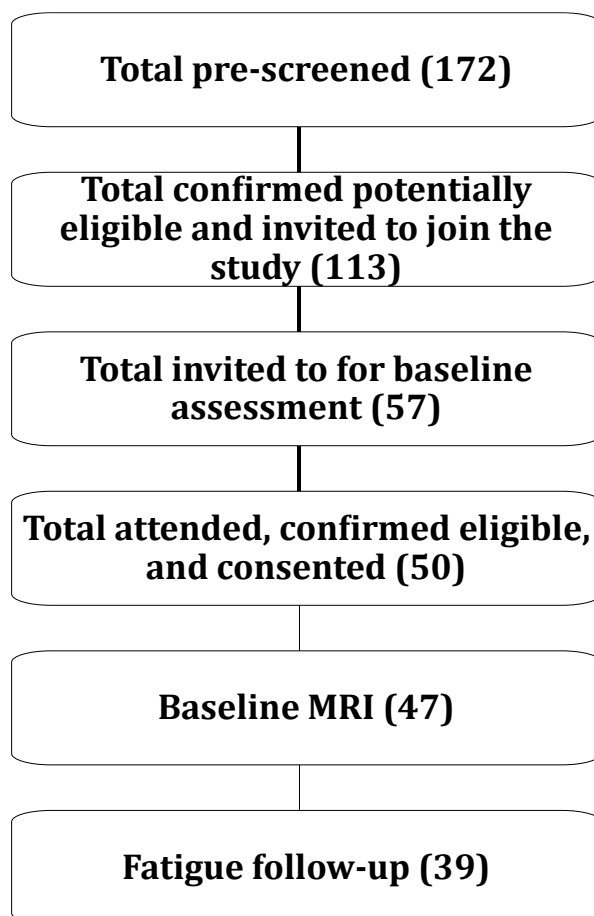


Figure 2-3: Flow chart of PsA study. Patients with PsA were approached, confirmed eligible and consented. Three patients did not tolerate MRI scanning and while a number of patients did not complete follow-up visits at six months to record their fatigue levels.

2.1.4 RA study dataset 2

Similar to the PsA study cohort, this study was also designed to characterise centralised pain in rheumatic disease patients. Patients in this dataset were recruited as part of a project funded by the National Institute of Health (NIH) and run by the University of Michigan (Grant code P50AR070600). Patients had to fulfil the 2010 American College of

Rheumatology classification. Similar to the PsA cohort, the RA patients in this dataset were due to start a new DMARD as part of their standard clinical care but were excluded if they had any MRI contraindications or peripheral neuropathy. They also had three- and six-month visits in which at least questionnaire measures were recorded, with recruitment shown in **Figure 2-4**. Fatigue was measured using the PROMIS Fatigue-FM Profile. The patients undertook compatible scans to ones from the PsA study but on a 3 Tesla Signa Discovery MR750 scanner (GE Healthcare) at the University of Michigan using a 32 channels phased-array head coil. The T1-weighted images were acquired using a spoiled gradient recalled echo sequence (TR = 4.9 ms, TE = 2 ms, inversion time (TI) = 1060 ms, flip angle = 8°, FOV = 256 mm, with 208 slices, 1 mm iso-voxel). Functional images at rest were acquired using the same T2*-weighted multiband EPI sequence as the PsA study cohort (TR = 800 ms, TE = 30 ms, flip angle = 52°, FOV = 216 mm, acceleration factor 6, 60 slices, 440 volumes at 2.4 mm iso-voxel).

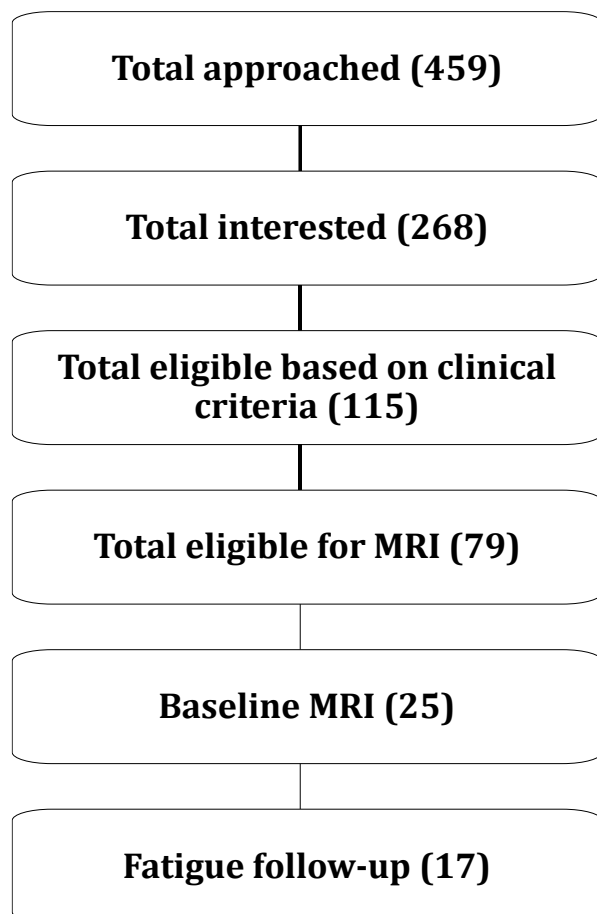


Figure 2-4: Flow chart of RA study 2. Patients with RA were approached, confirmed eligible and consented. A number of patients did not complete follow-up visits at three months to record their fatigue levels.

2.2 Preprocessing methods

Preprocessing of functional images was exclusively done using the functional connectivity (CONN) toolbox v19 (Nieto-Castanon, 2020), which is based on Statistical Parametric Mapping (SPM, <https://www.fil.ion.ucl.ac.uk/spm/>) and MATLAB (R2019a. The MathWorks, Inc., Natick, Massachusetts, United States).

2.2.1 Realignment

The procedure co-registers all functional volumes and resamples them to a reference volume (first scan) via a b-spline interpolation. This addresses potential susceptibility distortion-by-motion interactions by estimating the derivatives of the deformation field with respect to head movement and resampling the functional data to match the deformation field of the reference image.

2.2.2 Slice-timing correction

Temporal misalignment between different slices within one volume of the functional data is introduced by the sequential nature of the fMRI acquisition protocol. This is corrected by time shifting and resampling the functional data using sinc-interpolation to match the time in the middle of each acquisition time.

2.2.3 Outlier-detection

Tools detect potential outlier scans by thresholding the observed global BOLD signal and the amount of subject motion in the scanner that was quantified in the realignment phase. Framewise displacement is computed at each timepoint by considering a bounding box around the brain and estimating the largest displacement among six control points placed at the centre of this bounding box faces. Global BOLD signal change is computed at each timepoint as the change in average BOLD signal within the global-mean mask scaled to standard deviation units.

2.2.4 Coregistration

Functional data is co-registered to the structural data to improve localisation. This is done by estimating an optimal affine transformation between the reference functional image (mean BOLD signal) and the reference structural image (T1-weighted volume) that maximizes the mutual information between the two, storing this information in the functional image voxel-to-world mapping header information without resampling the data.

2.2.5 Segmentation and normalization (MNI-space)

Many analysis pipelines use a reference coordinate system. The most used reference space is the Montreal Neurological Institute and Hospital (MNI) 152 template based on an average of 152 people (Mazziotta et al., 1995). Functional and anatomical data are normalized into standard MNI space and segmented into grey matter, white matter, and CSF tissue classes using a unified segmentation and normalization procedure. This procedure iteratively performs tissue classification, estimating the posterior tissue probability maps from the intensity values of the reference functional/anatomical image, and registration, estimating the nonlinear spatial transformation best approximating the posterior and prior tissue probability maps, until convergence. Direct normalization applies this unified segmentation and normalization procedure separately to the functional data, using the mean BOLD signal as a reference image, and to the structural data, using the raw T1-weighted volume as a reference image. Both functional and anatomical data are resampled to a default bounding box, with 2mm isotropic voxels for functional data and 1mm for anatomical data, using 4th-order spline interpolation.

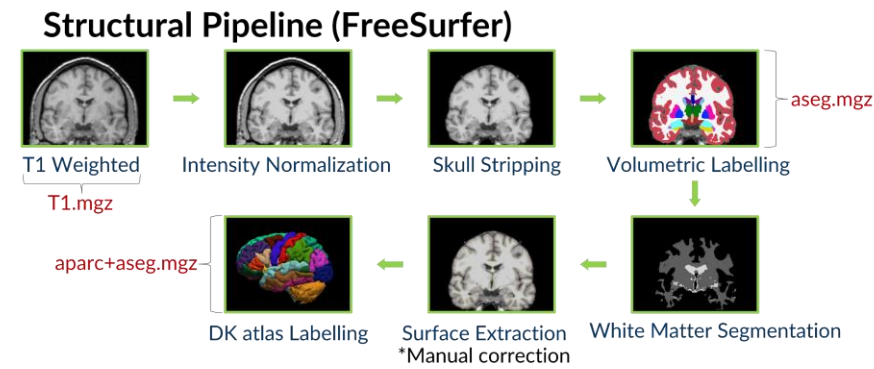
2.2.6 Smoothing

Functional data is smoothed using spatial convolution with a Gaussian kernel of 8mm full width half maximum, in order to increase the BOLD signal-to-noise ratio and reduce the influence of residual variability in function and anatomy across subjects.

2.2.7 Subject-space preprocessing

Alternative to MNI-based preprocessing, the data was also pre-processed in subject-space using the outputs the Freesurfer software package v6.0 (<http://surfer.nmr>.

mgh.harvard.edu/), with technical details described elsewhere (Fischl, 2012). Pre-processing included motion correction, intensity normalization, Talairach registration, skull stripping, subcortical segmentation and labelling, segmentation of white matter, delineation of grey matter/white matter and grey matter/CSF boundaries, automated topology correction, surface deformation, and cortical surface reconstruction. The reconstructed surfaces of each subject were visually inspected for segmentation inaccuracy, such as pial and white matter boundary errors, and manually corrected with subsequent repeated processing. Each subject's cortical surface was parcellated into 34 gyral-based ROIs per hemisphere according to the Desikan–Killiany atlas (Desikan et al., 2006). Surface area (in mm^2), cortical thickness (in mm) and cortical grey matter volumes (in mm^3) were calculated at the ROI level of each subject. Eight subcortical ROIs per hemisphere were segmented using a probabilistic atlas (Fischl et al., 2002), and their volumes were extracted. These maps could then be used in the CONN toolbox (**Figure 2-5**).



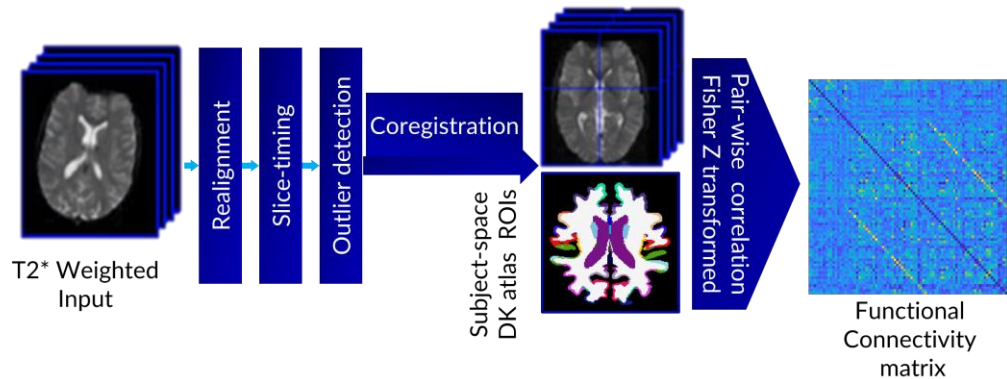
Inputs to functional and structural connectivity analyses

T1.mgz: Raw structural

aseg.mgz: Grey matter, White matter, CSF masks

aparc+aseg.mgz: Subject-space cortical and subcortical ROIs from DK atlas

Functional Connectivity (CONN toolbox)



Structural Connectivity (FSL)

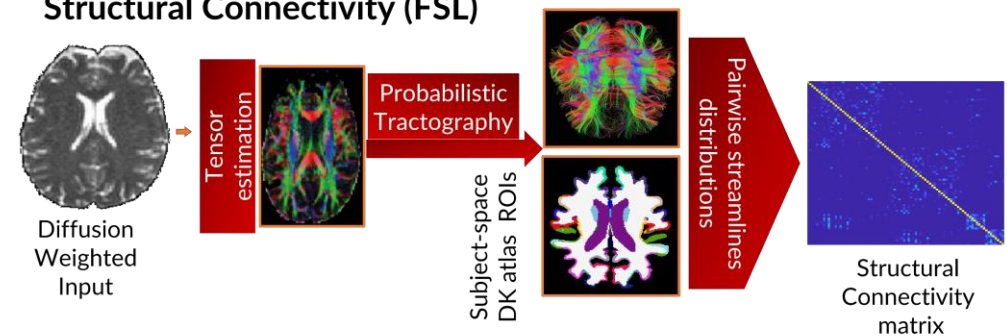


Figure 2-5: Subject-space preprocessing pipelines. The pipeline includes preprocessing of structural (T1-weighted) scan through FreeSurfer software (recon-all command). Its outputs are then used by functional connectivity software (CONN toolbox) and structural connectivity software (FSL) to extract matrices from DK atlas ROIs in subject-space without normalization. Abbreviations: CSF, cerebrospinal fluid; DK atlas, Desikan-Killiany atlas, ROIs, regions of interest.

2.2.8 Denoising

After the functional data has been pre-processed, the BOLD signal often still contains a considerable amount of noise or non-neural variability due to a combination of physiological, outlier, and residual subject-motion effects. These residual factors are particularly problematic in the context of FC analyses because they introduce very strong and noticeable biases in all FC measures. Because of this, conventional preprocessing steps in the context of FC have favoured considerably more conservative strategies than those often found in activation-based fMRI analyses, focusing on eliminating or at least minimizing the influence of these residual noise components in the BOLD signal. These additional strategies are often framed under the general umbrella term of denoising. They would include linear regression of potential confounding effects in the BOLD signal and temporal band-pass filtering.

Factors that are identified as potential confounding effects to the estimated BOLD signal are estimated and removed separately for each voxel and for each subject and functional run/session using ordinary least squares regression to project each BOLD signal timeseries to the sub-space orthogonal to all potential confounding effects. An anatomical component-based noise correction procedure (CompCor) would remove the noise components from cerebral white matter and CSF areas (Behzadi et al., 2007). They are defined by a one-voxel binary erosion step to the masks of voxels with values above 50% in white matter and CSF posterior probability maps. Within each area five potential noise components (Chai et al., 2012) are estimated: the first computed as the average BOLD signal, and the next four computed as the first components in a Principal Component Analysis of the covariance within the subspace orthogonal to the average BOLD signal and all other potential confounding effects. The regression would also remove subject-motion parameters (Friston et al., 1996), as well as identify outlier scans or scrubbing (Power et al., 2014).

Temporal frequencies below 0.008 Hz or above 0.09 Hz are removed from the BOLD signal in order to focus on slow-frequency fluctuations while minimizing the influence of physiological, head motion and other noise sources. Filtering is implemented using a discrete cosine transform windowing operation to minimize border effects, and performed after regression to avoid any frequency mismatch in the nuisance regression procedure (Hallquist et al., 2013). The denoising procedure can be run for both the MNI-space and the subject-space pipelines in the CONN toolbox (**Figure 2-6**).

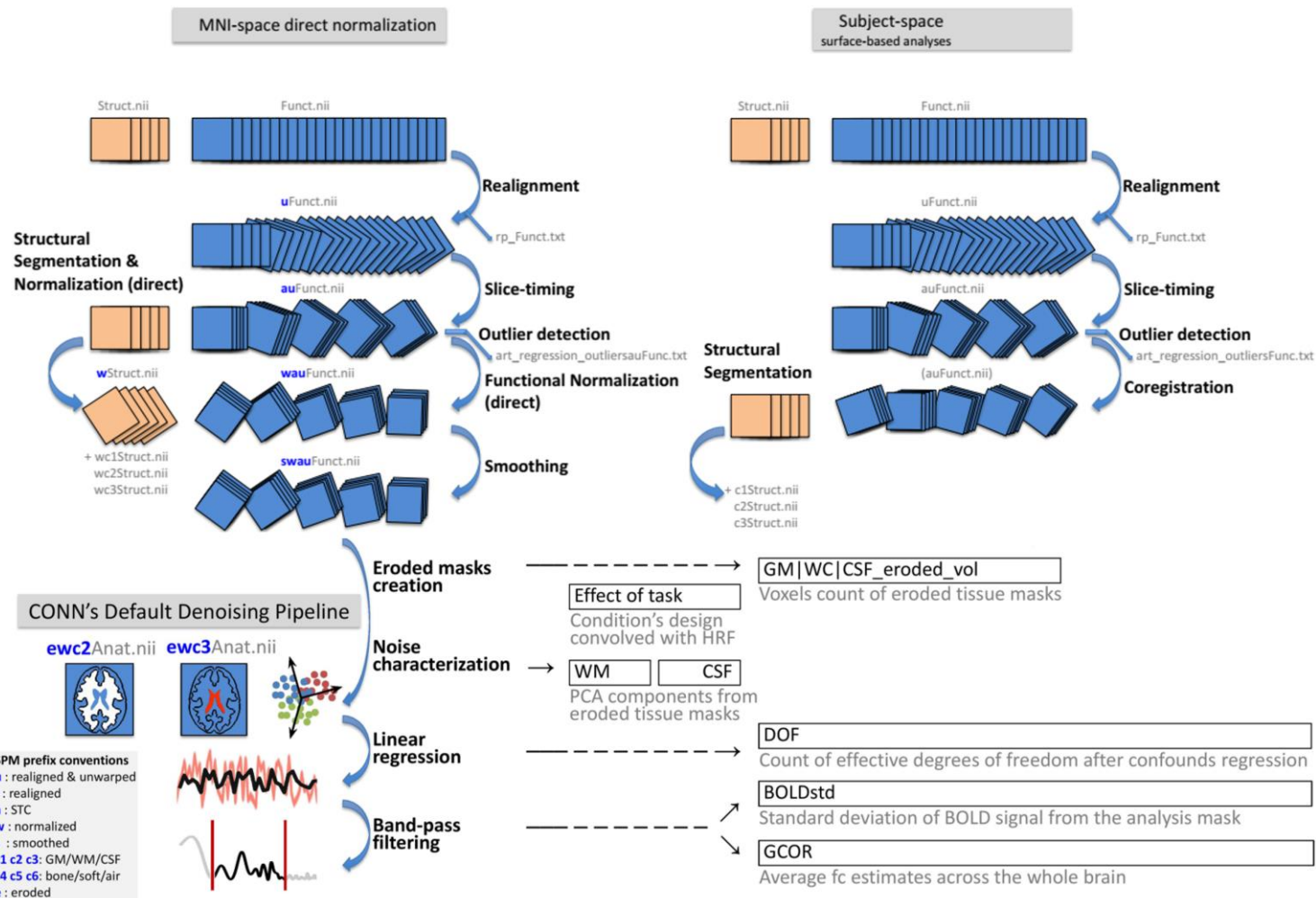


Figure 2-6: MNI-and subject-space processing of functional imaging data. A schematic displaying the differences between MNI-space and subject-space preprocessing and the common denoising pipeline in the functional connectivity (CONN) toolbox (Morfini et al., 2023).

2.2.9 Diffusion Tensor Imaging (DTI)

Diffusion data were pre-processed with the FMRIB Software package (FSL v6.0 and FSL Diffusion Toolbox (FDT) (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>). The pre-processing procedure included skull stripping, correcting for eddy current distortion and head movement, using affine registration of each diffusion-weighted image to the b=0 image, calculating FA, estimating the probabilistic distributions of fiber orientations from each voxel, using the GPU version of BEDPOSTX tool (Hernandez et al., 2013), and performing probabilistic tractography to estimate the connectivity probability among 84 cortical and subcortical regions using the PROBTRACKS tool (Behrens et al., 2007). Within each of the seed regions, 5000 fiber streamlines per voxel were initiated from the probability distribution of the principal fiber direction (estimated by BEDPOSTX) with fiber tracking parameters of 0.5 mm step size, $\pm 80^\circ$ curvature threshold, and tracking stopped after a maximum of 2000 steps. A white matter waypoint mask was provided to consider only streamlines that pass through. The connectivity probability between seed region i and target region j was then calculated as the ratio of the number of fibers passing through target region j to the total number of fibers from seed region i , which yields an 84X84 asymmetrical SC matrix for each subject, based on a 84-ROI brain atlas (**Figure 2-5**).

2.3 General Linear Model (GLM)

Linear regression is a simple and powerful method to solve regression problems (Hastie et al., 2009). Linear regression is defined by equation (2.1), where y denotes the target outcome variable, x the input features, β_0 the intercept (value of y when $x=0$), β the slope coefficients of the input features x , and ϵ the error term also called the residual (**Figure 2-7**).

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \epsilon \quad (2.1)$$

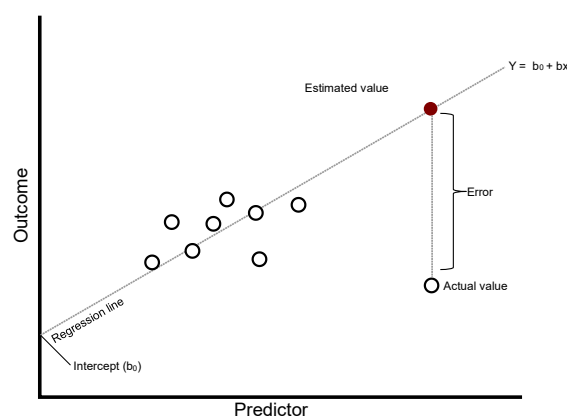


Figure 2-7: Elements of linear regression.

Given a simple linear regression between two variables, such as using height to predict weight), the GLM expands this framework so that each of the four terms (x , y , β_0 , β) can represent sets of variables rather than just single ones (Mardia et al., 1980). Consequently, the GLM defines a multivariate linear association between a set of explanatory measures (X) and a set of outcome measures (Y). In FC analyses, an outcome variable $y[n]$ typically comprises a row of FC values from the n -th subject, whereas an explanatory variable $x[n]$ consists of a row of behavioural, demographic, or group variables for that same subject (Nieto-Castanon, 2020). The matrix B captures the associations between the explanatory and outcome measures while the error term accounts for factors predictive of the outcome Y , unexplained by the explanatory measures of X . The GLM makes assumptions that the error term is independent across individuals and normally distributed, with a mean of zero and an arbitrary variance-covariance structure across outcome measures. As matrix B remains constant across subjects, having a sufficient dataset enables us to compute an unbiased estimate of B through an Ordinary Least Squares solution (2.2).

$$\widehat{B} = (X^t X)^{-1} X^t \quad (2.2)$$

GLMs specify and evaluate hypotheses of the form " $CBM'=D$ ", where users define arbitrary contrast matrices C , M and D . Because each column in matrix C corresponds to the same column in X , it is possible to formulate hypotheses that target specific combinations of explanatory measures X . The same principle applies to matrix M and its correspondence to outcome columns in Y to hypothesise about different combinations of Y . Finally, D is set to zero to test the null hypothesis within the explanatory and outcome measures.

When working with the C , M and D matrices, the hypothesis $CBM'=D$ is tested using the Wilks' Lambda statistic. The statistic compares the ratio of the residuals between the null-hypothesis model and an unconstrained model, with values spanning from 0 and 1. The resulting lambda value would then be set against the Wilks' Lambda distribution, defined as the expected lambda values if the null hypothesis was true, and accept the alternative hypothesis if the value is over a predetermined threshold of statistical significance (e.g., $p < 0.001$). The functional CONN toolbox can implement this framework to apply an array of classical analyses, by using GLMs and specifying the X and Y measures as well as the C and M matrices (**Table 2-2**).

GLMs can be used as predictive models, which here we define as models that estimate a mapping function for a future outcome, given the input data recorded at baseline. In the

following chapters, GLMs will be mostly used as univariate models that use single imaging metrics as a Y column to identify the strongest independent predictors of fatigue in IA patients. As X columns, the outcome along with confounding factors like age and gender will be used within a C matrix, which controls for confounders while setting the outcome as the effect of interest. Such an approach is frequently used to identify predictors of clinical outcomes using neuroimaging data (Stoeckel et al., 2014). For example, one study used this framework to predict symptom recovery in psychiatric inpatients using baseline FC between the amygdala and frontal brain regions, while controlling for age, sex, and length of stay (Venta et al., 2018). Another study predicted the response to transcranial direct current stimulation for memory improvement in older adults using baseline FC between the hippocampus and the temporoparietal cortex (Antonenko et al., 2019). To cope with the false positive rate of the multiple comparisons that need to be made across imaging metrics, the false discovery rate (FDR) approach is commonly applied in neuroimaging studies (Genovese et al., 2002). It considers the proportion of false positives in all the rejected tests while assuming independence between tests (Benjamini and Hochberg, 1995). The FDR approach is flexible as it can be used with any valid statistical test and results in increased power compared to other approaches like the family-wise error rate.

Term	Definition	Example
Subject-effects	explanatory measures (X columns)	Subject-level covariates (e.g., age, IQ, treatment, placebo, behavioural outcome). Additionally, a variable containing the value 1 for every subject is used to control for constant effects across all subjects.
Between-subjects contrasts	explanatory measures to be evaluated (C matrix)	A vector with as many elements as explanatory measures, 1's for effects of interest and 0's for all other elements (e.g., treatment and placebo, a [1, 0] contrast specifies evaluating the treatment effect only. A [1, -1] contrast compares the effect of treatment to that of placebo.
Conditions	outcome measures (Y columns)	A combination of choosing which first-level functional connectivity measures, and which experimental conditions (if applicable), we would like to evaluate.
Between-conditions contrasts	Outcome measures to be evaluated (M matrix)	Similar to between-subject contrasts, but across conditions instead of across-subject-effects (e.g., pre- and post-intervention functional connectivity).

Table 2-2: General linear model design in the functional connectivity CONN toolbox.

2.4 Prognostic model building and evaluation

2.4.1 Bias-variance trade-off

Several challenges arise when the goal is to build a model that can best predict an outcome on unseen data given a complex relationship between the outcome and the predictors. Simple linear models make strong assumptions about the relationship between predictors and outcomes, which facilitates the learning and understanding of the input-output mapping function but increases model bias (Hastie et al., 2009). In this context, bias refers to a model's inability to capture the true relationship between input and output variables. High-bias models typically underfit the data, which indicates they perform poorly on problems that do not fulfil the algorithm's simplifying assumptions. To reduce this bias, one can add more predictor variables that improve model accuracy, learn more complex data by expanding the training sample size, or employ more flexible algorithms that make fewer assumptions and can learn non-linear relationships (Greener et al., 2022).

Model bias manifests in training performance, while model variance—predictions in new data. High variance indicates unreliable models that heavily rely on the training data, including the number and types of parameters, rather than effectively extracting the underlying patterns. Minor changes in training data significantly impact the output of high-variance models, limiting their ability to generalise to new observations. Therefore, variance describes how the input-output mapping function adapts to different training datasets. To reduce model variance, one can reduce the number of predictor variables, augment the training data diversity to balance model complexity or use simpler models like linear regression.

High-bias models offer consistent but inaccurate predictions on average, while high-variance models provide inconsistent but accurate predictions on average. Model bias and variance are interdependent; increasing bias reduces variance, and increasing variance reduces bias. The trade-off between these two characteristics requires an ideal model to strike a balance between a reasonable amount of both bias and variance (**Figure 2-8**). The choice of algorithms and their configuration can adjust the trade-off to align with the underlying problem. In practice, real data does not allow us to directly calculate the bias and variance due to the unknown nature of the data's underlying mapping function. Nevertheless, the concept of the bias-variance trade-off informs why ML algorithms exhibit certain behaviours and offers guidance for improving their performance.

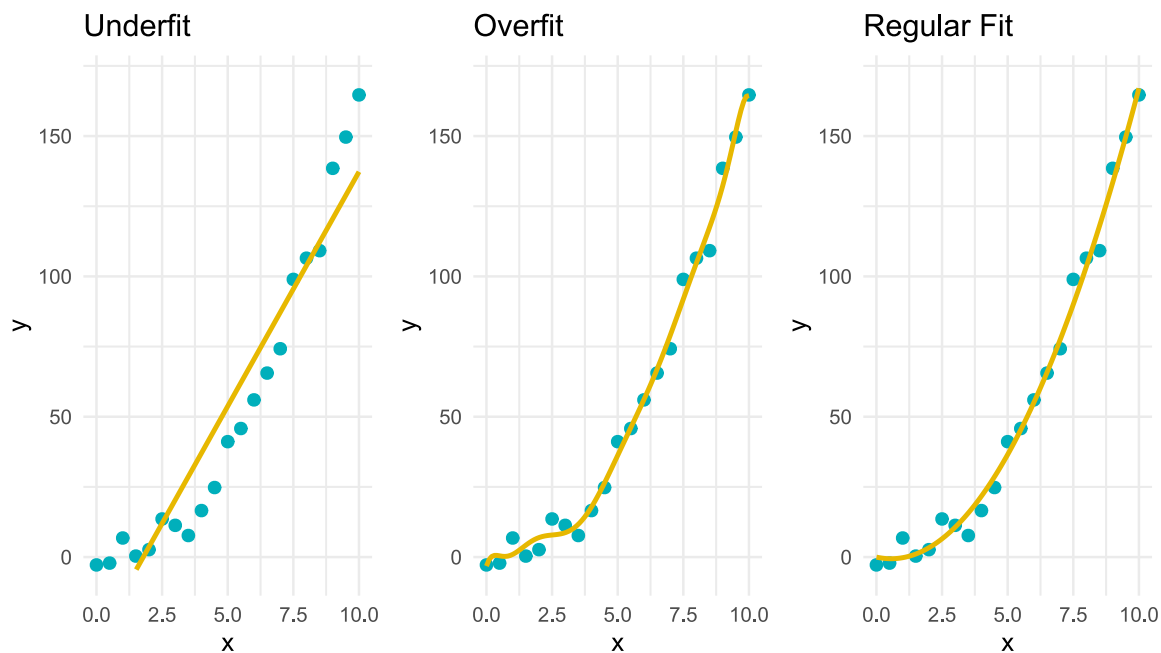


Figure 2-8: Bias-variance trade-off based on model fit. The figure depicts the fits of polynomial regression models on synthetic data using first-degree (underfit), tenth-degree (overfit), and second-degree (regular fit) polynomials. These plots act as examples of high bias (underfit) and high variance (overfit) models and the necessity for a trade-off between them (regular fit). The figure was created in R Studio (code: https://github.com/krisbg95/LIFT/blob/main/ML/Bias_variance.R).

2.4.2 Feature selection

Feature selection is the process of reducing the number of input variables when developing a predictive model. It addresses the “curse of dimensionality”, the exponential growth in computational cost as the dimensionality of data increases (Venkatesh and Anuradha, 2019). Feature selection reduces dimensionality by removing irrelevant or redundant features to select the best subset of features in a certain prediction problem, leading to an improvement in the learning performance and a reduction of the computational cost. Feature selection algorithms are divided into three categories: filter, wrapper, and embedded methods (Pudjihartono et al., 2022).

Filter methods perform feature selection independently from the ML algorithm. They evaluate the statistical relationship between each input variable and the target variable and then select the variables with the strongest relationship to the target variable. The choice of statistical measures depends on the data type of both the input and output variables. Filter methods can be fast and effective in improving learning performance and have a low risk of overfitting. However, they ignore the interaction between features due to their independence from the prediction model (Bommert et al., 2020). Wrapper methods rely on the ML algorithm performance to select the best subset of features. They create many models with

different subsets of input features and select those features that result in the best-performing model according to a performance metric. Since they assess the model during the selection process, they can detect feature dependencies and consequently generalize better than filter methods. The drawbacks of wrapper methods are that they are more likely to overfit the data because they involve training ML models with different combinations of features and require both higher computational power and longer training times. Finally, some ML algorithms perform feature selection automatically as part of learning the model. These models use embedded feature selection and include regularization algorithms like lasso and elastic net regression, which introduce additional constraints into the optimization of a predictive algorithm that bias the model toward lower complexity (fewer coefficients). Instead of additional constraints, random forest models sample both observations and features and keep only a random subset of them to build the tree. This sampling results in all trees looking at different information to make their decisions, which reduces the correlation between the different returned outputs and potentially improves the learning performance in the process.

Feature selection in this project was performed using Relief F algorithm. Relief algorithms are filter feature selection methods that rank features according to how well they discriminate an observation from the neighbour of a different class. Relief F is a more recent version which addresses more noisy data by using multiple neighbours (**Figure 2-9**), while the regression version differentially weights each observation based on their distance from the other observations. As a filter method, it offers faster processing than other feature selection methods and functional independence that allows integration with any modelling algorithm (Urbanowicz et al., 2018). This is crucial, given that there's no one-size-fits-all modelling algorithm, particularly in clinical applications (Watson et al., 2019). Relief algorithms can adapt to feature sets of different sizes, as they don't fixate on identifying an optimal minimum subset size. While feature construction is another approach for dimensionality reduction, where new features are defined as a function of two or more features, it alters the original features to the point of making them unrecognizable, posing challenges for downstream model interpretability. Unlike other individual evaluation methods that do feature weighting, Relief algorithms can capture feature dependencies or interactions. Instead of exhaustively searching through feature combinations, they employ the concept of nearest neighbours to derive feature statistics that indirectly address interactions. Relief algorithms possess the advantages of being nonmyopic, providing estimates of feature quality in the context of other features, and non-parametric, making no assumptions about population distribution or sample size. The ReliefF algorithm also estimates values that are missing in a dataset based

on the values of existing data points. On the flip side, Relief algorithms don't address feature redundancies, selecting all relevant features regardless of strong correlations.

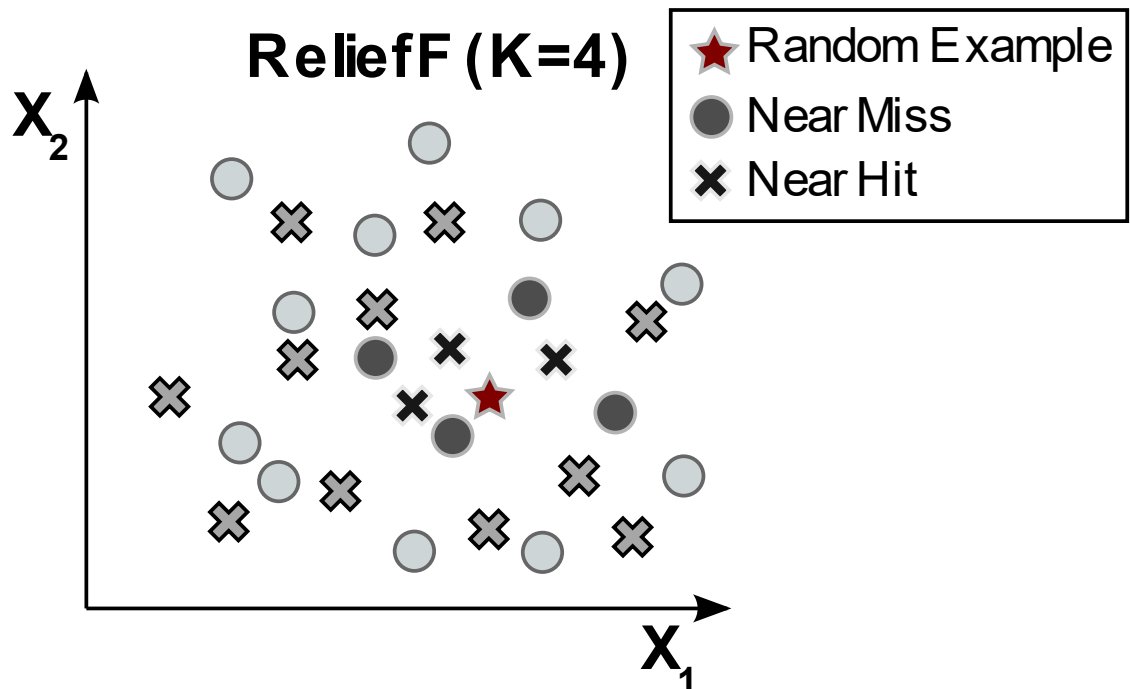


Figure 2-9: Relief feature selection. The process of Relief algorithms consists of starting with p different features and a binary target. Then, we're starting by initializing a weights vector of length p with all 0s. Then, for user-defined m out of n observations, we (1) draw a random example from our data, say from class 0, (2) find the closest example in terms of distance, which belongs to the same class (called near hit), (3) take the closest point, which belongs to the opposing class (called near-miss), (4) update the weight vectors for m iterations and p features. ReliefF algorithm considers user-defined k nearest hits and nearest miss neighbours instead of one. This adds weight estimate reliability, concretely in noisy data. Additionally, the parameter m is equalized to the number of observations n , so all observations in the dataset are selected to be the target observation once, which also reliably increases the weight estimates. The figure was created using Inscape (2020), adapted from Urbanowicz et al. (2018).

2.4.3 Machine learning algorithms

The number of different ML algorithms continues to grow, as computing resources continue to improve, and new datasets are acquired. Deep learning is at the forefront of this expansion but due to its complexity and lower interpretability, it requires much larger datasets than the neuroimaging datasets of interventions in IA, described in **Chapter 2.1**. The final aim of this project was to predict individual fatigue scores in RA patients six months and one year after receiving usual care or fatigue-specific treatments using baseline neuroimaging and clinical data. Therefore, the selected regression models needed to: (1) use characteristics of the brain that are associated with fatigue (e.g., functional and structural connectivity), (2) incorporate treatment allocation and clinical data, and (3) identify which predictors were most important for the model predictions. I selected models with different underlying methods that can

differently detect linear and non-linear relationships (**Table 2-3**), with similar selections used in studies using brain imaging to predict age (Sone and Beheshti, 2022) and depression (Lee et al., 2022) or using clinical data to predict outcomes in IA (Madrid-Garcia et al., 2023). These will be described below and include regularized linear regression, three-based methods (random forest, gradient boosting), support vector regression, and the Bayesian method of Gaussian process regression.

Algorithm Class	Examples	Description
Classic regression	Linear regression	Linear regression models the relationship between one or more features and a continuous outcome by fitting a regression line that minimizes the sum of all the residuals, which are the distances between each feature in the training data and the line being fitted to model them.
Regularized regression	Lasso, ridge regression, elastic net	An extension of the classic regression algorithms in which a penalty is imposed on the fitted model to reduce its complexity and decrease the risk of overfitting.
Tree-based	Classification and regression trees, random forest, gradient boosted trees	Decision trees are a sequence of “if-then-else” splits that are derived by iteratively separating the data into groups based on the relationship of the features with the outcome. Random forest and gradient boosted trees are example of ensemble tree models. Ensemble models combine the output of many trained models to estimate an outcome.
Support vector machines	Linear, polynomial, radial basis kernel	Represents the data in a multidimensional feature space and then fits a “hyperplane” that best separates the data based on the outcomes of interest.
Bayesian	Naive Bayes, Gaussian Process Regression	Use Bayes’ theorem of conditional probability, which is the probability that something will happen given that something else has already occurred. Bayesian algorithms work by iteratively updating the probability of an outcome (or posterior belief) given new data.

Table 2-3: Description of prediction algorithm types. The table is adapted from Sanchez-Pinto et al. (2018).

Linear regression overfits data when the number of variables surpasses the number of observations and inflates the variance of the model. Regularization techniques like ridge and lasso address this by shrinking the estimated coefficients towards zero, which in effect reduces the variance and performs feature selection (Hastie et al., 2009). Ridge regression shrinks coefficients if they are too far from zero, thus enforcing them to be small in a continuous way, which decreases model complexity while keeping all variables in the model. Alternatively, the lasso regression forces some of the coefficient estimates to be exactly equal to zero, discarding predictors with low contributions to the model. Both models use tuning parameter λ to enforce their penalties on the coefficients. Due to how they enforce their penalties, the lasso regression is optimal when only a few predictors influence the outcome while the ridge regression is superior when most predictors impact the outcome. The two approaches tackle correlated predictors in different ways. In ridge regression, the coefficients of correlated predictors are similar while in lasso regression one of the correlated predictors has a larger coefficient, while the rest are zeroed. Elastic net uses a combination of both approaches by finding the ridge coefficients, then applying a lasso-type shrinkage that doubles the amount of shrinkage, and finally rescales the coefficients to improve the prediction performance (**Figure 2-10**). Therefore, the elastic net adds another tuning parameter of α that sets the mixing between ridge ($\alpha = 0$) and lasso ($\alpha = 1$) in addition to λ . Elastic net tackles multicollinearity, and it can generally outperform both lasso and ridge regression under situations with high-dimensional data (Sirimongkolkasem and Drikvandi, 2019).

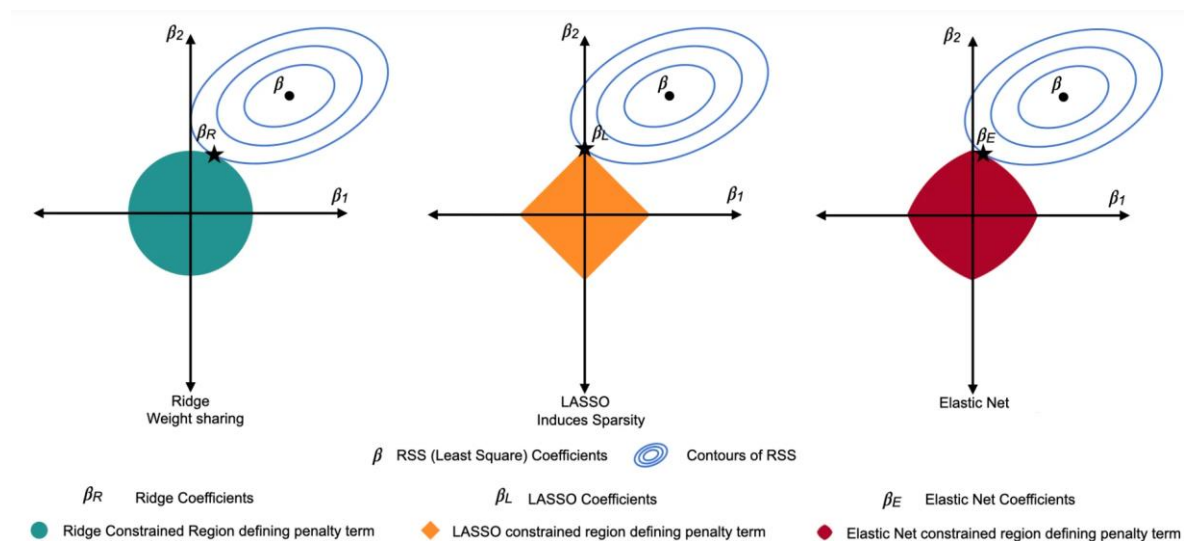


Figure 2-10: Regularisation techniques. The figure displays the penalty-constrained regions and coefficients for ridge, lasso, and elastic net regression in the context of the residual sum of squares (RSS). It visualises how the lasso penalty can force coefficient estimates to be exactly zero compared to the ridge penalty and how the elastic net is a mixture of the two. The figure is adapted from Qazi (2021).

Support vector machines (SVMs) are ML algorithms that perform classification by identifying an optimal decision boundary, known as a hyperplane, which effectively separates two classes in feature space. SVMs use support vectors, observations that lie closest to the hyperplane, while ignoring those farther away (**Figure 2-11A**). SVMs then maximize the margin, the distance between the hyperplane and the support vectors, while minimizing the classification error. However, because SVMs rely on a subset of observations, noise affecting those observations is likely to compromise the hyperplane for the majority of the data. To address this, soft margins are introduced, which relax conditions, allowing some points to fall within the margin. This is controlled by the tuning parameter C , where lower values result in less penalisation of samples within the margins. When a linear boundary proves insufficient, you can expand feature space to a higher dimension, leading to a nonlinear boundary in the original space (**Figure 2-11C**). Kernels can perform this expansion, such as the radial kernel, which computes the similarity or how close observations are to each other. The radial kernel is optimized by the tuning parameter γ , where a large value encompasses a bigger range within which two observations can be considered similar. Support vector regression (SVR) extends SVMs to regression tasks (**Figure 2-11B**), fitting a hyperplane within a specified margin of tolerance ϵ to minimize errors beyond ϵ . Unlike linear regression, which aims to minimize overall error, SVR focuses on minimizing errors caused by support vectors outside the margin. SVMs and SVRs excel with relatively few examples, as they only consider support vectors, a strength compared to other algorithms.

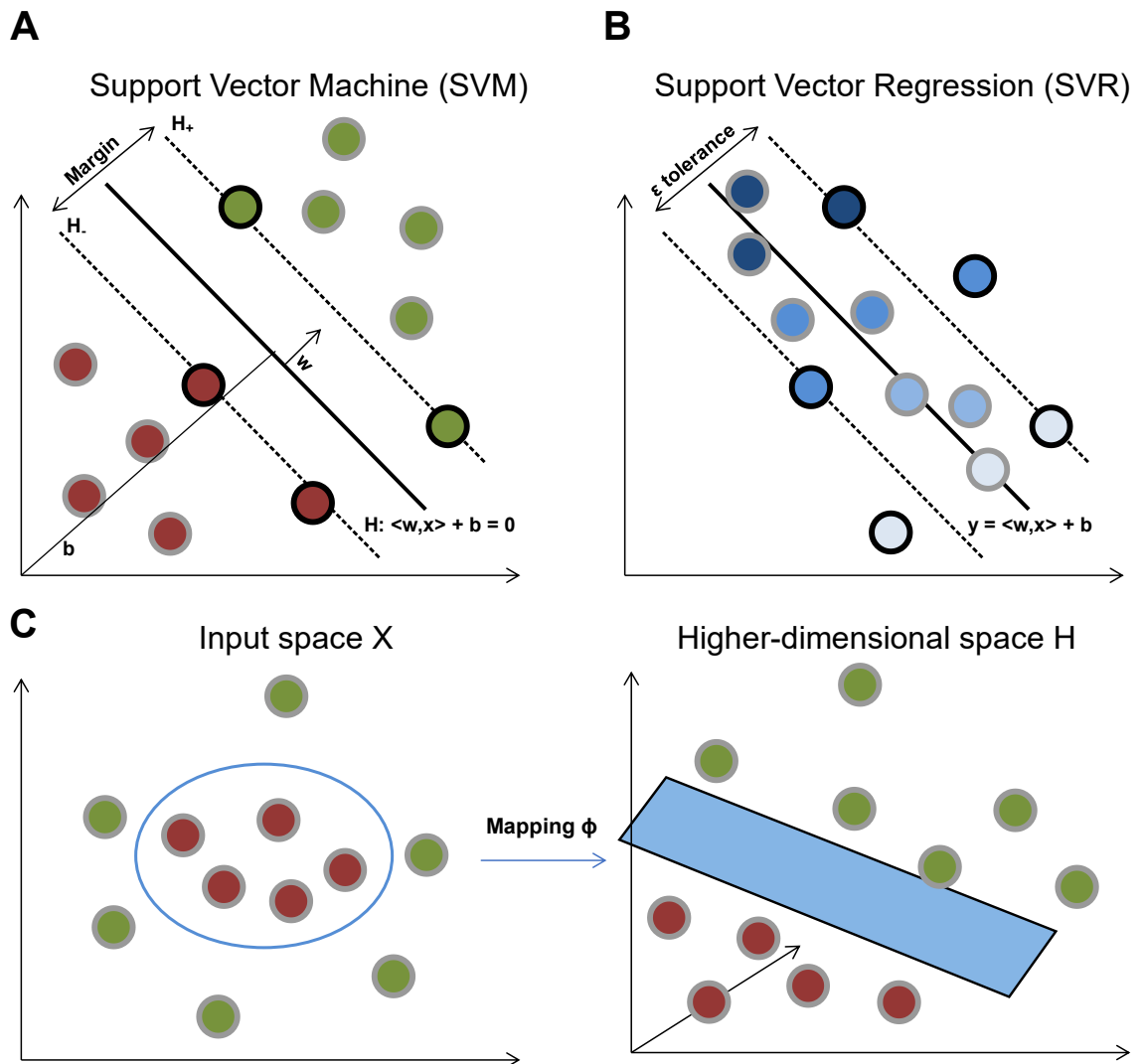


Figure 2-11: Support vector methods. Panel A displays a support vector machine, which identifies a hyperplane (H) and a maximal margin that jointly separate two classes. Comparatively, support vector regression (panel B) minimizes the difference between observed and predicted values, whose gradient goes from dark to light blue as the numerical values decrease. Support vectors in both panels are indicated by black circles but while they are located on the margin in the support vector machine, support vectors may be outside of the margin of tolerance in support vector regression. Panel C displays that if two classes of objects cannot be linearly separated within a feature space X , a non-linear mapping ϕ can project the data points into a higher-dimensional space H in which a linear hyperplane separating the two classes might be found. Adapted from Rodriguez-Perez and Bajorath (2022).

Instead of support vectors, decision tree classifiers identify a decision boundary by iteratively dividing the feature space into subregions. A decision tree, consisting of a root node, hidden nodes, and terminal nodes, specifies an optimal feature and corresponding threshold that divides the feature space into two subregions for each root and hidden node of the tree. It's computationally infeasible to consider every possible partition of feature space, so a top-down approach is used where at each step, the best split is made at the current step, rather than forecasting for a better tree in some future step. The decision tree is recursively constructed until all training samples in a subregion have the same class or the

minimum number of samples per subregion is reached. Regression trees work similarly where the average of the values found in the samples represented by the terminal node is assigned to the sample. However, decision tree models are sensitive to noise and prone to overfitting. Random forests (RFs) address these limitations by building ensembles of classification or regression trees. RFs build decision trees on bootstrapped training samples and each time a split in a tree is considered, a random selection of predictors is chosen as split candidates from the full set of predictors. A fresh selection of predictors is taken at each split, where typically the number of predictors considered at each split is approximately equal to the square root of the total number of predictors (e.g., 4 out of 12). Through this process, RFs decorrelate the trees, reducing the variance as we average across tree predictions, thereby increasing the generalization capabilities. Finally, it is possible to rank each feature with respect to its importance for the task by identifying the features that are most often selected.

In boosting, trees are grown sequentially: each tree is grown using information from previously grown trees. You update your model by adding in shrunken version of the new tree. The shrinking factor (λ) is what you tune for. Given the current model, we fit a decision tree to the residuals from the model. We then add this new decision tree into the fitted function to update the residuals. Each of these trees can be rather small, with just a few terminal nodes, determined by the parameter d in the algorithm. By fitting small trees to the residuals, we slowly improve the model in areas where it does not perform well. The shrinkage parameter λ slows the process down even further, allowing more and differently shaped trees to reduce the residuals. The number of trees B : Unlike random forests, boosting can overfit if B is too large, although this overfitting tends to occur slowly if at all. Very small λ can require using a very large value of B to achieve good performance. The number of splits d in each tree controls the complexity of the boosted ensemble. Often $d = 1$ works well, in which case each tree is a stump, consisting of a single split and resulting in an additive model. More generally d is the interaction depth, and controls the interaction order of the boosted model, since d splits can involve at most d variables. In terms of variable importance for bagged/RF regression trees, we record the total amount that the residuals are decreased due to splits over a given predictor, averaged over all trees.

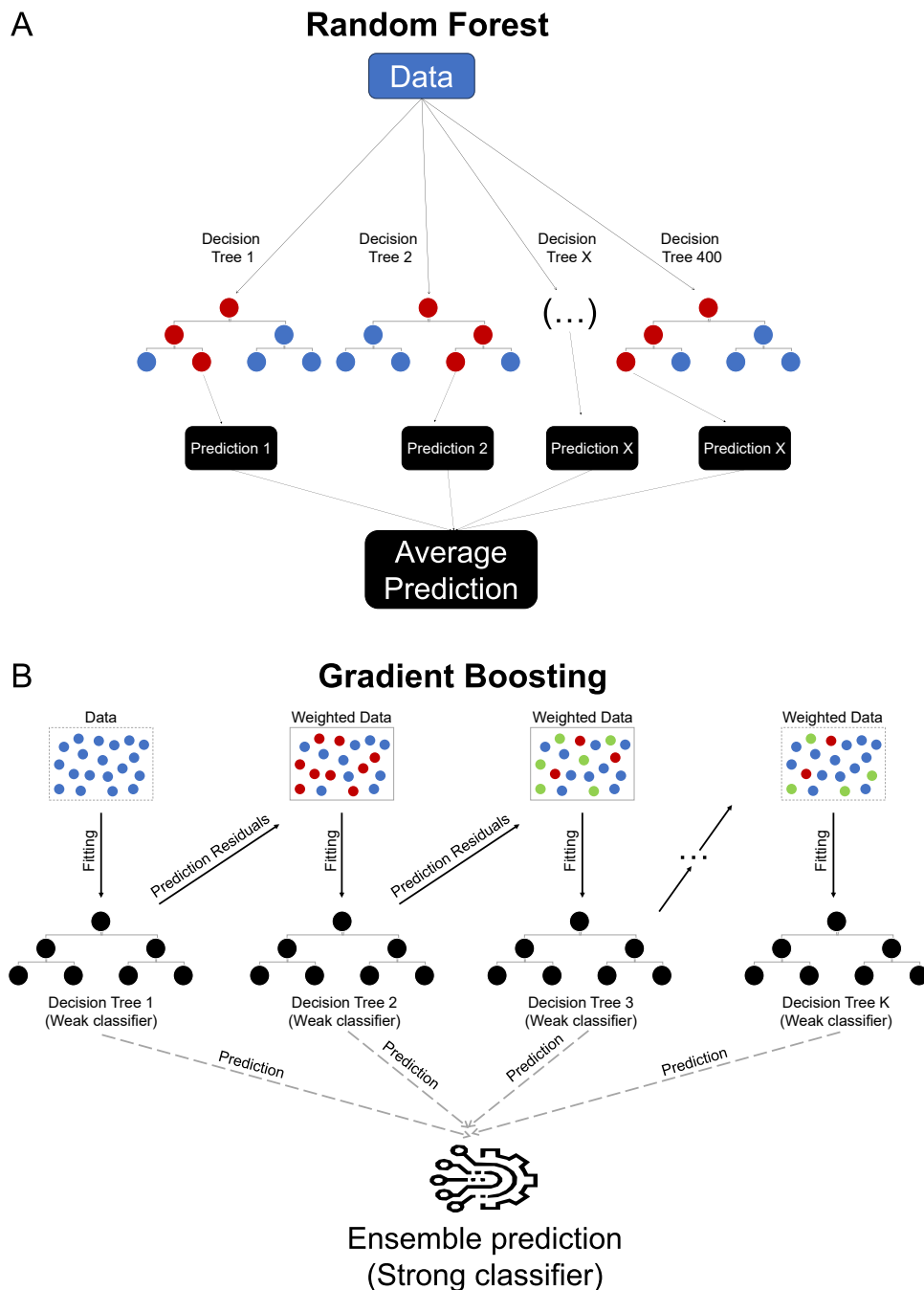


Figure 2-12: Tree-based methods. Panel A shows the random forest approach in which multiple decision trees are constructed on bootstrapped samples of all the data with subsamples of all the available features in each tree. The predictions of each tree are averaged, reducing the variance of the final model and reducing the chance of overfitting the data. Panel B displays gradient boosting in which the constructed trees are not as independent as they are in a random forest model. Instead, it starts by building a decision tree and assigning equal weights to all the data points but then increases the weights for all the points which are not fitted and lowers the weight for those that are easily fitted by using the residuals. A new decision tree is made for these weighted data points. The decision trees are normally only a stump (a tree with only one split), whose predictions are then combined into an ensemble prediction. Adapted from Rajesh et al. (2023).

Bayesian algorithms solve classification or regression problems by predicting the most probable hypothesis, given the input data. They produce non-parametric models, which assume that the data distribution cannot be defined in terms of a finite set of parameters.

Unlike many popular supervised ML algorithms that learn exact values for every parameter in a function, the Bayesian approach infers a probability distribution over all possible values. Given a linear function ($y = wx + \epsilon$), the Bayesian approach would specify a prior distribution, $p(w)$, on the parameter w , and relocate probabilities based on the observed data using Bayes' Rule (2.3) explained by (2.4).

$$p(w|y, X) = \frac{p(y|X, w)p(w)}{p(y|X)} \quad (2.3)$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \quad (2.4)$$

The updated distribution $p(w|y, X)$, called the posterior distribution, thus incorporates information from both the prior distribution and the dataset. Gaussian processes regression is a Bayesian approach that begins with a prior distribution and updates this as data points are observed, producing the posterior distribution over functions (**Figure 2-13**). We select a prior distribution over the function f and condition this distribution on our observations, using the posterior distribution to make predictions. Therefore, compared to linear regression, Gaussian process regression does not match its target function to a specific model (e.g., linear, quadratic, or cubic) but instead tries to infer how all the measured data is correlated. Given a set of training data (x, y) , the Gaussian process regression model uses Bayesian inference to learn the distribution of f that is most likely to have generated the data. This involves computing the posterior distribution of f given the data, defined by (2.5):

$$p(f|x, y) = \frac{p(y|x, f)p(f)}{p(y|x)} \quad (2.5)$$

where $p(y|x, f)$ is the likelihood of the data given the function f , $p(f)$ is the prior distribution of f , and $p(y|x)$ is the marginal likelihood of the data

Once the posterior distribution of f has been learned, the model can make predictions at new test points x^* by computing the posterior predictive distribution (2.6).

$$p(f^*|x^*, y, x) = \int p(f^*|x^*, f)p(f|y, x)df \quad (2.6)$$

The Gaussian process regression can also use kernels where for each input variable, the kernel function performs a transformation and takes the summation of the transformed

values. The distance between each pair of observations is then simultaneously determined by the nonlinear transformed difference across all features (Guan et al., 2019).

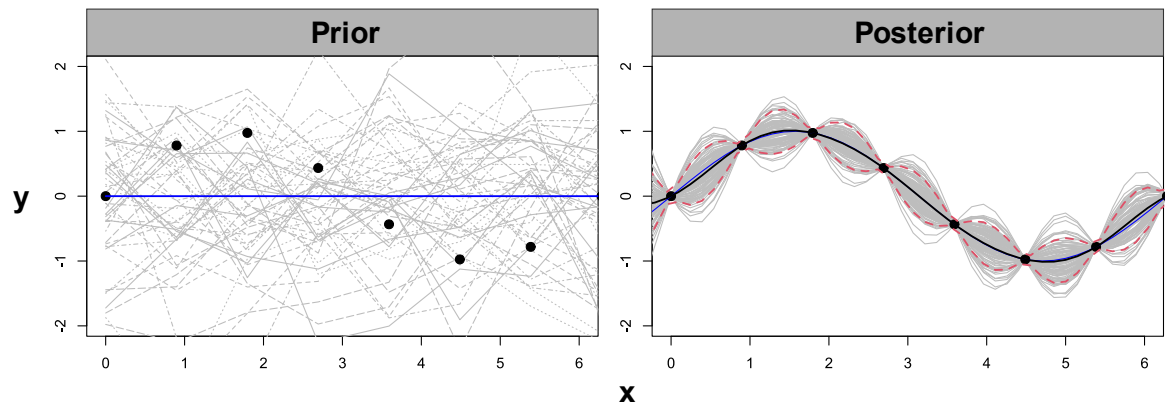


Figure 2-13: Samples from a Gaussian process prior and posterior. The prior shows samples from a Gaussian process prior with a zero-mean, shown as a blue line. The posterior plots each of the random predictive, finite samples as grey curves. Training data points are overlaid, along with true response at the x locations as a blue line. The predictive mean is in black, and 90% quantiles in dashed-red, are added as thicker lines. Adapted from Gramacy (2020).

In summary, ML algorithms predict labels or continuous outcomes and handle high-dimensional data through different methods such as regularisation, support vectors, decision trees, sequential learning (boosting), and posterior distributions. Some of these methods can use kernels to expand feature space to a higher dimension and utilize non-linear relationships between features. All of the mentioned ML algorithms require some form of tuning, where certain parameters need to be optimised and have both pros and cons for their use (**Table 2-4**).

Algorithm Class	Hyperparameters	Advantages (+) and disadvantages (-)
Elastic net	alpha = weighting between ridge (0) and lasso (1) penalties lambda = weighting of the sum of both penalties to the loss function	+ suitable for small datasets + explainable and interpretable + can produce sparse models - does not account for non-linear relationships
Support vector regression	C = penalisation of samples within the margins gamma = range of influence of kernel over observations (radial kernel)	+ robust to overfitting and outliers + can provide both linear and non-linear solutions + low generalization error - results vary based on kernel selection - harder to explain and interpret
Random forest	Number of branches after each split Number of variables sampled at each split	+ handles multicollinearity and non-linear relationships + easier to tune than Gradient Boosting + easy to parallelise - need larger datasets - training complexity can be high
Gradient boosting	Number of trees Shrinking factor Tree depth Minimum number of observations in terminal nodes	+ handles multicollinearity and non-linear relationships + performs well with large number of features - harder to tune due to many hyperparameters - small changes in feature set can greatly affect the models
Gaussian process regression	gamma = range of influence of kernel over observations (radial kernel)	+ suitable for small datasets + can capture non-linearity + explainable and interpretable + can quantify the uncertainty of the prediction - cannot handle high-dimensional data as well as others

Table 2-4: Overview of advantages and disadvantages of selected machine learning algorithms. The table is based from Juarez-Orozco et al. (2018).

The major attributes to consider when selecting algorithms are interpreting how the model generates predictions, handling multicollinearity and a large number of features, being robust to overfitting and outliers, and incorporating both linear and non-linear associations. Each of the algorithms differently balances these attributes while requiring lower training complexity and sample size relative to deep neural networks. Deep learning tends to have too many parameters or a structure that is too complex to be effectively trained on the described IA datasets because will likely result in high variance as there is not enough data to capture the proper patterns with so many parameters. The selected algorithms in this project are better equipped at handling smaller datasets and are the following:

- Elastic net (linear): a linear regression with an imposed penalty to reduce model complexity and decrease the risk of overfitting, making it more interpretable and suitable for small datasets.
- Support vector regression: a model that projects data in multidimensional space to find the best hyperplane to predict outcomes which makes it robust to outliers and produce low generalisation error.
 - Linear kernel
 - Radial (non-linear) kernel
- Gaussian process regression: a model that updates the probability of an outcome given new data, which makes it interpretable and suitable for small datasets
 - Linear kernel
 - Radial (non-linear) kernel
- Random forest (linear and non-linear): a model that combines the outputs of many “if-then-else” tree models that can handle collinearity between features.
- Gradient Boosting regression (linear and non-linear)
 - gbm: similar to random forest but where each subsequent tree model focuses on minimizing the errors made by the previous models.
 - xgbDART: incorporates dropout regularization to gbm, in which it randomly drops a subset of trees during each iteration, which helps to prevent overfitting and can lead to better generalization but requires more tuning due to additional hyperparameters

Beyond selecting ML algorithms that suit the data, methods of evaluating the models need to be robust to overfitting and give reliable estimates of error.

2.4.4 Model evaluation

After statistical models learn the mapping function between the input and output, they predict the training observations used to define the function and estimate the training error with the actual observations. However, we evaluate predictive models by predicting observations that were not used in their training and computing the test error with their actual values. The training error rate often differs and underestimates the test error. The best solution of a large independent test set to externally evaluate model performance is often not available. Instead, we internally evaluate the model where the available data is divided into a set for training the model and a set for validating its performance model, with multiple potential strategies to construct the two sets (**Figure 2-14**).

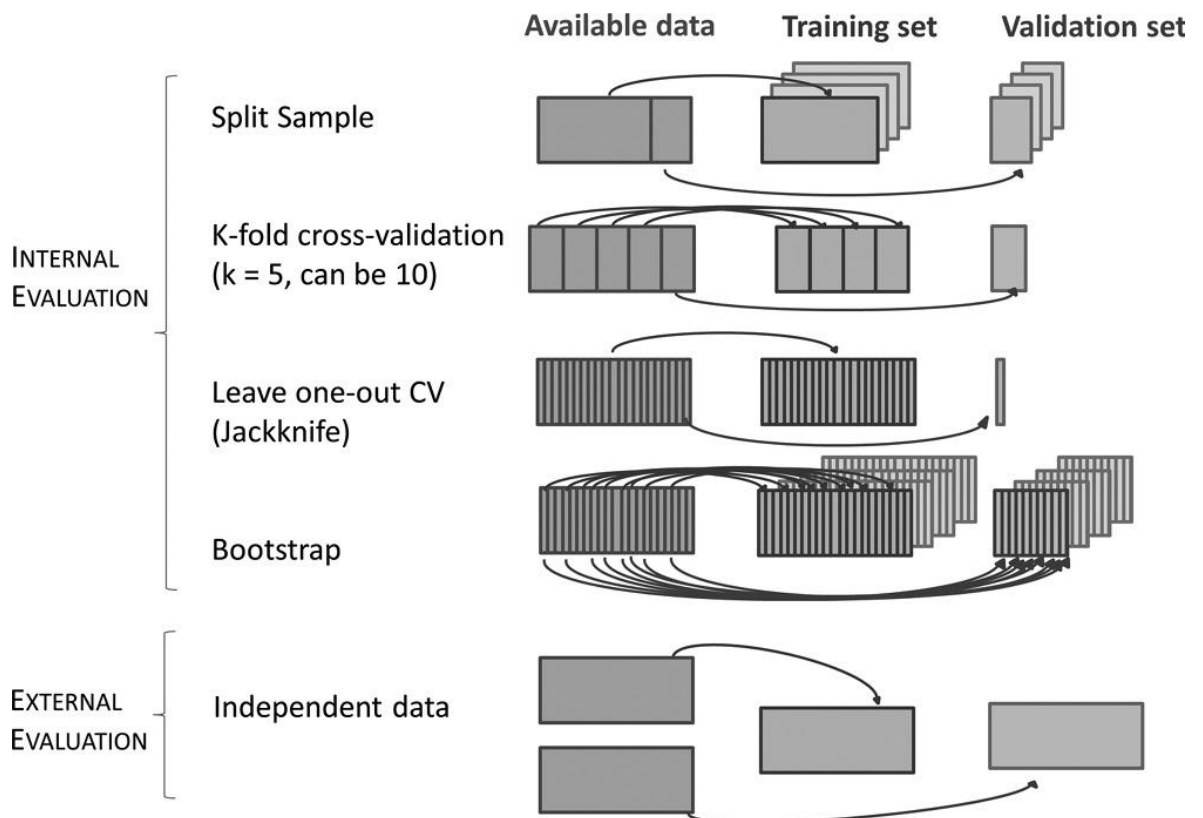


Figure 2-14: Strategies to evaluate models on predictive performance. The figure displays how models are built on training sets and evaluated on validation sets. Internal evaluation is used when a large independent dataset is not available. Split sample is the simplest approach where you randomly split part of the data to fit the model and then use the rest of the data to calculate model performance but repeat the process multiple times and aggregate the results. K-fold cross-validation is when you randomly divide the data into K equal-sized parts, leave out part k, fit the model to the other $K - 1$ parts, and then obtain predictions for the left-out kth part. In leave-one-out cross-validation, also called the jackknife procedure, only one observation is left out. Finally, the bootstrap to evaluate model performance resamples with replacement from the available data to create multiple datasets with the same size. These samples are used to evaluate the model and, at each iteration, use the non-sampled observations as the independent dataset for evaluating the model fitted with the bootstrap sample. Sourced from Guisan et al. (2017).

The drawbacks of internal validation are that the estimate of the test error can be highly variable, depending on precisely which observations are included in the training set and which observations are included in the validation set. This suggests that the validation set error may tend to overestimate the test error for the model fit on the entire dataset. The approach that provides a good compromise between the bias-variance trade-off (**Figure 2-8**) is five- or ten-fold cross-validation (Hastie et al., 2009). A ten-fold cross-validation means that the data is split into ten folds, in each iteration, a single fold is used for testing while the other nine folds are used for training the ML model. In the next iteration, another fold is used for testing while all other folds, including the one used for testing in the first iteration, are used for training. This process is repeated until each fold has been used for testing at least once. The results from each run are summarized into an average metric. In comparison, when only one observation is left out for validation, the data is not shaken up enough and the estimates from each fold are highly correlated thus their average can have high variance.

One concern during prediction modelling is double dipping: when one portion of the data is used twice (Ball et al., 2020). If the entire dataset is used for feature selection and then split into folds, the validation set is not independent as some information of it would be leaked from the selected features. The same applies to optimising the model hyperparameters because when the same cross-validation procedure and dataset are used to both tune and select a model, this optimistically biases the model evaluation and likely results in failing to generalise to new data. Therefore, you need to divide a sample into a set to train the model parameters (e.g., the support vectors), a set to apply feature selection and optimise the hyperparameters of the model (e.g., the type of kernel and its parameters), and then a set to validate the performance. If there is not enough data available, a nested cross-validation can be used (**Figure 2-15**). The procedure again splits the data into K different iterations of training and validation folds but is called the outer loop of cross-validation. The training folds are used for feature selection and then a second split into K folds is made with all the data set for training called the inner loop. This time, instead of validation, the left-out fold is used to select the optimal values for the model hyperparameters. The model is then trained with the optimal parameters and tested using the validation fold. Finally, this procedure can be repeated multiple times (e.g., 50 or 100) by using different splits of the full data as a single run may produce noisy model evaluations due to the specifics of the selected partition. The scores from all repetitions can be averaged to get a final model assessment score that is a more “robust” model evaluation than performing cross-validation only once while producing lower variance compared to other methods such as bootstrapping (Kuhn and Johnson, 2013).

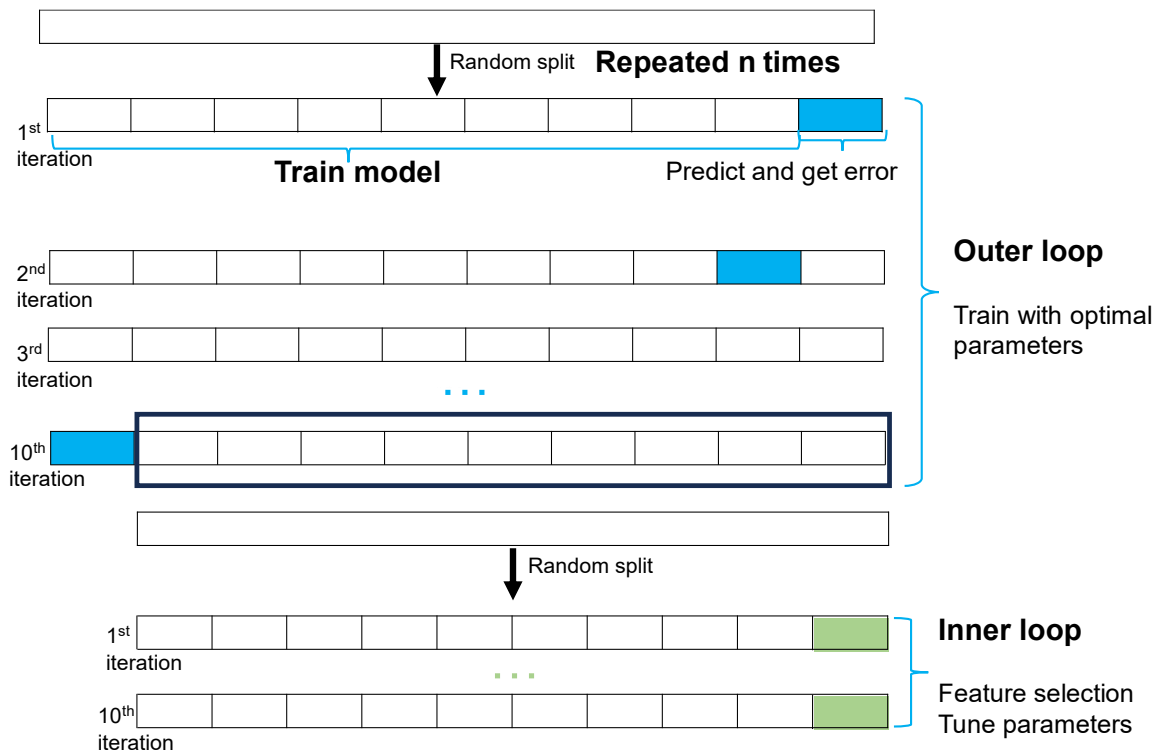


Figure 2-15: Repeated nested ten-fold cross-validation. The figure displays how feature selection, hyperparameter selection, and model evaluation are embedded within inner and outer loops of cross-validation. The data is split into ten folds for the outer loop with nine folds for training and one (shown in blue) for model evaluation. These training folds are combined and split into another ten folds within the inner loop. Nine of those folds are then used for feature selection and model building. The tenth fold (shown in green) is used to select which set of hyperparameters produces the smallest difference between train and test performance. The optimal set of hyperparameters and features are then used to evaluate models in the outer loop built from the initial training folds and evaluated using the validation set (blue). The whole procedure can then be rerun using different random splits of the full data.

2.4.5 Performance metrics

We can quantify the performance of continuous prediction models by associating predicted and actual values and comparing the association to that of a perfect prediction. For example, ML studies have computed the correlation between the predicted and actual values and compared it to a perfect correlation. Such comparative measures are used because they can easily determine model success. However, correlations are not suited for most ML problems as they overlook data scaling and introduce bias, potentially yielding high correlations despite discrepancies between actual and predicted values. As an alternative to correlations, another relative metric is the fraction of explained variance, also referred to as the predicted R^2 . Here, model evaluations can again be compared to a perfect score of 1, whereas a model with random predictions will have a score close to 0. Unlike correlations, a negative R^2 value would denote a model performing worse than if it solely used the mean value of the outcome for predictions. The fraction of explained variance can be incorrectly computed by squaring

the correlation coefficient, which defines R^2 in linear regression. The linear regression definition reflects the error between the predicted values and the fit to the regression line, rather than the error between the predicted and observed values. Although the squared correlation coefficient is appropriate when the model is obtained using the same data, it is unsuitable for out-of-sample testing (Scheinost et al., 2019). Finally, despite being easily interpretable, the predicted R^2 cannot be used to compare performances across different datasets due to variations in the variance of outcome variables.

To compare different datasets, we can calculate error measurements by computing the difference between the predicted and observed values. Common measures include mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE).

$$MSE (observed, predicted) = \frac{1}{n} \sum_{i=1}^n (observed_i - predicted_i)^2 \quad (2.7)$$

$$RMSE (observed, predicted) = \sqrt{MSE (observed, predicted)} \quad (2.8)$$

$$MAE (observed, predicted) = \frac{1}{n} \sum_{i=1}^n |observed_i - predicted_i| \quad (2.9)$$

When comparing different regression models for the same problem, the superior model is likely to demonstrate lower values of each metric. These error measurements should also be similar when training and testing the models because discrepancies would signal potential underfitting or overfitting of the data. The MAE quantifies the error in the units of the original measure and is recommended when the prediction error distribution is highly nonsymmetric due to its robustness to outliers. Instead of providing the absolute error, the MSE squares the error, which places greater weight on extreme differences between the predicted and true outcomes. However, the MSE is difficult to interpret, as the values will not have the same unit as the output. The RMSE is widely used because it remains in output units yet, like MSE, penalises extreme errors while being more likely than MAE to reveal the superior model when the errors follow a Gaussian distribution (Hodson, 2022).

Cross-validation can evaluate predictive power but cannot indicate the statistical significance of the performance of a model or whether it performs better than chance (Eriksson et al., 2003). To address the significance of a performance metric, multiple parallel models can be created by fitting them to randomly reordered outcomes, and then assessing the real performance metric against a distribution of reordered response samples (Malley et

al., 2011). Permutation testing involves fitting a model to a perturbed version of the outcome data, yielding a new estimate of the "permuted" performance value. Repeating this process establishes a null distribution of performance metrics from random data, aiding in evaluating the statistical significance of the parent predictive model's performance values. Statistical tests can then inform of the validity of a model by assessing if the actual performance values (e.g., RMSE, predicted R^2) fall outside the null distribution as the fraction of samples that are greater than or equal to the performance observed when using the correct outcomes (Pereira et al., 2009). However, if the original outcomes are minimally affected by permutation, the permuted outcomes will naturally be highly correlated with the original outcomes. Therefore, evaluating correlations between the original and permuted outcome variables is required to account for occasional high-performance values in permuted outcome data (Eriksson et al., 2003).

2.5 Model interpretation

Clinical practice necessitates predictive models that not only exhibit high performance but are also comprehensible and can guide meaningful actions (Scott, 2021). Moreover, users may require an understanding of how ML models arrive at their predictions, given that these predictions can potentially be spurious or biased without explicit insight from the models themselves. This understanding should encompass, at the very least, the transparency regarding which clinical populations these models may not be applicable to, based on the methods used for model training. Tools that provide such interpretability can offer either global or local explanations. The distinction between the two lies in the scope of explanation: the global type explains the overall functioning of a model, while the local type explains how the model arrived at a specific decision (Molnar, 2020). Most often, both explanations are done post-hoc and involve training a more interpretable model to approximate the behaviour of the task model. These post-hoc models can subsequently uncover crucial features and illuminate interactions among them for the entire model, or how they influenced the prediction for a specific observation.

Measuring the explanatory power of features aids in understanding their influence on predicted outcomes and facilitates comparisons with clinical prior knowledge. These techniques, called attribution methods, typically perturbate the data as is the case in feature importance and accumulated local effects (ALE) plots. Both methods can then provide global explanations of model behaviour. Feature importance calculates the statistical

contribution of features to the performance of a model (Fisher et al., 2019). This is typically done by permuting individual feature values, reapplying the model, and ranking features based on the resulting increase in prediction error (Altmann et al., 2010). A significant decrease in predictive performance from randomly shuffling a variable then implies its higher importance to the predictions of the model. Feature importance has assessed clinical, biological and sociodemographic predictors of cardiovascular disease risk ((Alaa et al., 2019) and mental health outcomes (Bokma et al., 2022), as well as fMRI predictors of antidepressant response (Nguyen et al., 2022). Feature importance through permutation offers clear, easily understandable explanations and is computationally efficient as it does not require repeated model training. However, permutation importance may yield unreliable results when a model is using highly correlated variables while also not being able to show non-linear relationships between features and outcomes (Strobl et al., 2008).

ALE plots address the limitations of permuted feature importance by depicting the average effect of input features on the output (Apley and Zhu, 2020). They specifically illustrate how model predictions fluctuate within a small "window" of the feature around specific values for observations within that window. These plots present the "window" values along the x-axis and display prediction differences on the y-axis (**Figure 2-16**). ALE plots remain unbiased, ensuring their effectiveness even when dealing with correlated features. Being centred at zero, they are easily interpretable, as each point on the ALE curve signifies the deviation from the mean prediction. They also excel at depicting non-linear relationships between variables and outcomes, as demonstrated in a study discerning clinical variables non-linearly classifying individuals with mild cognitive impairment (Basta et al., 2023). Such plots have also revealed the relationship between cognitive ability and fMRI activity from multiple brain regions, acting as predictors in ML models (Pat et al., 2023). However, it's important to note that ALE effects might deviate from the coefficients outlined in a linear regression model in cases where features interact and are correlated, as ALE defines first-order effects differently compared to a linear formula (Molnar, 2020). Therefore, constructing a model could provide better insight into how predictions are made for individual cases.

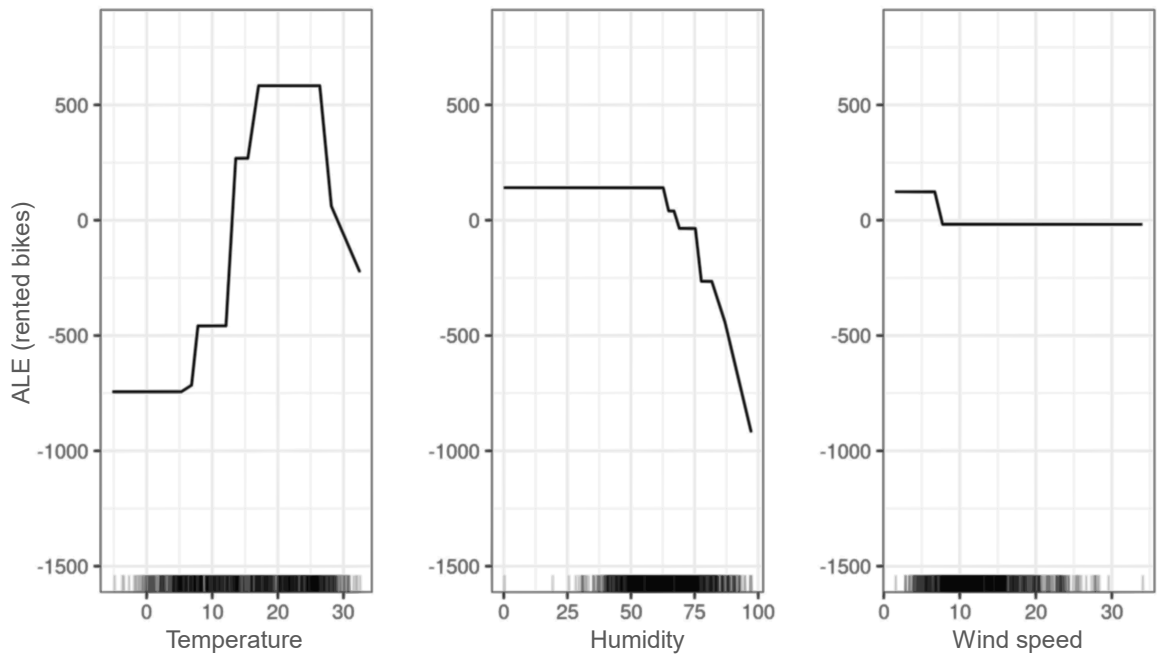


Figure 2-16: Example of ALE plots. The figure displays ALE (accumulated local effects) plots temperature, humidity, and wind speed that act as predictors in a model that predicts bike rentals. The temperature greatly affects the prediction. It has a non-linear effect because while the average prediction rises with increasing temperature, it falls again above 25 degrees. In comparison, humidity negatively affects the prediction. While initially not having an effect, when it goes above 60%, increasing humidity relates to lower rented bike difference in the mean prediction. Unlike temperature and humidity, wind speed does not meaningfully change the predictions at any point. Adopted from Molnar (2020).

Compared to feature importance and ALE plots, Local Interpretable Model-agnostic Explanations (LIME) develop simpler, more interpretable models that mimic the behaviour of the model for predicting individual outcomes (Ribeiro et al., 2016). Specifically, the LIME model is trained to approximate the predictions of the underlying model by altering the input for each observation and tracking how the predictions change. LIME generates a new dataset with perturbed samples from the relative distribution of the features along with the corresponding predictions of the task model. On this new dataset, LIME then trains an interpretable model, weighted by the proximity of the sampled instances to the instance of interest. The interpretable model can be any interpretable model type such as a regularised regression or a decision tree. LIME produces a list of explanations, indicating the contribution of each feature to the prediction of an observation. These results are visualized as deconvolution graphs, highlighting the most influential features in predicting a case. This method offers local interpretability, and it also alludes to which feature alterations will have the most influence on the prediction. For example, Ghafouri-Fard et al. (2019) used LIME to identify the most protective and risky genotypes for autism spectrum disorders. What's important to note is that while LIME models should be a precise approximation of ML model predictions individually, they don't necessarily need to be a precise global approximation.

2.6 Summary

Overall, the subsequent chapters will follow a method structure of preprocessing the neuroimaging data from the specific dataset (**Chapter 2.1**) and then applying GLMs, unsupervised or supervised ML to address the specific research aims and hypotheses. Depending on the aims and hypotheses, neuroimaging data will be pre-processed in either MNI or individual space with appropriate for each modality preprocessing steps (**Chapter 2.2**). Methods, specific to the research aim such as the brain atlas used to define the ROIs will be more extensively described in the dedicated chapter (**Figure 2-17**).

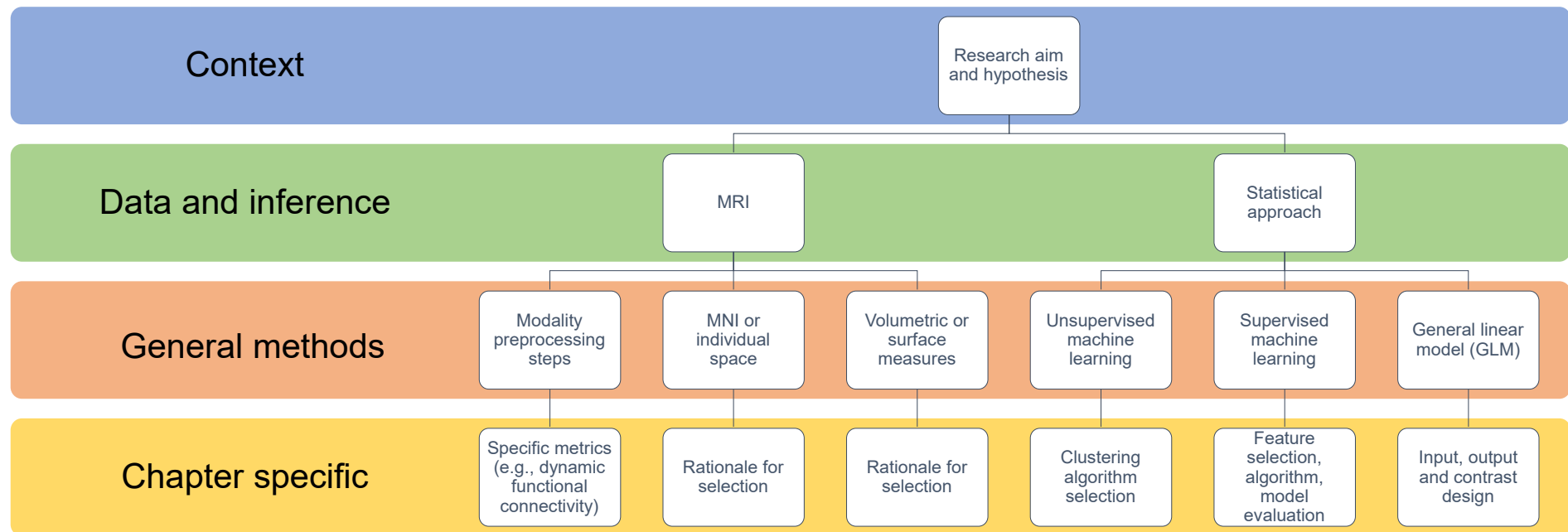


Figure 2-17: Summary of general methods and chapter specific additions.

Chapter 3 Identifying subgroups of fatigue

3.1 Introduction

This chapter will explore subgroups identified through neuroimaging data and assess how they differ phenotypically. The rationale behind this aim stems from the significant heterogeneity within people with RA that impedes identifying subgroups of patients who are more likely to benefit from one therapy over another. These patients differentiate in terms of disease pathogenesis, development, and symptom profile (Townsend, 2014, Verheul et al., 2015). To address this variability, studies have attempted to partition patients based on clinical or biological information using data-driven approaches (clustering) rather than intuition or experience alone. Lee et al. (2014) clustered patients on pain, fatigue, inflammation and psychosocial factors and found three distinct subgroups. However, clustering on symptomatology may not reflect any underlying biological differences and cannot reliably predict treatment response on an individual level. Conversely, biological data more often than not is dominated by uninformative signals and may reflect groups of people with similar brain size or gender rather than elements of pathophysiology (Snoek et al., 2019). Therefore, studies may require both clinical and biological metrics to infer stable and relevant subtypes in RA.

This lack of clinically relevant biological subtypes is also evident in psychiatry and was approached by Drysdale et al. (2017) for depression. In patients with major depressive disorder, the study incorporated fMRI and clinical data to identify subgroups of these patients. The advantages of such subtypes can be two-fold: the authors (1) predicted response to a specific transcranial magnetic stimulation (TMS) treatment while a review on biomarkers in psychiatry suggested using the described heterogeneity to (2) inform larger prediction studies of what to expect when validating results on external data (Schnack and Kahn, 2016). In RA, depression levels relate to fatigue as they share correlates of higher functional interactions between specific brain regions (Basu et al., 2019). Therefore, this approach can potentially describe the present RA datasets and advise on how to construct and validate prediction models, as described in **Chapter 7**.

To validate this approach in RA, a similar methodology to Drysdale et al. (2017) was conducted on two RA datasets but modified based on a replication study by Dinga et al. (2019). The aim was to identify neurobiological subtypes of fatigue with distinct signatures

of brain connectivity patterns using Drysdale and colleagues' pipeline and validate this approach in an independent RA dataset. I hypothesise that:

- There is a significant brain-behaviour relationship between fatigue subdimensions and brain connectivity, illustrated by a significant canonical correlation between fMRI connectivity and subdimensions of fatigue.
- RA patients can be stratified to groups with distinct brain connectivity. This will be demonstrated by a cluster analysis on fatigue-related brain connectivity. The statistical distribution of the brain connectivity will be different for each subgroup.
- The subgroups will be statistically different in terms of disease activity, inflammation, pain, sleep, and depression as these variables have been shown to stratify RA patients with fatigue and relate to centralised symptoms of sickness behaviours.
- These findings will be replicated in a different RA cohort.

3.2 Methods

3.2.1 Patient recruitment and MRI preprocessing

Patient recruitment and both clinical and MRI data descriptions for the RA study have already been described in **Chapter 2.1**. Pre-processing procedures are described in **Chapter 2.2** and implemented by the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012). This analysis used the default MNI pipeline to be compatible with Drysdale et al. (2017): discarding the first 4 volumes; realignment; functional and structural centralisation to (0,0,0) coordinates; slice-timing correction in ascending order; ART-based motion outlier detection using a z threshold (global signal) of 9 and movement threshold of 2 mm; functional and structural segmentation, MNI normalization and 8-mm smoothing. Subjects' data were excluded if they had over 20% of their total volumes (222) beyond the 2 mm movement threshold (invalid volumes ≥ 44 in either session). To remove any spurious correlations due to physiological or MRI sources of noise a denoising step on the fMRI data was applied and included the regression of 5 dimensions for white matter and CSF and their first-order derivatives (10 dimensions) using the CompCor algorithm (Behzadi et al., 2007); motion parameters and their first derivatives (12 dimensions); scrubbing; and band-pass filtering (0.008 – 0.09 Hz).

3.2.2 Dynamic fMRI connectivity

Both fMRI and clustering analyses were based on a previous study that used FC and clinical symptom scores to find neurophysiological subtypes of depression (Drysdale et al., 2017). The methods (**Figure 3-1**) also included recommendations from a study which attempted to replicate Drysdale and colleagues' results in a similar cohort (Dinga et al., 2019). The functional data that was further analysed included the “on periods” of the PASAT task. CONN's default atlas defined ROIs that included the FSL Harvard-Oxford Atlas and cerebellar areas from the AAL atlas, consisting of 132 ROIs in total.

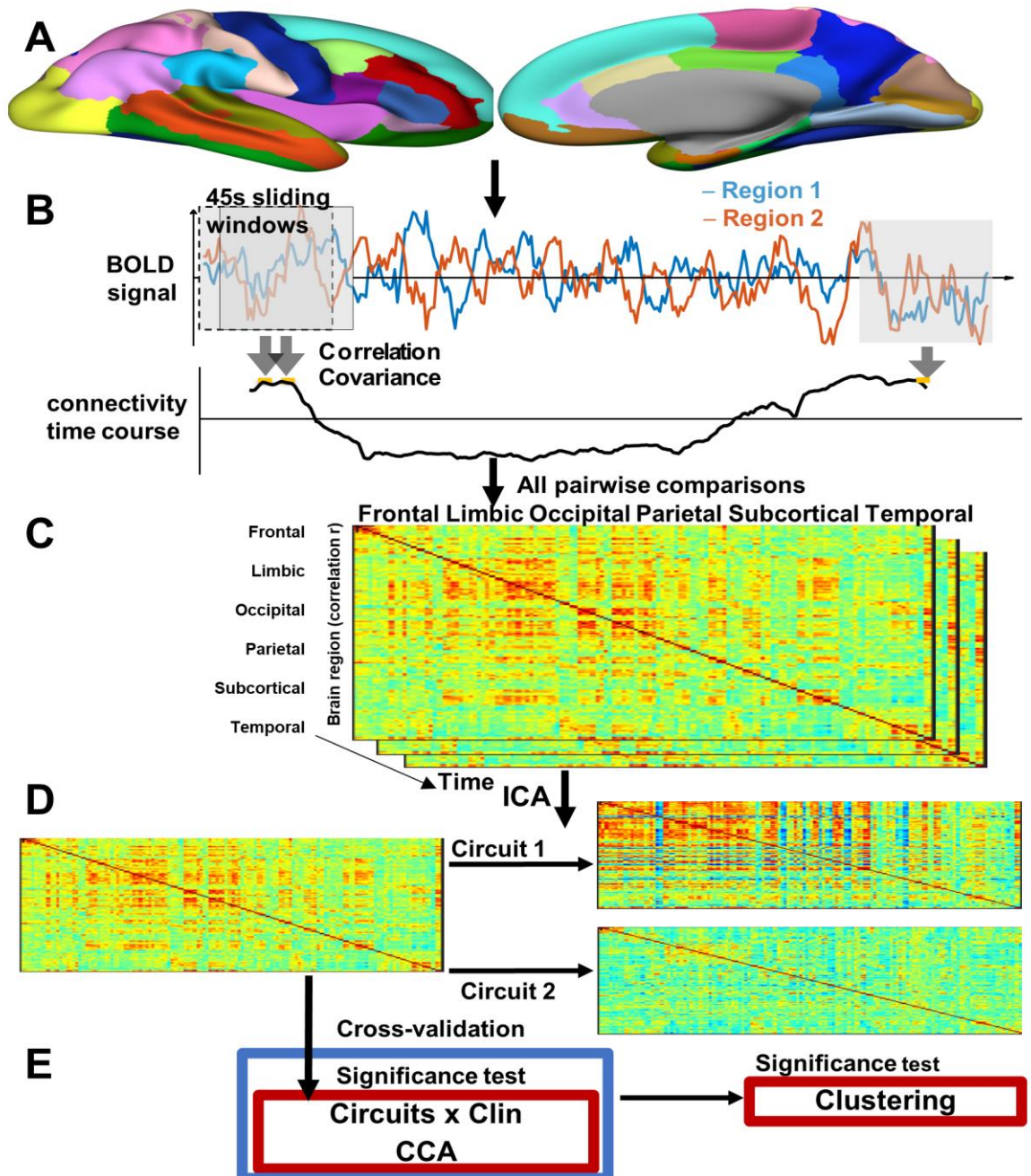


Figure 3-1: Analysis pipeline for identifying patient subgroups. **A:** the BOLD signal change (activity) is estimated for 132 brain regions-of-interest (ROIs) based on the CONN toolbox default atlas which covers cortical, subcortical, and cerebellar areas; **B:** for each ROI pair (e.g. region 1 and 2 on the graph), a correlation value over their activity is calculated every 45 seconds (sliding window) of the Paced Auditory Serial Addition Test (PASAT) task; **C:** for each window (consisting of 41 volumes), a connectivity matrix is created from correlation values for all pairwise comparisons. The outputs of this dynamic connectivity are 12 matrices for the whole PASAT task for each patient; **D:** the CONN dyn-ICA analysis performs independent component analysis of the connectivity matrices for all patients, returning a predetermined number of brain circuits with similar fluctuations in correlation strength across time; **E:** the variability of 4 circuits are used combined with sub-scores of fatigue using a canonical correlation to create 2 summary connectivity values for each patient relevant to fatigue. These values are then used by hierarchical clustering algorithms as the basis of differentiating subgroups (clusters) with these patients. The significance of the canonical correlation and clusters are tested separately, while a cross-validation procedure establishes the stability of the correlation with different subsamples of the patients. Abbreviations: BOLD, blood-oxygen-level-dependent; ICA, independent component analysis; CCA, canonical correlation analysis.

Dynamic FC was conducted to create brain measures, previously unexplored in RA patients, which could further inform about the fatigue state. While studies using stationary FC have aided in identifying resting-state fMRI networks, they often fall short of capturing their intricate dynamic changes (Hutchison et al., 2013). Unlike stationary FC, dynamic FC reflects the dynamic changes over time within brain states (Zhang et al., 2016) and is typically estimated by the variance of FC between networks/regions (**Chapter 1.5.2**). This added variability has been suggested to offer greater insight into the properties of brain networks (Allen et al., 2014). Dynamic FC has proven beneficial in explaining individual behavioural and cognitive traits, as well as classifying neurological diseases, outperforming stationary FC (Peng et al., 2022). This advantage may stem from dynamic FC capturing latent functional boundaries, particularly in regions with high flexibility and adaptability.

The disadvantages of dynamic FC include the complexity of estimating such measures and the lack of a standard approach compared to static FC (Prete et al., 2017). Additionally, it requires larger volumes of data, although reliable results have been demonstrated with a duration of 10 minutes (Zalesky et al., 2014, Betzel et al., 2016). In contrast, static FC is limited in highlighting only areas with stable connectivity throughout the acquisition time, while dynamic FC can capture the inherent dynamic nature of FC alterations, describing changing cognitive states (Prete et al., 2017). Therefore, dynamic FC is particularly suited for conditions where excessive variability occurs at different times for the same individual, which may potentially be applicable to RA patients and their fluctuating levels of fatigue during a fatigue-inducing task.

Dynamic FC also allowed to reduce the dimensions of the data in an unbiased way, as suggested by Dinga et al. (2019). The procedure groups ROI pairs in circuits that have similar changes in correlated activity across time (Anteraper et al., 2020). The window length for estimating each connectivity time-point was set to 45s (Leonardi and Van De Ville, 2015). Dynamic connectivity was estimated for the RA study (baseline and follow-up) and LIFT (baseline), in the form of mean frequency and variability of the circuits for each subject. Whether frequency or variability of dynamic circuits was used for clustering in the trial data depended on which metric resulted in the most similar clusters between baseline and follow-up of the RA study.

3.2.3 Canonical correlation and clustering analyses

Clinically relevant variables were extracted from connectivity estimates by a canonical correlation analysis (CCA, **Figure 3-2**) with fatigue scores, and then used for partitioning patients in R Statistical Software (R Core Team, 2021) based on Dinga et al. (2019). The CCA used 4 sub-scores from the Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire (BRAFMQ) which records subdimensions of physical, impact on living, cognitive, and emotional fatigue dimensions (**Appendix A**). To perform reliable CCAs, 4 connectivity circuits were correlated with the 4 BRAFMQ sub-scores (total of 8 variables) for the 82-subject LIFT trial, meeting a standard of 10 subjects per variable (Leach and Henson, 2014). After age, gender, and scanning site (LIFT data) were regressed, CCA resulted in two linear combinations of connectivity estimates (connectivity variates) that maximally correlated with the fatigue variables. Ward's D linkage method (hierarchical clustering) with Euclidean distance partitioned patients into clusters based on the two connectivity variates (**Figure 3-2D**). To identify the best number of clusters, indexes are used to compare clustering solutions on certain characteristics such as maximizing the ratio of between-cluster to within-cluster variance. After going over all available index criteria (27), the majority of indexes selected the number of clusters to find. The identified clusters at baseline were then described in terms of the other clinical variables at baseline.

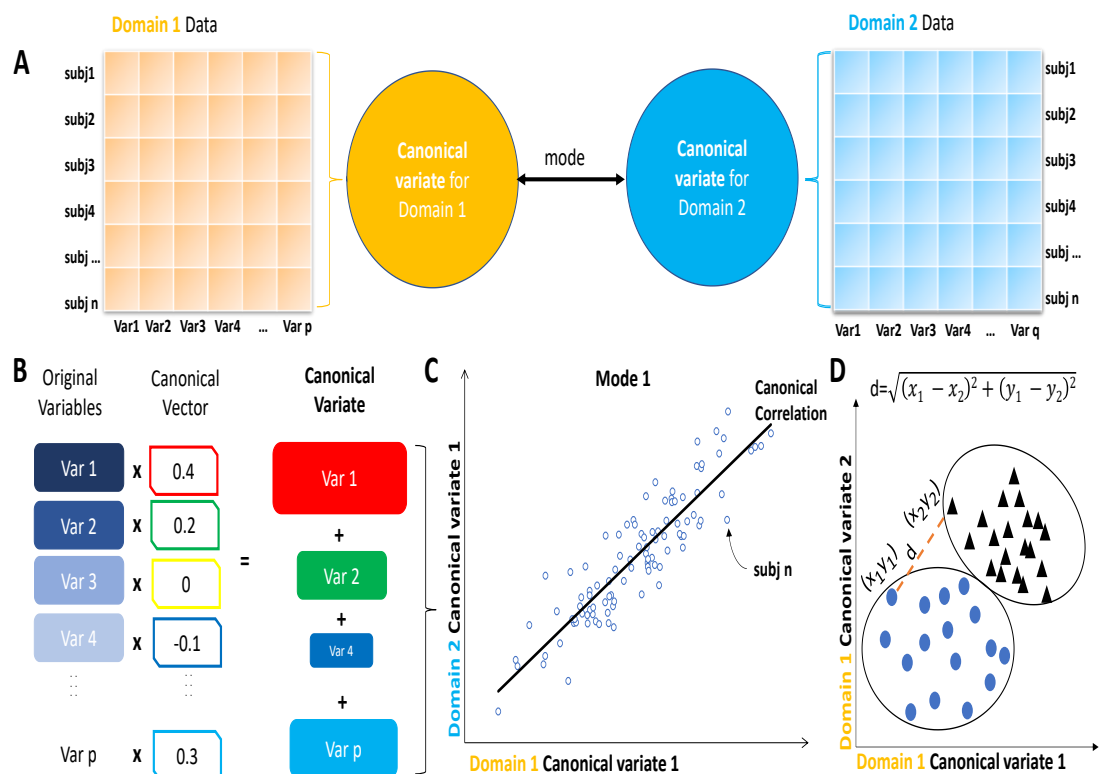


Figure 3-2: Summary illustration of a canonical correlation analysis (CCA) and clustering. A: If several domains of data are acquired from the same subjects, with p and q number of variables

respectively, they can be co-decomposed by CCA. In this study, the domains reflect dynamic connectivity and BRAF fatigue sub-scores with 4 variables each ($p=q$). CCA re-expresses the variables as latent linear combinations called canonical variates. Each variate from domain 1 is maximally correlated with the corresponding variate from domain 2 but completely uncorrelated to the next variate. Each pair of variates is called a mode with the number of modes corresponding to the number of variables in the lowest-variable domain; **B**: In each domain, the canonical variate is composed of the weighted sum of variables by a canonical vector; **C**: In a CCA with two domains, each subject can be described by two canonical variates per mode, and their canonical correlation coefficient (r value). Each consecutive mode has a lower in strength r value. In this study context, there are 4 modes with 4 canonical variates for both the dynamic connectivity and fatigue domains. **D**: Hierarchical clustering was based on the first two canonical variates from dynamic connectivity depicted in the illustrative plot using Euclidean distance(d) to identify patient subgroups (blue circles and black triangles). Adapted from Wang et al. (2018).

3.2.4 Statistical analyses

Mann-Whitney U test (non-normal distribution) estimated whether the RA study patients differed from LIFT patients at baseline for inflammation (ESR) and fatigue (Chalder fatigue). Each CCA outputted a correlation coefficient and a Wilk's lambda value to test for significance. To evaluate the stability of CCA, a 10-fold cross-validation (Varoquaux et al., 2017) was implemented and the average r value was recorded. Based on Dinga et al. (2019), significance testing informed whether the values of the indexes which picked the cluster number are significantly different than what would have been expected under the null hypothesis of data with no underlying clusters. To create the data, a covariance matrix between the two canonical variates was estimated and used to generate random samples of the number of observations in the dataset from a bivariate Gaussian distribution defined by this covariance matrix. The same clustering procedure was performed on each random sample, thus obtaining an empirical null distribution of index criteria values. The optimal value for some of these indexes is their maximal value and for some, it is their minimal value. Therefore, the p -value was defined as the proportion of the calculated indices in the null distribution smaller/bigger (relative to the index) than what is observed in the real data. The corrected Rand index (Hubert and Arabie, 1985, Steinley, 2004) was used to estimate the similarity between baseline and follow-up clusters, through the `fpc` R package (Hennig, 2019). The index assesses the similarity between the two cluster solutions and is a value from -1 (no agreement) to 1 (perfect agreement). In terms of clustering based on structural and functional MRI, a value of over 0.8 has been shown to be replicated in a reliability study on a multi-centre dataset (Hawco et al., 2018). This was the criterion for selecting mean frequency or variability when replicating the CCA and clustering steps in LIFT. Finally, to investigate differences between clusters on disease activity, ESR, sleep, pain, fatigue or depression, a series of Kruskal Wallis tests were implemented.

3.3 Results

3.3.1 Patient characteristics

From the RA study 1, four subjects were excluded due to excess motion while further analyses were carried out in the remaining 50 patients [38 females, mean (SD) age 55.13 (11.36)]. No patients exhibited excess motion from LIFT, in which 82 subjects completed clinical assessments and neuroimaging scans at baseline [60 females, age 59.6 (11.57)]. Although comparable on some clinical traits (**Table 3-1**), the patients of the RA study had more inflammation, but less fatigue severity compared to LIFT patients (**Figure 3-3**). The excluded four subjects were not significantly different for age, gender, inflammation or fatigue severity (**Table 3-2**).

Factor	Baseline RA study	6-month RA study	Baseline LIFT
RA disease activity ^a	3.58 (2.18)	3.16 (2.07)	4.28 (1.65)
ESR (mm/h)	16.5 (14)	13 (16)	10.5 (11.25)
Fatigue ^b	35.5 (16.25)	31 (26.75)	40.22 (16)
Sleep disturbance ^d	16 (7.75)	16 (8.75)	12 (9)
Pain severity ^e	4 (3)	3 (2.75)	6 (3.75)
Depression ^f	7.5 (5)	6 (6)	6 (6)

Table 3-1: Clinical characteristics of patients in the RA study and LIFT with PASAT fMRI as Median (IQR). ^aDAS-28; ^bBristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFF MDQ) total score; ^dJenkin's sleep scale; ^eCurrent pain NRS; ^fHospital Anxiety and Depression Scale; NRS: numerical rating scale.

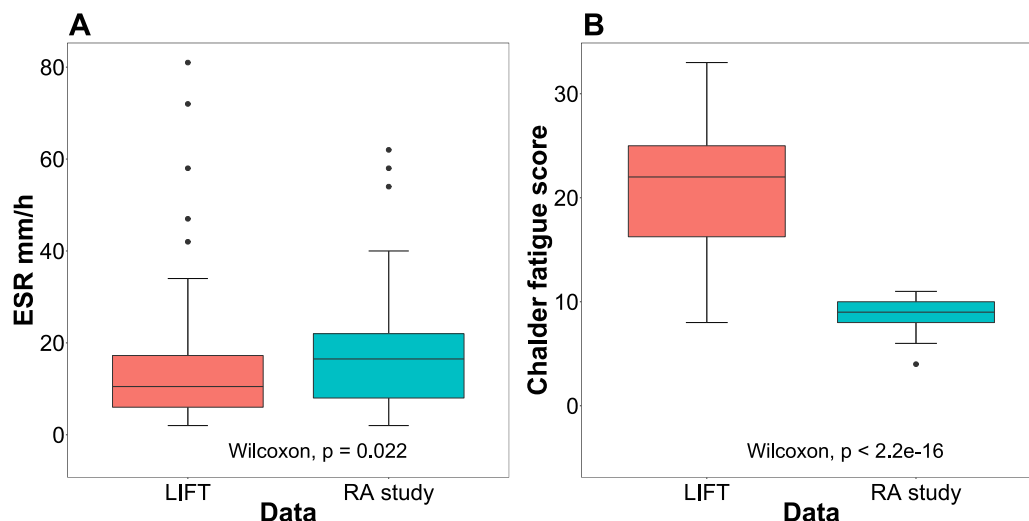


Figure 3-3: Patients in the RA study are different from patients in the RA trial (LIFT). The boxplots in both panels represent the following: the centre line (the median value), the lower bound of the box (25th percentile), the upper bound of the box (75th percentile), and the whiskers (minimum and maximum values, or 1.5 times the interquartile range in the presence of outliers, indicated by dots). In **panel A**, a Mann-Whitney U test indicates the patients from the RA study had significantly more inflammation ($U=1520$, $p=0.022$, $rB = -0.24$). In **panel B**, the same test indicated more fatigue severity in LIFT patients ($U=4057$, $p<0.0001$, $rB = 0.98$), as measured by the Chalder Fatigue Scale. The effects are small and large, respectively.

Factor	Included (Baseline)	Excluded (Baseline)
Sample size	50	4
Gender (female)	76%	75%
Age	56 (14)	47.5 (12.55)
RA disease activity	3.58 (2.18)	5.08 (2.84)
ESR (mm/h)	16.5 (14)	41 (14.25)
Chalder Fatigue	9 (2)	11 (0)

Table 3-2: Differences between the included and excluded patients from RA study 1 as Median (IQR). The four patients excluded due to too much movement during MRI scanning did not significantly differ (Mann-Whitney U test) from the fifty included patients on age, gender, DAS28 disease activity, Chalder Fatigue score, and inflammation in the form of erythrocyte sedimentation rate.

3.3.2 Dynamic connectivity is associated with dimensions of fatigue in the RA study

The 4 dimensions of BRAF MDQ fatigue (Physical, Living, Cognitive, Emotional) were all combined with the 4 dynamic connectivity estimates (mean frequency or variability) using CCA. The CCA with the mean frequency of the 4 dynamic circuits led to a significant correlation at baseline (**Figure 3-4A**) and trend-level correlation for follow-up (**Figure 3-4B**), with cross-validation means of 0.38 and 0.28, respectively.

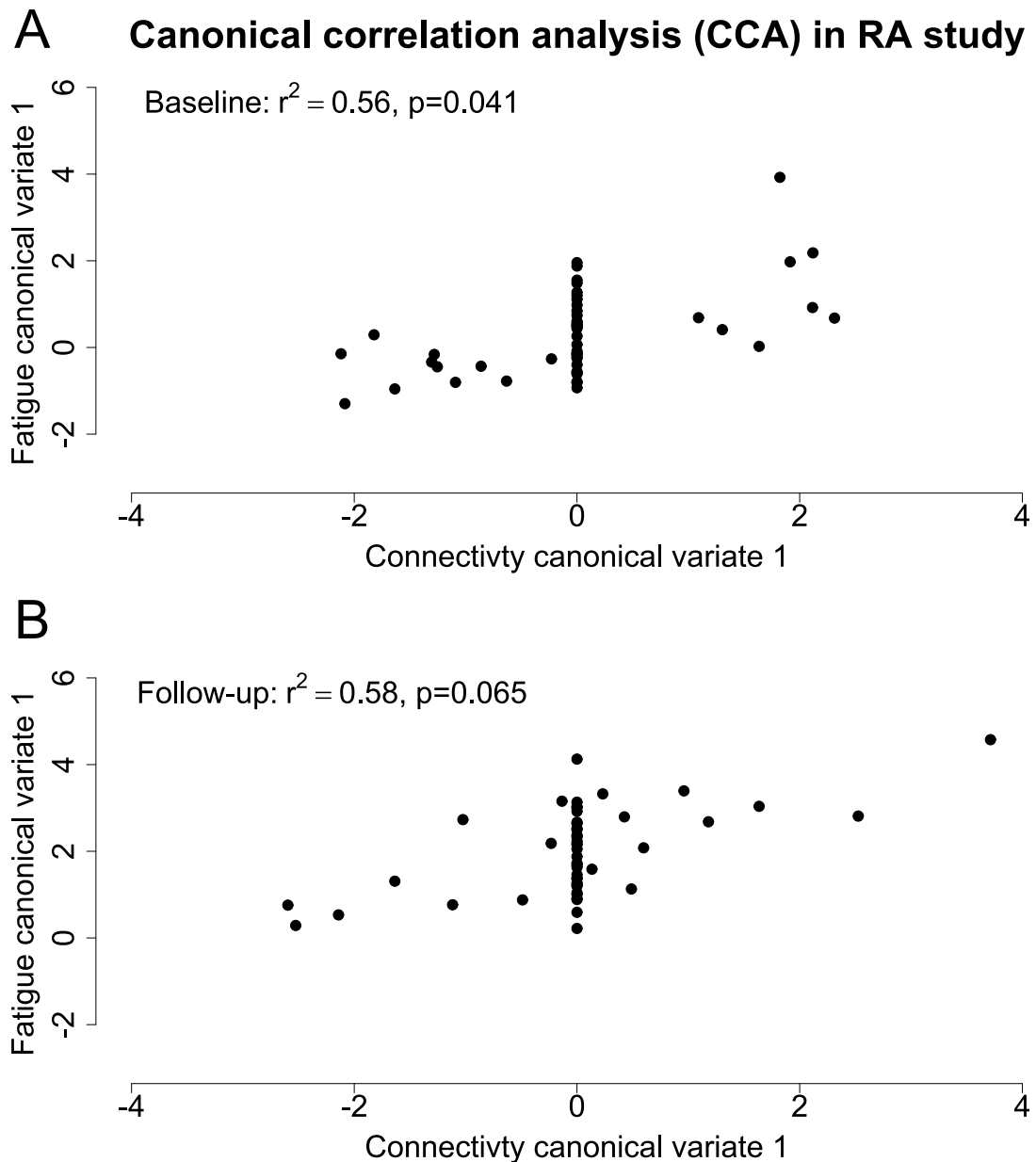


Figure 3-4: Significant canonical correlations at baseline and trend level at follow-up. A: Linear combinations (canonical variates) of dynamic connectivity features (mean frequency) that are maximally correlated with a linear combination fatigue sub-scores (BRAAF MDQ) resulted in a significant correlation ($r^2=0.56$, $p=0.041$) at baseline in the RA study. **B:** The canonical correlation for follow-up data showed a similar trend but was not significant ($r^2=0.58$, $p=0.065$). Both CCAs were conducted after age and gender were regressed from the dynamic connectivity features.

3.3.3 Reliable clusters from canonical connectivity scores

The hierarchical clustering yielded four clusters by majority rule for both baseline (13 indices, 31:8:6:5 members ratio) and follow-up (16 indices, 31:8:6:5 members ratio) data. For baseline, 8 out of 13 indices were significant ($p < 0.001$), while for follow-up: 8 out of 16 were significant ($p < 0.001$) (**Table 3-3**). The null distributions for the ptbiserial index at baseline and follow-up are illustrated in **Figure 3-5**. The members of clusters were highly similar (corrected Rand Index=0.88) with 41 out of the 50 patients in the RA study staying in the same clusters from baseline to follow-up (**Figure 3-6**).

Index	Baseline (index value; p-value)	Follow-up (index value; p-value)
ch	27.87; 0.8825	29.28; 0.799
hartigan	46.55; 0.0005***	49.07; 0.0005***
cindex	0.04; 0.1605	0.04; 0.2435
db	0.88; 0.0005***	0.91; 0.0005***
silhouette	0.68; 0.935	0.68; 0.743
ratkowsky	0.39; 0.0005***	0.4; 0.0005***
ptbiserial	0.8; 0.0005***	0.79; 0.0005***
mcclain	0.11; 0.0005***	0.11; 0.0005***
gamma	0.93; 0.0005***	0.93; 0.0005***
gplus	9.85; 0.0005***	9.92; 0.0005***
tau	279.38; 0.65	279.23; 0.0685
sdindex	1.96; 0.779	1.78; 0.3355
sdbw	0.94; 0.8825	0.64; 0.799
ccc		4.25; 0.806
dunn		0.17; 0.3955
friedman		5.75; 0.0005***

Table 3-3: Cluster index criteria values and p-values after comparisons with simulated data. The index column lists all the cluster index criteria which choose the 4-cluster solution for the baseline or the follow-up data. The columns for the baseline and follow-up data annotate the index value from the real data and the p-value based on the null distributions created from repeating the clustering procedure on simulated data with the same mean and covariance. For baseline, 8/13 indices were significant while for follow-up: 8/16. Based on the NbClust package in R statistical software (Charrad et al., 2014); *** annotates values below $p < 0.001$.

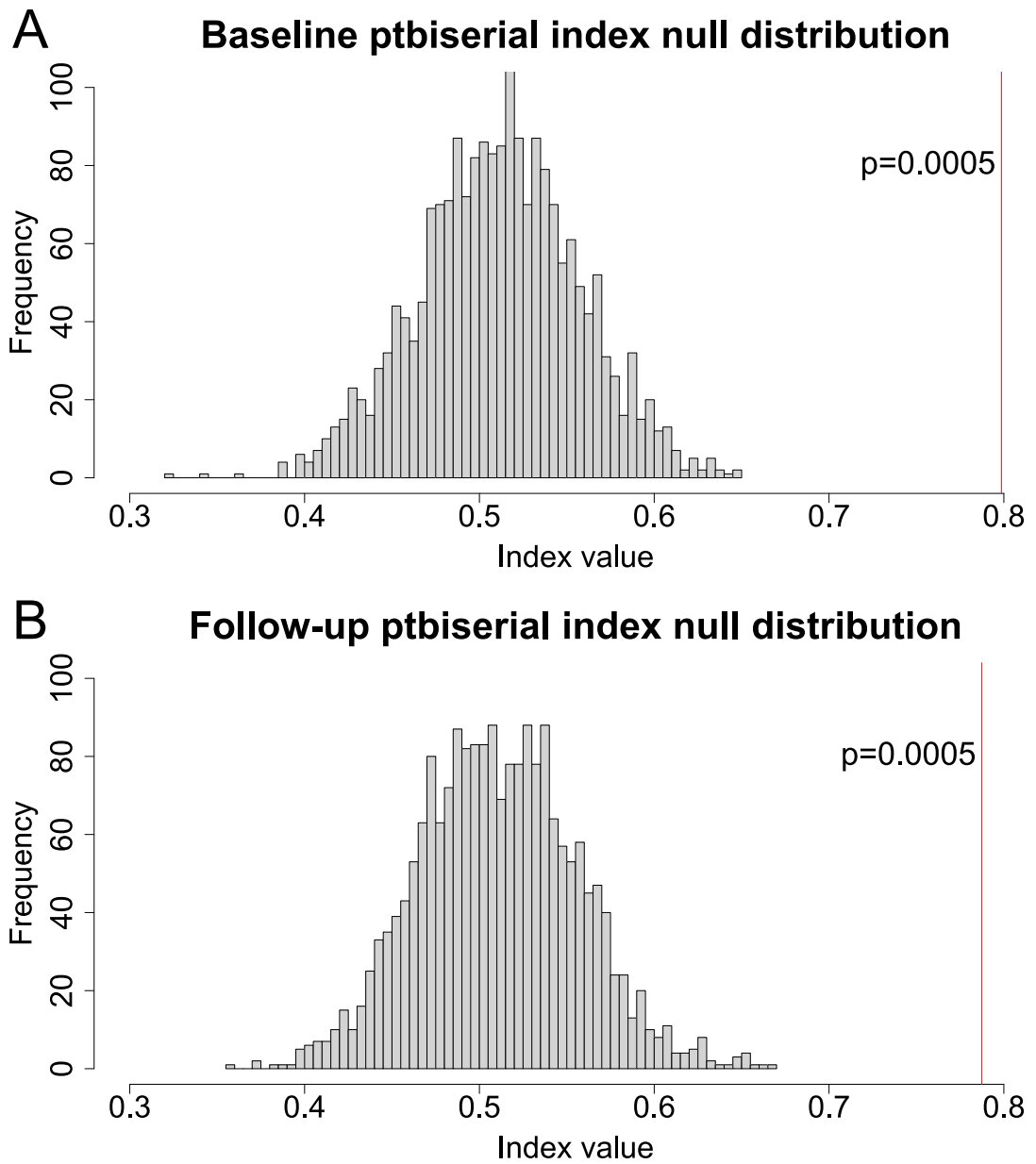


Figure 3-5: Null distributions and real values of the ptbiserial clustering index for baseline and follow-up data. A: The index value (0.8) from the actual data at baseline, annotated by the red line, was significantly different compared to data simulated from a distribution with no clusters. **B:** The index value (0.79) from the actual data was also significantly different at follow-up.

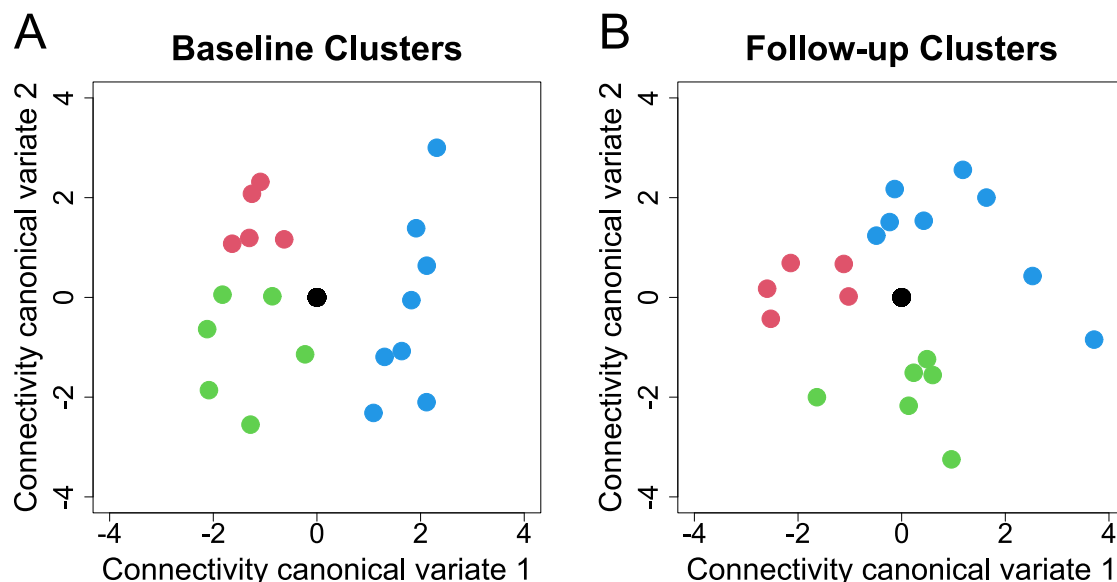


Figure 3-6: The obtained 4-cluster solution for the RA study. A: The 4 clusters observed after hierarchically clustering the first two connectivity variates at baseline. The black dot depicts the biggest cluster (31) where all the patients had values close to 0 for variates 1 and 2. The other clusters are depicted in blue (8), green (6), and red (5). **B:** The 4 cluster-solution using the same pipeline at follow-up with the same 31:8:6:5 ratio, in which 41 out of 50 patients stayed in the same clusters as at baseline.

3.3.4 Baseline clusters did not differentiate on clinical traits

The clusters did not differ based on baseline disease activity, ESR, sleep, pain, or depression (**Table 3-4**) as tested by individual Kruskal Wallis tests.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number of members	31	8	6	5
RA disease activity	3.24 (1.8)	4.28 (2.53)	2.91 (1.39)	4.46 (1.46)
ESR (mm/h)	14 (13)	20 (13)	22 (13.5)	22 (8)
Fatigue	35(10.5)	49 (30.5)	34 (30.5)	37 (9)
Sleep disturbance	16 (8.5)	15 (8.25)	15.5 (4)	15 (6)
Pain severity	4 (3.5)	3 (5.25)	4.5 (3.25)	2 (3)
Depression	6 (5.5)	7 (4.5)	8.5 (4.75)	9 (3)

Table 3-4: Clinical characteristics of RA study clusters Median (IQR). The clusters were not significantly different on any of the clinical characteristics based on Kruskal Wallis tests (RA disease activity $H(3) = 2.2$, $p = 0.5$; ESR $H(3) = 2.8$, $p = 0.4$; Fatigue $H(3) = 1.3$, $p = 0.7$; Sleep $H(3) = 0.7$, $p = 0.86$; Pain $H(3) = 1.26$, $p = 0.7$; Depression $H(3) = 1.01$, $p = 0.8$). ^aDAS-28; ^bBristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ) total score; ^dJenkin's sleep scale; ^eCurrent pain NRS; ^fHospital Anxiety and Depression Scale; NRS: numerical rating scale.

3.3.5 Failed connectivity-fatigue association and cluster replication in LIFT

The CCA in baseline LIFT data yielded a non-significant correlation (**Figure 3-7A**: $r^2=0.3$, $p=0.8$) which was unstable (cross-validation mean=0.07). The hierarchical clustering led to a 3-cluster solution (**Figure 3-7C**: 37:27:18 ratio) with none of the 8 cluster indices being significant above the $p<0.001$ level (**Figure 3-7B**).

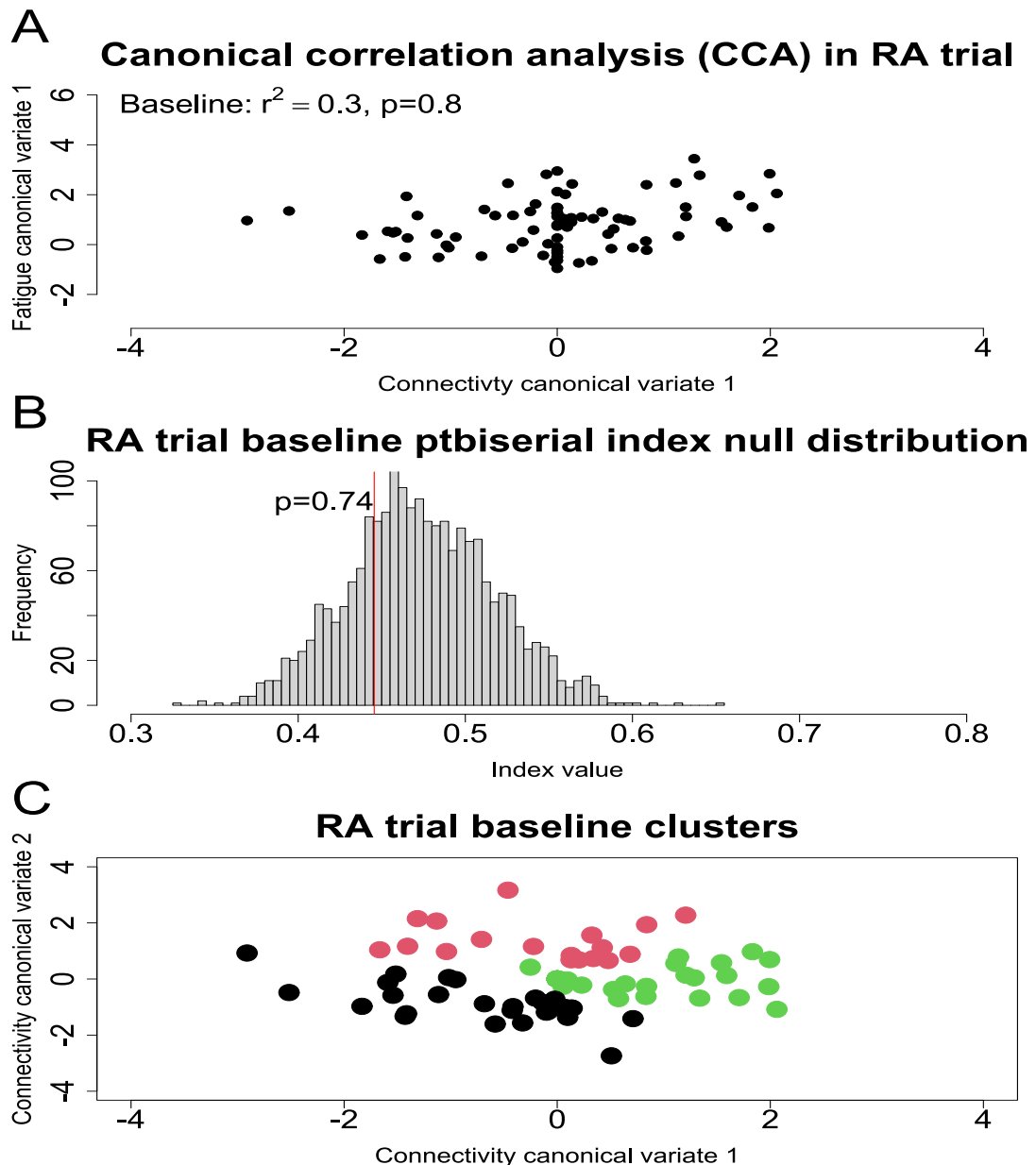


Figure 3-7: Failed replication in the LIFT (RA trial). **A:** The CCA plot shows an insignificant correlation ($r^2=0.3$, $p=0.8$) between connectivity (mean frequency) and fatigue variates. The CCA was conducted after age, gender, and scanning site were regressed from the connectivity estimates. **B:** Hierarchical clustering of the first two connectivity variates led to insignificantly different ($p=0.74$) clusters to ones from simulated data with no underlying clusters based on the ptbiserial index and the other 7 indices (not shown). **C:** The majority of indices (8) set a 3-cluster solution of 37:27:18 ratio displayed on the plot.

3.3.6 Integral brain regions for clustering patients

Two out of four groups of brain connections (circuits) mostly formed the clustering variables at both timepoints. At baseline (**Figure 3-8**), these connections included connectivity between the precuneus and the left posterior temporal fusiform cortex and between the left precentral gyrus and the left anterior parahippocampal gyrus. At follow-up (**Figure 3-9**), connectivity of the precuneus, precentral gyrus, and parahippocampal gyrus among others again contributed to the two circuits, which mainly defined the clustering variables. Additionally, regions like the right posterior supramarginal gyrus and the left and right insular cortex also contributed to the most influential circuits either at baseline or follow-up.

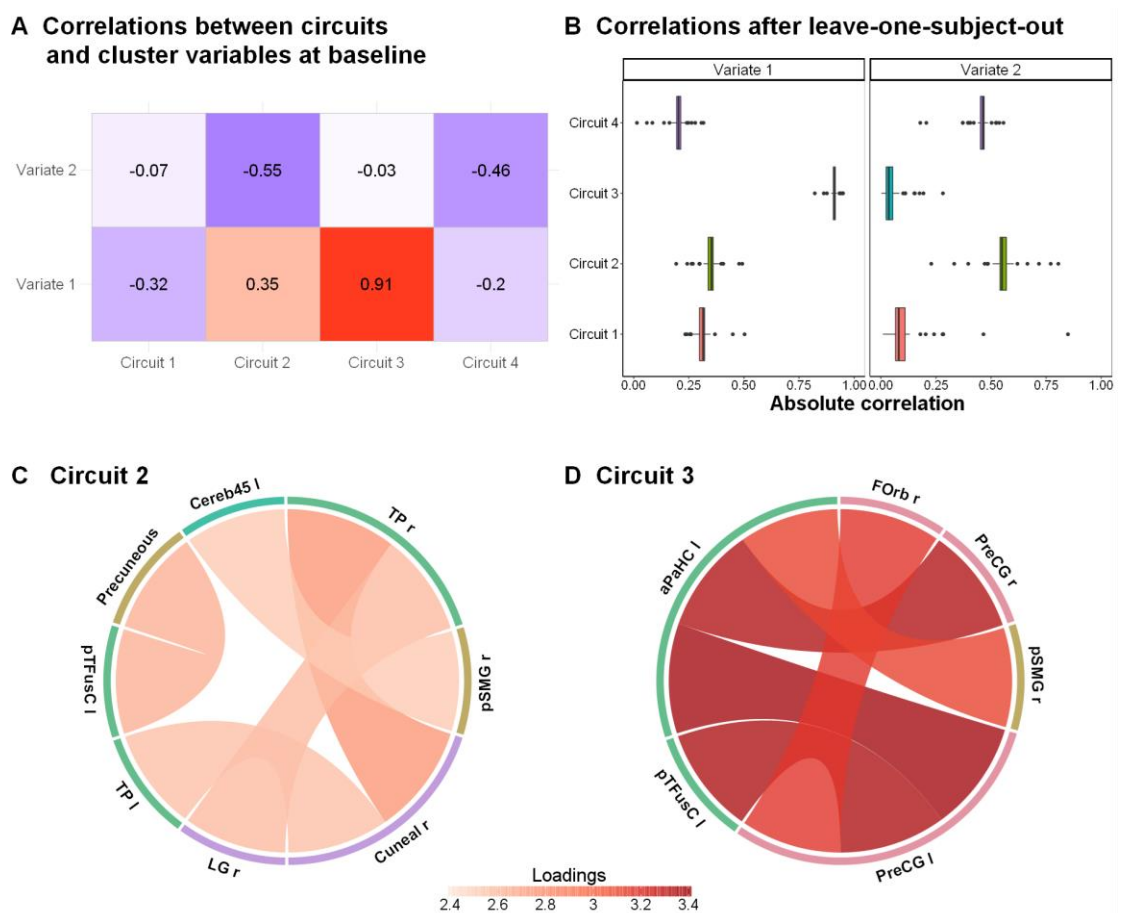


Figure 3-8: Brain connections that contribute to patient clustering at baseline. Panel A displays correlations between the clustering variables and circuits. These circuits are groups of brain connections whose connectivity values change in a similar way across the PASAT blocks of the functional imaging sequence. The correlations identify circuits 2 and 3 as contributing the most to these clustering variables. Panel B displays the stability of these correlations (absolute value) when rerun while leaving one subject out. Panel C and D depict the 5 brain connections that contribute the most to each circuit represented as z-scored loadings. Abbreviations: aPaHC l, left anterior Parahippocampal Gyrus; Cereb45 l, left Cerebellum 4 5; Cuneal r, right Cuneal Cortex; Forb r, left Frontal Orbital Cortex; LG r, right Lingual Gyrus; pSMG r, right posterior Supramarginal Gyrus; PreCG l/r, left/right Precentral Gyrus; pTFusC l, left posterior Temporal Fusiform Cortex; TP l/r, left/right Temporal Pole.

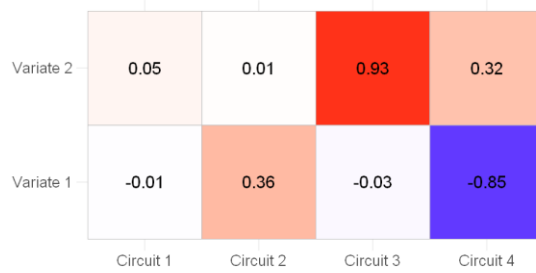
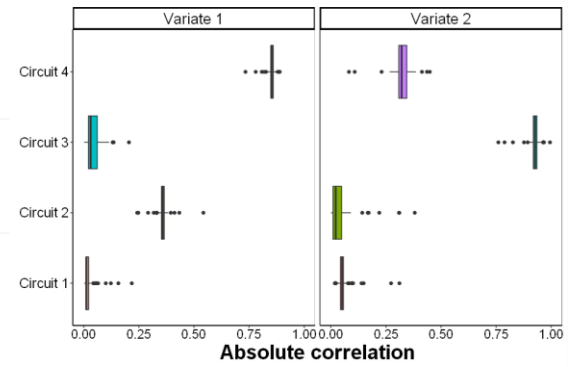
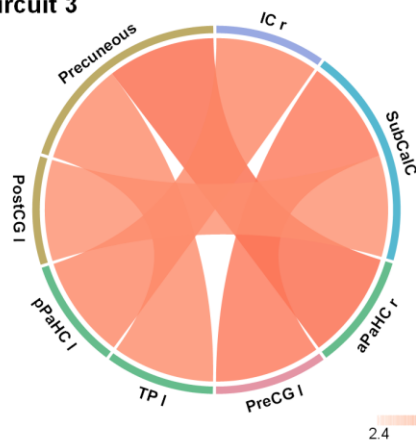
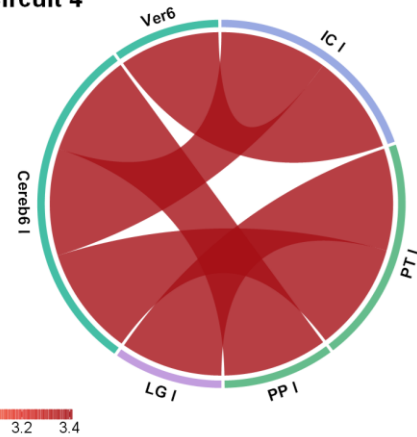
A Correlations between circuits and cluster variables at follow-up**B Correlations after leave-one-subject-out****C Circuit 3****D Circuit 4**

Figure 3-9: Brain connections that contribute to patient clustering at follow-up. Panel A displays correlations between the clustering variables and circuits. These circuits are groups of brain connections whose connectivity values change in a similar way across the PASAT blocks of the functional imaging sequence. The correlations identify circuits 3 and 4 as contributing the most to these clustering variables. Panel B displays the stability of these correlations (absolute value) when rerun while leaving one subject out. Panel C and D depict the 5 brain connections that contribute the most to each circuit represented as z-scored loadings. Abbreviations: aPaHC r, right anterior Parahippocampal Gyrus; Cereb6 l, left Cerebellum 6; IC l/r, left/right Insular Cortex; LG l, left Lingual Gyrus; PP l, left Planum Polare; PT l, left Planum Temporale; PostCG l, left Postcentral Gyrus; pPaHC l, left posterior Parahippocampal Gyrus; PreCG r, right Precentral Gyrus; SubCalC, Subcallosal Cortex; TP l, left Temporal Pole; Ver6, Vermis 6.

3.4 Discussion

This is the first instance of using ML to combine brain metrics with clinical data that produced neurobiological subgroups of RA patients, but which may not generalise to all RA populations. In this study, dynamic brain connectivity correlated with fatigue (BRAFMQ) in RA patients confirming the hypothesis of a significant relationship between brain connectivity and fatigue. This association partitioned the patients of the study into 4 distinct clusters that replicated using data after 6 months. The clusters were similar in terms of disease activity, inflammation, pain, sleep, and depression, which suggests that these variables did not significantly affect the neurobiological differences between the subgroups. However, both the correlation and cluster results failed to emerge using the baseline data of the LIFT trial and externally validate these neurobiological subgroups.

Although novel in RA, dynamic connectivity has previously been investigated in MS, another autoimmune disorder with multiple potential mechanisms of fatigue (Manjaly et al., 2019). Studies have linked dynamic connectivity of specific brain regions with fatigue (Zhou et al., 2016), amongst others like physical and cognitive disability (Valsasina et al., 2019). Moreover, cluster analysis on such links has identified aberrant brain networks specific for accepted subtypes of MS (Eijlers et al., 2019). In comparison, the cluster analysis in this study did the reverse of finding new subgroups of patients using neuroimaging and clinical traits. In psychiatry, such an approach using clinical variables has been long established. For example, which antidepressant should be selected for a patient would be informed by whether a patient exhibits symptoms of atypical or other subtypes of depression (Quitkin et al., 1993). More recently, neuropsychiatric research has implemented brain-behaviour associations to identify potential “biotypes” and then search for therapeutic relevance in psychosis (Planchuelo-Gomez et al., 2020), autism (Hong et al., 2018), attention deficit hyperactivity disorder (Barth et al., 2018), depression and anxiety (Williams, 2017). However, these biotypes are yet to improve long-term treatment outcomes as hoped (van Loo et al., 2014). An alternative approach would be to start with a therapeutic response to an anatomically targeted intervention. For instance, differences between anatomical sites of TMS were exploited in a study of MDD (Siddiqi et al., 2020). The authors found retrospectively the best improvement in two distinct symptom clusters when two different brain circuits were stimulated that did not occur in a placebo condition. These findings could then be prospectively applied in future clinical trials of this treatment. In a similar fashion,

TMS targets in RA could be planned in advance depending on whether patients' fatigue is connected to an inflammatory or cognitive origin.

Some brain regions overlapped in governing how the clustering algorithm grouped patients at baseline and follow-up, regions previously linked to fatigue, pain, or systemic inflammation. These brain regions included the precuneus and parahippocampal gyrus, integral regions of the default mode network (DMN), as well as the precentral gyrus, a main region of the somatosensory network (SMN). DMN activity indicates introspective behaviour, which can enhance cognitive functions, but excessive DMN use may result in fatigue. The DMN and SMN are critical networks in pain processing, and a systematic literature review identified pain as one of the strongest predictors of RA fatigue (Nikolaus et al., 2013). The bilateral insula was also involved in clustering patients. This further strengthens the link between fatigue and pain, as connectivity between the insula and the DMN in the same patients was previously associated with fibromyalgia—a measure of centralised pain (Basu et al., 2018). Unlike other common predictors, pain clusters with fatigue across almost all subjects (Basu et al., 2017). Thus, it is unsurprising that fatigue and pain appear to share neurobiological mechanisms. Systemic inflammation also consistently clusters with fatigue, and in the same RA patients was associated with connectivity of the inferior parietal lobule, made up of the supramarginal gyrus and angular gyrus (Schrepf et al., 2018). The connectivity of the supramarginal gyrus contributed to identifying fatigue clusters of patients at baseline. Direct associations also exist between the inferior parietal lobule and fatigue. In MS patients, fatigue was associated with the atrophy of the right inferior parietal lobule (Hanken et al., 2016), while fatigue in chronic fatigue syndrome has been positively correlated with hyperconnectivity between the inferior parietal lobule and the sensorimotor cortex (Boissoneault et al., 2016). Overall, the brain regions identified in fatigue clusters of RA patients reinstate how interconnected fatigue is with factors such as pain and systemic inflammation and suggest some common neurobiological origins.

Despite the brain regions that separated the subgroups being previously associated with pain or systemic inflammation, these subgroups were not clinically separable. Ambrosen et al. (2023) also did not find clinical differences among schizophrenia subgroups based on brain connectivity. They highlighted the limitations of cluster analysis as a data-driven method, lacking the ground truth of the problem. In this RA analysis, the ground truth is the neural circuits implicated in fatigue during patient stratification. Therefore, subgroups may still reflect irrelevant traits, even with CCA identifying fatigue-relevant information in neuroimaging data. Another study used morphometric MRI from large datasets, Generation

Scotland (GS subsample, N=980), and UK Biobank (UKB, N=8,900), clustering participants and testing whether they differ in depression status and cognitive measures (Yeung et al., 2021). No depression status associations were found, attributed to few individuals having current depression, but associations with general cognitive ability were observed. RA patients in this analysis may not have substantially differed in pain and systemic inflammation for the clustering approach to detect. Additionally, unbalanced cluster sizes may have reduced statistical power to identify such differences. Overall, clustering limitations and potential lack of statistical power may have led to clinically inseparable patient clusters despite their underlying neurobiological differences being linked to those clinical characteristics in previous research.

There are several limitations of the methods concerning the high variability of the RA datasets and the adopted pipeline. The clustering analysis by Drysdale et al. (2017) was used on a homogenous sample of depressed patients that may not be suited for more heterogeneous data (Grosenick et al., 2019). Such was the case for this analysis, in which patients from the RA study had more inflammation but less fatigue compared to the LIFT patients (**Table 3-1, Figure 3-3**), and the replication study by Dinga et al. (2019). This is consistent with observations of considerable heterogeneity in RA. Even amongst selected patients for severe fatigue, four different clusters have emerged and validated using variables reflecting pain, fatigue, mental health, disability and inflammation (Basu et al., 2017). The pipeline of using dynamic connectivity and ICA may have avoided the overfitting of the original paper but possibly resulted in insufficient brain metrics (only 4 variables) to extract enough clinically relevant data. This is especially relevant if the underlying biological mechanisms differ between the two cohorts. The power to find significant effects would be decreased if specific fatigue dimensions have different effects on dynamic connectivity in different groups of patients. A possible solution to both limitations is applying contemporary versions of CCA that can combine structural and functional brain metrics with behavioural symptoms (Ing et al., 2019). If incorporated into rigorous statistical and clinical validation (Mihalik et al., 2020), such an approach could extract enough brain-related signals that are sufficient to take advantage of the variance in the sample. Once neurobiological subgroups of fatigue are identified that can generalise across different IA cohorts, homogeneous connectivity patterns can emerge that underlie the brain dysfunction related to fatigue. Animal studies could then dissect how manipulating the dysfunctional brain circuits affects animal behaviour with molecular techniques such as optogenetics, whereas human studies can modulate connectivity via non-invasive neurostimulation and develop patient-specific therapeutics. The subsequent findings could then be integrated into clinical practice, such as

using clusters like in this study to determine if the brain connectivity profile of a patient can be directly affected to improve their fatigue or can be addressed with pharmacological, exercise, or psychological interventions.

Chapter 4 Neuroimaging markers of subdimensions of fatigue

4.1 Introduction

This chapter builds on **Chapter 3**, approaching the same question from a different perspective. While the prior chapter explored neurobiological subgroups using neuroimaging and assessed their phenotypic variations, this one begins with different fatigue phenotypes that can potentially be neurobiologically distinct. Brain imaging can identify neurobiological correlates of central fatigue, which can be divided into different subtypes, such as ‘motivational’, ‘physical’, and ‘cognitive’ fatigue (Karshikoff et al., 2017). This raises the question of whether neurobiological associations differ among central fatigue subtypes. Currently, no formal definition exists for fatigue subtypes and questionnaires assessing facets of fatigue do not reveal clear subsets of patients with subtypes of fatigue (Chorus et al., 2003). Subtypes of fatigue may not underly clinically distinct subgroups, but instead represent different aspects of the same symptom. Each fatigue facet could be linked to the disruption of a specific neural circuit, such as the mesolimbic (motivational), nigrostriatal (physical), and mesocortical (cognitive) dopamine pathways (Korte and Straub, 2019). Although these pathways are distinct, a common stressor can drive the brain towards a fatigued state, affecting multiple neural circuits. In the context of IA, inflammatory cytokines can interfere with neurotransmitter release in multiple brain regions (Soliven and Albert, 1992), or chronic psychosocial stress can change noradrenaline metabolism (Matsumoto et al., 2021). However, no studies have yet compared brain imaging correlates of subdimensions of fatigue in IA. This study will use neuroimaging data to identify descriptors of fatigue phenotypes and subsequently examine their differences. In practical terms, while **Chapter 3** outlined potential biological subgroups for mechanistic studies, this chapter offers phenotypic neurobiological descriptors.

To identify brain imaging correlates of fatigue, studies can agnostically test for associations across the brain for links with fatigue. However, the agnostic approach suffers from a curse of dimensionality since looking at brain structure or connectivity at every voxel would result in over 60 billion individual statistical tests. Instead, larger regions of interest (ROIs) from brain atlases can be used, followed by ROI-to-ROI connectivity analyses, providing sufficient sensitivity while reducing the multiple-comparison problem to focus on a limited number of ROIs. Although this approach may compromise spatial specificity, subject-

specific ROIs extracted from surface-based analyses can partially alleviate this limitation (Nieto-Castanon, 2022). Brain imaging analysis can identify fatigue correlates by exploring associations across a limited number of subject-specific ROIs that cover the whole brain.

Another methodological question is whether to use FC from the resting-state or the PASAT task. Although initial FC studies emphasised task-related states (Friston, 1994), resting-state FC prevails in the field due to the concept of an "intrinsic" functional network, bypassing the need to consider numerous task states (Fox and Raichle, 2007). There can be two perspectives on this topic (Gratton et al., 2018). One posits that functional networks mirror cognitive, perceptual, and motor processes, adapting significantly to context, task demands, moods, and fleeting thoughts. However, this implies that networks may be less reliable for clinical applications, being strongly influenced by specific measurement conditions like the scanner environment, rather than the underlying disease. The alternative view suggests that functional networks are fundamentally stable. Therefore, an fMRI scan could measure network properties informative about a person's stable traits (e.g., disease status or personality) regardless of their thoughts, mood, or task during the scan. If accurate, this indicates that functional networks can gauge individual traits and monitor diseases, holding potential for personalized medicine, but may have limited utility for assessing cognitive content.

Evidence indicates that fMRI FC undergoes only minimal reconfiguration during tasks compared to rest. For instance, one study examined FC between numerous brain regions across 64 task states and rest in individual subjects and found an "intrinsic" functional network architecture present across many tasks that highly resembles the resting-state architecture (Cole et al., 2014). While task-evoked changes in this architecture were common across tasks, task-specific FC alterations were relatively small but significant overall. The authors replicated their findings in a larger dataset with seven distinct tasks. Additionally, a similar study demonstrated a correlation of 0.75 between resting-state FC and FC across 14 tasks, ranging from passive fixation to increasingly demanding classification tasks (Krienen et al., 2014). Sources of variation between rest and task state FC are also differentially distributed across the brain, and state-based modulations vary widely across individuals. This implies that functional networks are best suited for measuring stable individual characteristics, and state changes should be considered in network comparisons (Gratton et al., 2018). Furthermore, Hearne et al. (2017) found that increased reasoning demands lead to selective patterns of connectivity within cortical networks, supplementing the previously mentioned general task-induced architecture. These findings, along with others in healthy

and MS cohorts (Gianni et al., 2021, Gonzalez-Castillo and Bandettini, 2018), suggest that task-based FC surpasses resting-state FC in task-related networks. Consequently, resting-state FC might not directly correlate with a behavioural function in the same way that a task-based measure of connectivity does. Convergent findings from both types of FC have already been shown to improve our understanding of adolescent brain network maturation (Stevens, 2016). Finally, studies have used both resting-state and task FC to complement each other (Harrewijn et al., 2020, Bolt et al., 2017, Cohen and D'Esposito, 2016), as was also done in this study.

This chapter will compare brain imaging correlates of fatigue subdimensions in IA. These will include morphometric features, FC extracted from fMRI during rest and the PASAT task. I will agnostically identify associations with baseline scores of subdimensions of fatigue using the same BRAF MDQ questionnaire in RA patients from the LIFT trial. Although the previous chapter found no phenotypical differences among the neuroimaging subgroups, I expect fatigue phenotypes to differ in neuroimaging correlates in this analysis. The subgroups might have exhibited internal heterogeneity, while the fatigue phenotypes in this context are inherently distinct and are linked to the functions of different neural circuits. Additionally, the methodological approach will examine each fatigue phenotype independently to better identify potential differences. I hypothesise that:

- There are statistically significant correlates of subcomponents of fatigue (physical, living, cognitive, emotional)
- Some significant components will be common across fatigue subcomponents and others will be distinct for the different fatigue subcomponents.

4.2 Methods

Patient recruitment, clinical and MRI data descriptions for LIFT have already been described in **Chapter 2.1.2**. T1-weighted scans quantified grey matter morphometric features (volume, thickness, surface area) of brain regions while DTI estimated their SC as white matter structural integrity between each pair of regions. Functional scans from both resting-state and PASAT depicted FC between brain regions. The Desikan-Killiany anatomical atlas (Desikan et al., 2006) defined 84 brain regions for all modalities. These regions can also be related to specific brain networks (Error! Reference source not found.). The pre-processing of the resulting images included surface-based segmentation of grey and white matter (Dale et al., 1999) using the Free Surfer 6 software. This was followed by the default surface-based CONN pipeline for fMRI which included the same steps as **Chapter 3.2.1** apart from centralisation and normalization. The FMRIB software packages (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>) pre-processed the DTI data which was then fed into probabilistic tractography to estimate SC between atlas regions using the BEDPOSTX and PROBTRACKS tools (Behrens et al., 2007, Hernandez et al., 2013). The outputs were measures of grey matter volumes, thickness, and surface area with two functional and one structural connectivity matrices made of values for each pair of regions.

This study used FC during both rest and the PASAT task. Firstly, this enabled comparisons between results in this thesis as the LIFT trial had both modalities but others like the RA study dataset 2 and the PsA dataset had only resting-state fMRI while the RA study dataset 1 had only PASAT fMRI. Secondly, while rest and task FC offer similar information, their distinct characteristics have been shown to provide unique insights into the investigated behaviour (Schultz and Cole, 2016). Specifically, the PASAT task has previously been shown to cause temporary mental fatigue in conditions such as chronic fatigue syndrome (Cook et al., 2007) and the autoimmune disease granulomatosis with polyangiitis (Basu et al., 2014), potentially reflecting connectivity relevant to clinical features such as 'mental fog'. However, the intention here is not to differentiate patterns of the "intrinsic" functional network architecture from task-specific architecture but rather to offer complementary information. Also, while the PASAT was chosen for its fatiguing nature, the experience of the MRI scanner itself can be fatiguing and may contribute to patient fatigue. The task may also engage other cognitive dimensions, and it remains unclear whether the fatigue observed is clinically significant or similar to normal tiredness. Nevertheless, previous research has indicated that the PASAT task may more effectively engage attention networks than resting-

state (Basu et al., 2019, Audoin et al., 2005), which could be a key contributor for one or several subcomponents of fatigue, central to the focus of this study.

To analyse the data in a univariate way, I employed GLMs that associate fatigue levels with a single neuroimaging variable while controlling for variables of no interest (age, gender, MRI imaging site). GLMs are an extension to the simple linear regression that can explore non-linear relationships and whose error distribution of the target variable is not limited to a Gaussian distribution (Schmettow, 2021). The resulting analysis tests the null hypothesis that the coefficient for the selected variable is zero. Because each variable is modelled individually, all tests are controlled for multiple comparisons. This method was implemented using a custom MATLAB code based on the “conn_glm” function of the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012) that can be found: https://github.com/krisbg95/LIFT/tree/f086a8b0bb4981c19d675299f7ba9b6f2c0f74f0/Univariate_KS/BRAF

Baseline neuroimaging variables from the same 84 brain regions from Desikan-Killiany atlas were individually tested whether they associate with sub-scores of BRAF fatigue (**Appendix A**) that measure the physical, impact on living, cognitive and emotional fatigue. To achieve this, the covariates (mean effect, gender, age, site, fatigue) were used in a design matrix within a general linear model. For grey matter features only, total intracranial volume (TIV) was also used as a covariate. Because this analysis was an agnostic approach, all p values were adjusted for multiple comparisons using FDR. The significance level for the morphometric features and FC was set at $p < 0.05$. For SC, this was set at $p < 0.025$ because twice as many connections are tested per region.

4.3 Results

Before RA patients were allocated to the UC, PEP, or CBA treatment groups, they completed the BRAF MDQ questionnaire, which measures fatigue levels as a sum of physical (median = 15.87, IQR = 5), impact on living (median = 10.04, IQR = 6), cognitive (median = 7.31, IQR = 5), and emotional (median = 5.96, IQR = 4.25) dimensions of fatigue. All sub-scores were significantly correlated with each other although the strength of their association varied (**Figure 4-1**). The physical and cognitive dimensions shared the least variance while the emotional scores had the strongest correlations with both the living and the cognitive scores. Each of these sub-scores was then used as outcome variables in general linear models to find correlates of subdimensions of baseline fatigue using grey matter volumes, thickness, and surface area with two functional and one structural connectivity matrices.

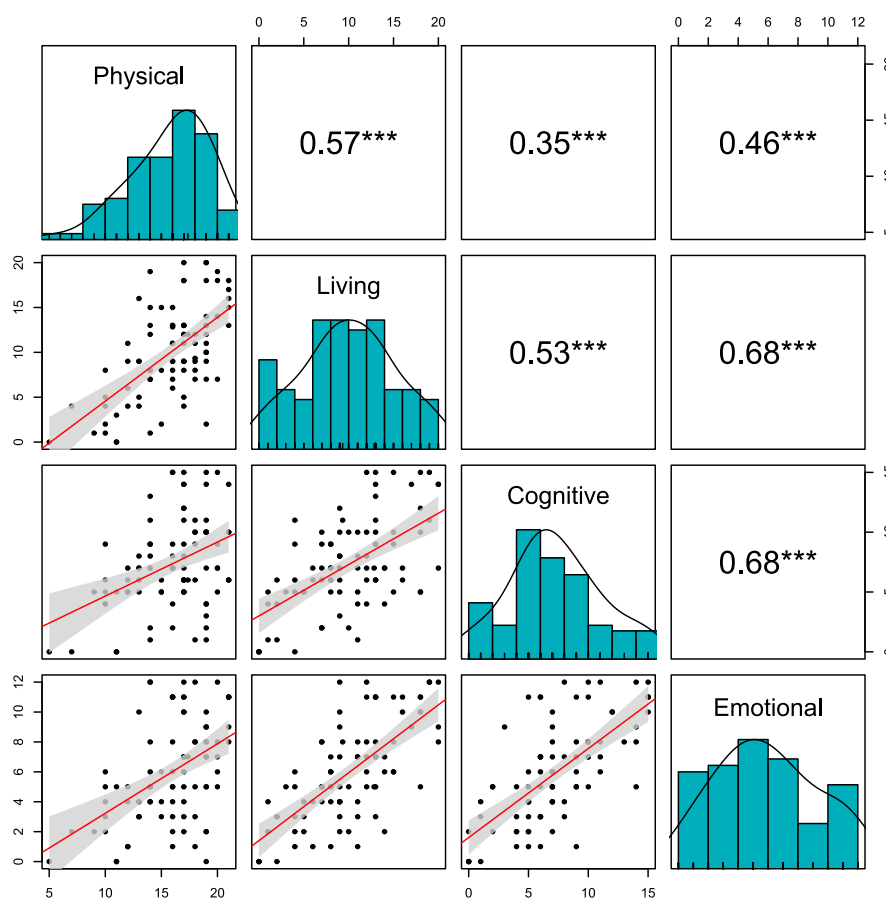


Figure 4-1: Correlation matrix and scatterplots of fatigue subdimensions. The figure displays the correlation matrix (Spearman's rank correlation) and scatterplots of the four sub-scores of the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ): Physical, Living, Cognitive, and Emotional. Each scatterplot represents the relationship between two sub-scores, with regression lines and their 95% confidence intervals. The significance stars denote the strength of the associations while the histograms along the diagonal show the distributions of each sub-score. Density curves are displayed in turquoise. $p < 0.001$ ***

Looking at morphometric brain metrics, grey matter thickness of the left insula and right lingual gyrus negatively correlated with cognitive fatigue scores (**Table 4-1**). Structural connectivity of the right paracentral lobule to multiple areas was negatively associated with living subdimensions of fatigue, along with connectivity of the right and left pericalcarine cortex, the right lingual gyrus and the left middle temporal gyrus (**Table 4-2**). Connectivity of the right lingual gyrus also correlated with physical fatigue while that of the right pericalcarine cortex – with emotional fatigue scores.

All fatigue sub-scores correlated with resting-state FC of frontal areas (**Table 4-3**), including the left lateral orbitofrontal cortex, the right caudal middle frontal gyrus, and the left pars orbitalis. While most of the FC of the frontal areas was positively correlated with fatigue scores, connectivity between the left pars orbitalis and the left supramarginal gyrus was negatively correlated with both cognitive and emotional fatigue scores. In comparison, PASAT FC resulted in multiple positive correlates of physical fatigue with connectivity of sensorimotor regions, including the right and left postcentral gyrus and the right precentral gyrus (**Table 4-4**). Additionally, the living fatigue sub-scores were positively associated with connectivity of the left and right fusiform gyrus as well as the left pars triangularis, left nucleus accumbens, and left cuneus. FC of the left frontal pole with the left paracentral lobule was also associated with the living fatigue sub-scores while connectivity with the right paracentral lobule was correlated with the emotional fatigue sub-scores.

Grey Matter Thickness

ROI Seed	t statistic	p value	Effect size
Cognitive Fatigue Score			
Right Lingual	-3.76	0.02	0.15 (small)
Left Insula	-3.30	0.05	0.12 (small)

Table 4-1: Morphometric correlates of fatigue subdimensions. Significant results ($p < 0.05$) using general linear models on baseline grey matter thickness variables to associate with baseline subcomponents of fatigue scores in all RA patients from the LIFT study ($n=88$). The fatigue scores are based on the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ) which has sub-scores of physical, living, cognitive, and emotional dimensions. All analyses were adjusted for age, gender, imaging site, and total intracranial volume. The p values are after false discovery rate (FDR) correction. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).

Structural Connectivity

ROI Seed	ROI Target	t statistic	p value	Effect size
Physical Fatigue Score				
Right Lingual	Right Bank of the Superior Temporal Sulcus	-4.68	<0.001	0.21 (small)
Left Transverse Temporal Gyrus	Right Pericalcarine	-3.91	0.016	0.16 (small)
Living Fatigue Score				
Right Paracentral Lobule	Right Inferior Parietal Lobule	-3.62	0.012	0.14 (small)
	Right Parahippocampal	-3.62	0.012	0.14 (small)
	Right Hippocampus	-3.60	0.012	0.14 (small)
	Right Pericalcarine	-3.59	0.012	0.14 (small)
	Right Lateral Occipital Gyrus	-3.50	0.012	0.13 (small)
	Right Supramarginal Gyrus	-3.25	0.023	0.11 (small)
Right Pericalcarine	Left Transverse Temporal Gyrus	-4.17	0.006	0.18 (small)
	Left Pericalcarine	-3.86	0.012	0.15 (small)
	Left Isthmus Cingulate	-3.86	0.018	0.15 (small)
Right Lingual	Left Middle Temporal Gyrus	-4.05	0.010	0.17 (small)
	Left Superior Temporal Gyrus	-4.05	0.009	0.17 (small)
Left Pericalcarine	Right Cuneus	-3.80	0.012	0.15 (small)
Left Middle Temporal Gyrus	Left Amygdala	-3.61	0.022	0.14 (small)
Emotional Fatigue Score				
Right Pericalcarine	Right Paracentral Lobule	-3.68	0.017	0.14 (small)
	Left Hippocampus	-3.68	0.017	0.14 (small)

Table 4-2: Structural connectivity correlates of fatigue subdimensions. Significant results ($p < 0.025$) using general linear models on baseline structural connectivity variables to associate with baseline subcomponents of fatigue scores in all RA patients from the LIFT study ($n=87$). The fatigue scores are based on the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ) which has sub-scores of physical, living, cognitive, and emotional dimensions. All analyses were adjusted for age, gender, and imaging site. The p values are after false discovery rate (FDR) correction. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).

Resting-state Functional Connectivity

ROI Seed	ROI Target	t statistic	p value	Effect size
Physical Fatigue Score				
Left Lateral Orbitofrontal Cortex	Left Medial Orbitofrontal Cortex	3.60	0.045	0.14 (small)
Living Fatigue Score				
	Right Parahippocampal	3.67	0.024	0.14 (small)
	Left Bank of the Superior Temporal Sulcus	3.58	0.048	0.14 (small)
Right Caudal Middle Frontal Gyrus	Right Hippocampus	3.38	0.026	0.12 (small)
	Left Fusiform	3.27	0.027	0.12 (small)
Left Pallidum	Left Postcentral Gyrus	4.08	0.009	0.17 (small)
Right Ventral Diencephalon	Left Fusiform	3.61	0.043	0.14 (small)
Cognitive Fatigue Score				
Left Pars Orbitalis	Left Supramarginal Gyrus	-4.12	0.007	0.17 (small)
Emotional Fatigue Score				
Left Pars Orbitalis	Left Supramarginal Gyrus	-3.71	0.032	0.14 (small)

Table 4-3: Resting-state functional connectivity correlates of fatigue subdimensions. Significant results ($p < 0.05$) using general linear models on baseline resting-state functional connectivity variables to associate with baseline subcomponents of fatigue scores in all RA patients from the LIFT study ($n=87$). The fatigue scores are based on the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ) which has sub-scores of physical, living, cognitive, and emotional dimensions. The common regions between fatigue subdimensions are highlighted in grey. All analyses were adjusted for age, gender, and imaging site. The p values are after false discovery rate (FDR) correction. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).

PASAT Functional Connectivity

ROI Seed	ROI Target	t statistic	p value	Effect size
Physical Fatigue Score				
Right Precentral Gyrus	Left Postcentral Gyrus	3.67	0.018	0.15 (small)
	Right Paracentral Lobule	3.64	0.021	0.15 (small)
	Left Paracentral Lobule	3.34	0.028	0.13 (small)
	Right Fusiform	3.38	0.047	0.13 (small)
	Left Fusiform	3.07	0.036	0.11 (small)
Right Postcentral Gyrus	Left Fusiform	3.65	0.039	0.15 (small)
	Right Fusiform	3.55	0.047	0.14 (small)
	Right Paracentral Lobule	3.36	0.034	0.13 (small)
Left Postcentral Gyrus	Left Paracentral Lobule	2.78	0.043	0.09 (trivial)
	Right Paracentral Lobule	3.89	0.017	0.16 (small)
Left Amygdala	Left Paracentral Lobule	3.44	0.028	0.13 (small)
	Left Lateral Occipital Gyrus	3.82	0.022	0.16 (small)
Living Fatigue Score				
Left Fusiform	Right Cuneus	3.42	0.031	0.13 (small)
	Left Postcentral Gyrus	3.30	0.031	0.12 (small)
	Right Lateral Occipital Gyrus	3.30	0.031	0.12 (small)
Right Fusiform	Right Cuneus	3.59	0.042	0.14 (small)
	Left Cuneus	3.23	0.043	0.12 (small)
Left Pars Triangularis	Left Rostral Anterior Cingulate	4.42	0.003	0.2 (small)
	Right Rostral Anterior Cingulate	3.88	0.009	0.16 (small)
Left Accumbens	Left Bank of the Superior Temporal Sulcus	4.02	0.011	0.17 (small)
	Left Pars Opercularis	3.64	0.020	0.15 (small)
Left Cuneus	Right Middle Temporal Gyrus	3.69	0.035	0.15 (small)
Left Frontal Pole	Left Paracentral Lobule	3.67	0.037	0.15 (small)
Emotional Fatigue Score				
Left Frontal Pole	Right Paracentral Lobule	3.68	0.035	0.15 (small)

Table 4-4: PASAT functional connectivity correlates of fatigue subdimensions. Significant results ($p < 0.05$) using general linear models on baseline functional connectivity variables, extracted from activity during the Paced Auditory Serial Addition Test, to associate with baseline subcomponents of fatigue scores in all RA patients from the LIFT study ($n=87$). The fatigue scores are based on the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ) which has sub-scores of physical, living, cognitive, and emotional dimensions. All analyses were adjusted for age, gender, and imaging site. The p values are after false discovery rate (FDR) correction. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).

Overall, correlates of fatigue subdimensions differed within SC as well as FC extracted from activity during resting-state and the PASAT task (**Figure 4-2**) despite significant correlations between all of the fatigue sub-scores. The physical and living fatigue scores produced the majority of the correlates with many of them focused on the paracentral lobule. Connectivity of the paracentral lobule also correlated with emotional fatigue scores. The resting-state functional connection between the left pars orbitalis and the left supramarginal gyrus was the only connection that correlated with two different scores in the cognitive and emotional subdimensions.

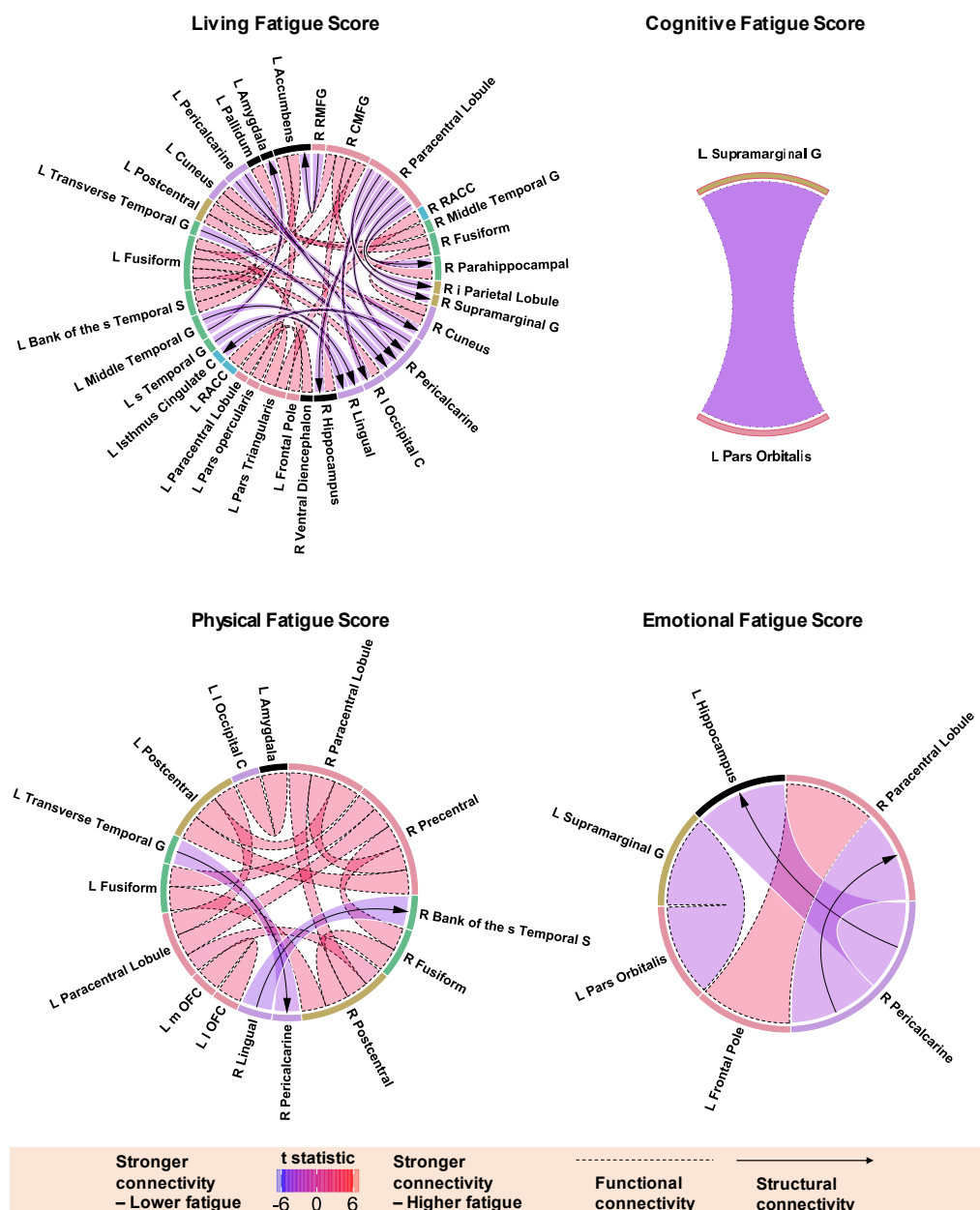


Figure 4-2: Brain connectivity correlates with subdimensions of fatigue. All analyses used a general linear model, adjusted for age, gender, and imaging site. Abbreviations: L/R, left/right; i/s, inferior/superior; m/l, medial/lateral; C, cortex; G, gyrus; CACC, caudal anterior cingulate cortex; CMFG, caudal middle frontal gyrus; RACC, rostral anterior cingulate cortex; RMFG, rostral middle frontal gyrus; OFC, orbitofrontal cortex. The brain images were created using BrainNet software v1.7 (Xia et al., 2013), and the diagram was constructed using circlize R package (Gu et al., 2014).

4.4 Discussion

This was the first study to compare brain imaging correlates of subdimensions of fatigue in IA. The fatigue subdimensions exhibited mostly distinct associations with various brain imaging metrics, despite significant correlations between the subdimensions themselves. These associations could be attributed to separate circuits underlying fatigue subdimensions. Brain circuits can classify central fatigue into motivational, physical, and cognitive subtypes (Korte and Straub, 2019), reflected by BRAF sub-scores. The living score assesses the impact of fatigue on patient motivation by asking whether fatigue affected the planning of activities (**Appendix A**), while it has been shown to correlate with other questionnaires that measure motivation, such as the Short Form 36 General Health Survey (Nicklin et al., 2010, Stansfeld et al., 1997). The living score correlated with FC of the left accumbens and frontal areas, including the caudal and rostral middle frontal gyrus, the frontal pole, the left pars opercularis and triangularis, as well as the rostral anterior cingulate cortex and the isthmus cingulate cortex. These findings may underlie changes in the mesolimbic pathway, which consists of dopamine neurones in the ventral tegmental area projecting to the nucleus accumbens, and subsequently to the orbitofrontal cortex and the anterior cingulate cortex (Koob and Volkow, 2016). This circuit determines the drive for incentives, shaping behavioural responses to rewards based on a cost-benefit analysis of internal and external information. Changes in this circuit are associated with anhedonia (Clery-Melin et al., 2019), which forms a major part of depression symptoms that also correlate with the BRAF living dimension (Nicklin et al., 2010). The brain circuit that induces motivational fatigue could thus increase the perception of energy costs of actions and/or decrease the expectation of rewards.

The living fatigue sub-scores also negatively correlated with SC from the paracentral lobule to six different regions, whereas physical fatigue levels were positively associated with FC between the paracentral lobule and the motor cortex (precentral gyrus) and the somatosensory cortex (postcentral gyrus). The paracentral lobule receives motor and sensory innervations and forms part of the SMN, which also includes the motor and somatosensory cortices. In healthy subjects, FC of the left paracentral lobule with both the insula and the ventromedial prefrontal cortex has previously been positively associated with fatigue, induced by a working memory task (Chen et al., 2020, Wylie et al., 2020). Weaker paracentral lobule connections may predispose people to fatigue, which some theoretical accounts assign to changes in brain systems involved in performing cognitive or motor

operations such as the paracentral lobule (Muller and Apps, 2019). This framework also suggests that the internal states of the task networks can be monitored by a more general network composed of the dorsal anterior cingulate cortex, the dorsolateral prefrontal cortex, and the anterior insula due to their widespread connectivity profile and involvement in interoception, cognitive control and metacognition. For example, interoceptive attention to heartbeats enhances activity in the anterior insula during subsequent judgements of emotional faces (Ernst et al., 2013). Cognitive control is the ability to adapt behaviour in pursuit of a goal. Activity in the dorsolateral prefrontal cortex has been associated with cognitive control in the form of switching between tasks (Swainson et al., 2003). Metacognition is a specific form of cognitive control in which control stems from self-awareness of one's own abilities (Shea et al., 2014). Activity in both the dorsolateral prefrontal cortex and dorsolateral anterior cingulate cortex was associated with metacognition by distinguishing participants who were informed that their cognitive ability was being assessed from subjects performing the same task but without that knowledge (Bengtsson et al., 2009). This general network of the insula, prefrontal and cingulate cortices may then amplify the effort a person perceives in carrying out subsequent behaviours. Changes in the paracentral lobule may thus induce physical fatigue due to its function within a task-related network such as the SMN but also associate with motivational fatigue because of the influence of regions like the insula and prefrontal cortex that monitor the internal state of the somatosensory network.

Among the fatigue sub-scores, only cognitive fatigue correlated with the structural properties of grey matter, wherein reduced thickness of the insula and lingual gyrus corresponded to higher fatigue levels. The lower thickness of the insula and lingual gyrus might suggest a consistently higher cognitive load, as FC of the lingual gyrus has been associated with fatigue in SLE (Barraclough et al., 2021) while anterior insula activity co-varies with the amount of cognitive and physical effort required in healthy cohorts (Chong et al., 2017b). The insula acts as a main hub for interoception—perceiving the physiological condition of the body (Craig, 2002). Interoceptive accuracy, an objective measure of behavioural performance on interoceptive tasks such as counting heartbeats, positively correlates with activity and grey matter volume of the insula (Critchley et al., 2004) as well as insular FC (Chong et al., 2017a). Inflammation is one pathway that may induce interoceptive dysfunction through functional or structural changes in the insula (Hanken et al., 2014). These insular changes may lead to abnormal representations of bodily signals that then exacerbate the rapid development of fatigue (Karshikoff et al., 2017). Interoceptive processing has already been associated with fatigue (Harrison et al., 2009), including in

chronic fatigue disorders like MS (Gonzalez Campo et al., 2020). However, fatigue has had mixed results in associating with interoceptive awareness, the correspondence between the accuracy and the confidence in interoceptive ability, as measured by a subjective questionnaire (Garfinkel et al., 2015). Although fatigue was negatively associated with interoceptive awareness in MS (Rouault et al., 2023), another MS study failed to find a statistically significant relationship (Ware et al., 2023). While both had a similar number of patients and used the Fatigue Severity Scale, one study (Ware et al., 2023) additionally selected participants who consider themselves physically active. Also, Ware et al. (2023) used the total score from the eight-subscale Multidimensional Assessment of Interoceptive Awareness questionnaire (Mehling et al., 2018) compared to Rouault et al. (2023) that used only the “not-worrying” and “trust” subscales to assess the feelings of being in homeostasis and control. In essence, while the link between interoception and fatigue is apparent, further work needs to elucidate the process of how interoceptive changes generate fatigue and which factors are pivotal for this relationship.

Cognitive fatigue in this study was not only associated with the cortical thickness of the insula but also with that of the lingual gyrus, a brain region involved in internally directed attention (Benedek et al., 2016). FC within an attention network significantly predicted the subjective intensity of an interoceptive task in a healthy cohort (Wu et al., 2019). Notably, Haruki and Ogawa (2021) found that while activity in the anterior insula correlated with interoceptive accuracy during a heartbeat counting task, activity in the mid insula differentiated attention to the interoceptive task from that of a control attention task. Interoceptive disturbances have also been associated with major depressive disorder (Dunne et al., 2021), which has established changes in the insula (Namkung et al., 2017). Furthermore, antidepressants have been shown to improve interoceptive function (Eggart and Valdes-Stauber, 2021), suggesting that affecting interoception may be a potential avenue to ameliorating sickness behaviours like depression and fatigue. In the context of this study, inflammation in some RA patients may promote changes in interoception and attention, reflected in neurobiological differences, and enhance and maintain a chronic fatigue state, which can potentially be mitigated by affecting interoceptive processes.

Cognitive fatigue was also negatively associated with the resting-state FC between the left pars orbitalis and left supramarginal gyrus, both integral regions of the DMN (Uddin et al., 2019). The same link was observed in the emotional scores, which showed the strongest correlation with cognitive scores, suggesting that emotional and cognitive fatigue may agglomerate into regions of the DMN, central to both affective and cognitive processes

(Yeshurun et al., 2021). DMN activity differentiated stable SLE patients with high fatigue from healthy controls using fMRI data from both a cognitive working memory task and an emotional facial recognition task (Barraclough et al., 2019). Activity in the DMN decreases progressively with task difficulty, thus acting as an indicator of attentional resources, whose overuse can lead to cognitive fatigue (Wylie and Flashman, 2017). Also, FC within the DMN negatively correlates with fatigue in conditions such as traumatic brain injury (Zhou et al., 2012) and MS (Sjogard et al., 2021). Additionally, traumatic brain injury patients with lower DMN FC and greater fatigue show more evidence of diffuse axonal injury in the form of mean diffusivity within the adjacent corpus callosum (Sharp et al., 2011). Overall, DMN connectivity correlating with fatigue in this study is consistent with the previous literature in autoimmune and neurological disorders, but future work needs to strengthen these findings in IA and determine whether this biomarker changes in response to treatment.

These findings are restricted to the clinical cohort, the BRAF measure, and the nature of univariate analyses. The LIFT cohort has relatively low systemic inflammation that is not representative of a typical RA population and thus certain brain circuits may be more affected than others. Also, the utilized BRAF MDQ questionnaire is a subjective measure of different aspects of fatigue rather than the severity of fatigue that treatments aim to target. Here, the univariate cross-sectional analysis found independent biomarkers that have the strongest effect on different dimensions of fatigue. Although these biomarkers were different between fatigue subdimensions, common patterns of structural or functional brain changes may emerge if tested in a multivariate way. This approach would replace testing independently each brain variable with testing brain patterns that account for dependencies between brain variables. These patterns are derived from dimensionality reduction techniques such as principal component analysis and can be set to best capture the variability across individuals in relation to different types of fatigue as it has been done for reduction of depressive symptoms after treatment (Paolini et al., 2023). Overall, how much fatigue interfered with daily living was associated with the connectivity of motivational brain regions, physical fatigue with SMN regions, and cognitive and emotional fatigue with DMN regions, implying that specific brain circuits may be responsible for these facets of fatigue.

Chapter 5 Predictors of fatigue in RA after non-pharmacological interventions

5.1 Introduction

In IA, fatigue is associated with structural and functional brain metrics, which describe the state of the brain relevant to fatigue. Cross-sectional studies have found these associations by correlating fatigue with different brain measures or differentiating patients with chronic fatigue from healthy cohorts. However, fatigue correlates may not be valuable in predicting how different treatments will affect fatigue. Management of fatigue in IA currently lacks predictive neurobiological markers that are associated with response to fatigue-specific treatments. Predictive markers would help identify patient subgroups benefiting most from treatment or at high risk of unfavourable outcomes. Brain imaging can thus offer neurobiological predictors of fatigue that may inform how IA patients should be treated. To identify brain imaging predictors of fatigue, studies can either use previously found correlates of fatigue or agnostically test for associations across the brain for links with fatigue or treatment response. Neuroimaging studies have used such methods to discover biomarkers of Alzheimer's disease (Ruan et al., 2016), markers of treatment resistance in schizophrenia (Molent et al., 2019) and treatment response in depression (Levy et al., 2019, Kang and Cho, 2020).

This chapter will focus on finding neurobiological markers that predict response to fatigue-specific treatments in IA. Potential predictors of fatigue in this study included FC derived from fMRI during rest or the PASAT task. This choice aligned with the rationales from **Chapter 4**, which were to provide complementary information and offer cross-study comparisons within this thesis. In addition, task-based FC can carry implications for generating predictions. Firstly, resting-state FC can predict task-induced BOLD activity in individuals (Mennes et al., 2010). Secondly, task FC enhances predictions of various traits and behaviours. For example, a study predicted fluid intelligence across two large independent datasets using cognitive task FC (Greene et al., 2018). They demonstrated that task-based fMRI data outperformed resting-state data in modelling trait-relevant individual differences in FC. The best model explained over 20% of the variance, compared to <6% for rest-based models. These results suggest that inducing the appropriate brain state during fMRI can better reveal brain-

behaviour relationships. This approach offers an opportunity to comprehensively characterise individual differences in the neural circuitry underlying complex traits, yielding valuable behavioural predictions. Lastly, a study highlighted the clinical utility of combining resting-state and task FC in predicting attention symptoms in attention deficit hyperactivity disorder (Rosenberg et al., 2016). They initially identified functional brain networks whose strength during a sustained attention task predicted performance differences. Models based on these networks generalised to new individuals, even predicting performance from resting-state connectivity alone. These models then predicted the clinical measure of attention symptoms from resting-state connectivity in an independent sample of children and adolescents. Overall, these findings highlight the potential of using both resting-state and PASAT FC in predicting fatigue.

To address the aim of identifying neuroimaging predictors of response to fatigue-specific treatments in RA, I constructed univariate general linear models based on a preselected and an agnostic approach using subject-specific ROIs. The former investigated fatigue markers from earlier findings (Basu et al., 2019) in RA patients who had 6 months of a personalised exercise programme (PEP) or a cognitive behavioural approach (CBA). The previous study found functional neuroimaging correlates of current fatigue in RA that replicated on repeat scanning after 6 months. It also identified grey matter volume associations with average fatigue over a week. Alternatively, the agnostic approach searched through every neuroimaging variable within the structural/resting-state/PASAT connectivity and structural grey matter metrics. I hypothesised that:

- The previous correlates will not predict fatigue as these are markers of current fatigue in patients who did not receive fatigue-specific treatments.
- There are statistically significant predictors of fatigue scores using the agnostic approach.

5.2 Methods

Patient recruitment, clinical and MRI data descriptions for LIFT have already been described in **Chapter 2.1.2**. A mixed ANOVA determined whether time (baseline, follow-up) and treatment groups interacted in reducing Chalder fatigue (**Appendix C**) adjusting for gender, age, and imaging site. Post-hoc t-tests with Bonferroni corrections then identified which groups had an effect. Brain imaging metrics were extracted using the same preprocessing pipeline as outlined in **Chapter 4.2** while similar general linear models were also used (https://github.com/krisbg95/LIFT/tree/f086a8b0bb4981c19d675299f7ba9b6f2c0f74f0/Univariate_KS/Chalder). Each neuroimaging variable acted as a predictor of post-treatment Chalder fatigue while controlling for baseline levels of fatigue as well as age, gender, and MRI imaging site. These models are useful for finding relevant associations but not individual predictions.

5.2.1 Pre-selected approach

In the previous study in RA (Basu et al., 2019), univariate models depicted that patients with greater grey matter putamen volumes reported higher average over-the-week fatigue levels. The study observed the same positive relationship with current fatigue for PASAT connectivity between the DAN and the bilateral medial prefrontal cortex. This result was replicated on repeat scanning after six months. General linear models tested whether these biomarkers predict follow-up Chalder fatigue in the PEP and CBA groups while correcting for confounding variables like age, gender, imaging site, total intracranial volume and baseline Chalder fatigue. As the neuroimaging variables were preselected based on a previous study, no correction for multiple comparisons was applied (Streiner, 2015). The findings from **Chapter 4** were not included in the pre-selected approach as they represented correlates of subdimensions of fatigue rather than fatigue severity that treatments target.

5.2.2 Agnostic approach

Before constructing predictive models, some adjustment for baseline outcome values is necessary because subjects with extreme scores at baseline tend to improve more than those with average scores, a phenomenon termed regression to the mean. For this reason, studies either adjust for the baseline and predict the raw outcome or predict the difference between the outcome and baseline (change scores). Adjusting for baseline is sometimes preferred

because change scores do not control for baseline imbalances between groups. The ANCOVA adjusts each subject's follow-up score for their own baseline score but is unaffected by chance baseline differences and regression to the mean that can occur in adjusted models (Vickers and Altman, 2001). Even if imbalances are controlled for by randomisation in clinical trials, the expected marginal but non-zero imbalance still has the potential to impact power, type I error rate, and bias in marginal intervention effect estimates (Ciolino et al., 2019). It is recommended to adjust for baseline measures correlated with the outcome, including by the European Medicines Agency (Committee for Medicinal Products for Human Use, 2015). Changed scores negatively correlate with baseline scores so if any potential predictor is correlated with the baseline, that predictor will tend to have a spuriously negative relationship with changed scores.

Proponents of change scores exhibit that adjusting for baseline can inflate ANCOVA estimates (Glymour et al., 2005), known as Lord's paradox (Kim, 2018). They additionally argue that if the predictor is uncorrelated with the transient component of baseline scores, the change score approach is superior (Allison, 1990). Guidelines observe that if treatment is not randomised, change scores should be considered. This is especially the case if selection into treatment is correlated with baseline value, and change scores are preferred because any time-invariant predictors of the outcome are controlled for. However, if change scores are used, they should not be used while adjusting for baseline. This is because, by construction, the baseline score is correlated with the error term when the change score is used as the dependent variable, hence the estimated effect of the baseline on the change score is uninterpretable.

Using the data from the two treatment groups only, baseline neuroimaging variables from the same brain regions of the Desikan Killiany atlas were individually tested whether they predict future post-treatment Chalder fatigue. To achieve this, the covariates (mean effect, gender, age, imaging site, baseline, raw outcome) were used in a design matrix within a general linear model. For morphometric features, TIV was also used as a covariate. Because this analysis was an agnostic approach, all p values were adjusted for multiple comparisons using FDR. The significance level for the morphometric features and FC was set at $p < 0.05$. For SC, this was set at $p < 0.025$ because twice as many connections are tested per region. The results of using change scores are reported in the appendix (**Appendix D**). Additionally, the same analysis was performed for PEP and CBA separately as well as looking into correlates of baseline Chalder fatigue in all RA patients (**Appendix D**).

5.3 Results

5.3.1 Patient characteristics and treatment effects

In LIFT, 88 subjects completed clinical assessments (**Table 5-1**) and at least a T1-weighted scan at baseline [66 females, age 59.3 (11.76)], 87 completed resting-state and DTI, and 82 completed a PASAT scan.

Factor	Baseline	6-month
	LIFT	LIFT
RA disease activity ^a	4.29 (1.64)	4 (1.82)
ESR (mm/h)	11 (10.75)	11 (11.5)
Fatigue (severity) ^b	22 (8.13)	16 (10)
Sleep disturbance ^c	12 (8.5)	9 (8.25)
Pain severity ^d	6 (3)	5 (4)
Depression ^f	6 (6)	5 (5)

Table 5-1: Clinical characteristics of patients in the LIFT trial as Median (IQR). ^aDAS-28 (disease activity score); ^bChalder fatigue scale; ^cJenkin's sleep scale; ^dCurrent pain NRS; ^eHospital Anxiety and Depression Scale; NRS: numerical rating scale.

Patients were balanced (N, missing at follow-up) across the UC (29, 4), PEP (31,6), and CBA (28,5) groups. The 15 patients who did not complete follow-up data did not differ by age, gender, baseline inflammation or baseline Chalder Fatigue but were significantly different on disease activity compared to the 73 patients who completed follow-up (**Table 5-2**).

Factor	Complete	Missing
Sample size	73	15
Gender (female)	76%	66%
Age	62 (19)	58 (9.5)
RA disease activity**	4.25 (1.56)	5 (1.1)
ESR (mm/h)	10 (11)	13 (9)
Chalder Fatigue	21 (7)	22 (8.75)

Table 5-2: Comparison between patients with completed and missing follow-up data as Median (IQR). The 15 patients with missing follow-up did not significantly differ (Mann-Whitney U test) from the 73 patients who completed follow-up on age, gender, Chalder Fatigue score, and inflammation in the form of erythrocyte sedimentation rate (ESR). They did differ in terms of DAS28 disease activity (U = 303, p = 0.008, rank biserial effect size rB = -0.44). ** p < 0.01.

Treatment allocation affected Chalder fatigue scores over the difference between baseline and follow-up with a medium effect size (**Figure 5-1**). Specifically, PEP and CBA but not UC significantly reduced fatigue.

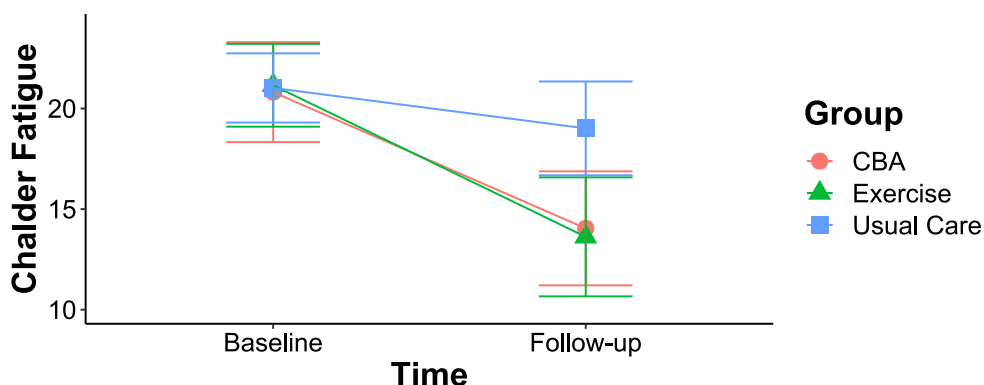


Figure 5-1: Treatment by time interaction in LIFT. The plot displays Chalder fatigue means and 95% confidence intervals of the 3 groups at baseline and 6-month follow-up. Based on a mixed ANOVA, there was a significant Time x Group interaction ($F(2, 67) = 3.88, p = 0.025$, effect size Partial Eta squared $\eta_p^2 = 0.104$), adjusted for age, gender, and imaging site. Bonferroni corrected post hoc testing showed that PEP ($t(24) = 4.863, p < 0.001$) and CBA ($t(22) = 4.355, p < 0.001$) significantly reduced fatigue but not Usual Care ($t(24) = 1.466$, non-significant). Results are averaged over levels of gender and imaging site. Abbreviations; PEP, personalised exercise programme; CBA, cognitive behavioural approach.

5.3.2 Previous fatigue correlates were not predictive of post-treatment fatigue

Previous findings of larger right putamen volumes associating with higher current fatigue failed to predict Chalder fatigue after six months in the active treatment patients (**Figure 5-2**).

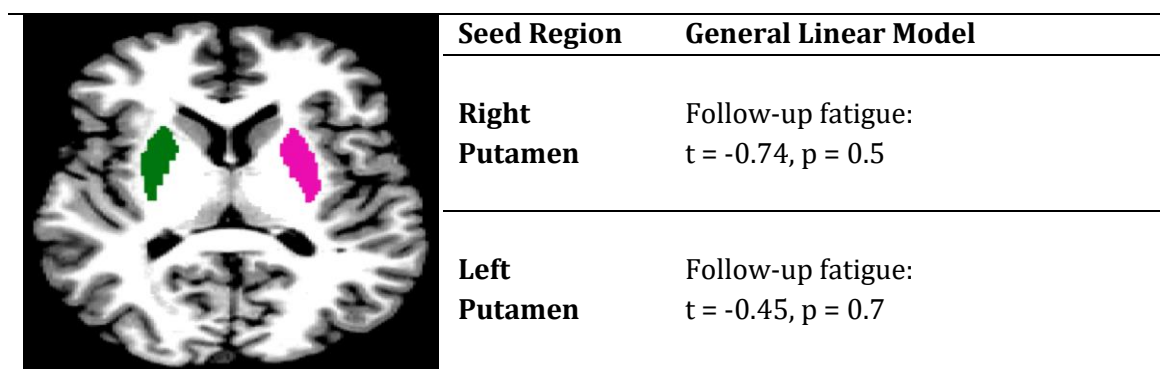


Figure 5-2: Putamen volume analyses based on the RA study. The figure illustrates the left (purple) and right (green) putamen volumes based on the Free Surfer segmentation scheme in the software’s provided subject “Bert”. The table describes putamen volume models of future fatigue based on the previous RA study. General linear models that test whether putamen volume is predictive of follow-up Chalder fatigue in patients with active treatment ($n = 48$) without multiple comparison corrections

Previous findings of higher current fatigue correlating with higher PASAT connectivity between the DAN and the bilateral medial prefrontal cortex also did not predict Chalder fatigue after six months (**Figure 5-3**).

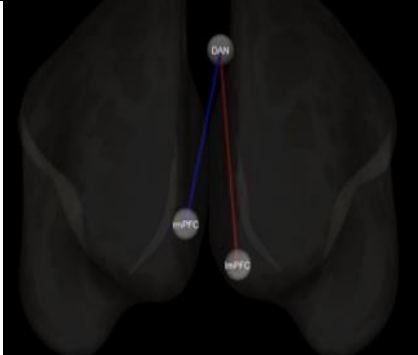
	Seed Region	General Linear Model
	Right medial prefrontal cortex	Follow-up fatigue: $t = 0.13, p = 0.9$
Left medial prefrontal cortex	Follow-up fatigue: $t = -0.65, p = 0.5$	

Figure 5-3: Dorsal attention network (DAN) connectivity with medial prefrontal cortex (mPFC) and fatigue. The figure illustrates the regions-of-interest of DAN and bilateral mPFC and their functional connectivity. The table describes the results of GLM analyses on PASAT functional data in MNI space using the connectivity between the DAN and 8-millimetre spheres of the bilateral mPFC based on coordinates from the previous RA study (left -10, 50, right -12; 8, 54, 0). It lists models that attempt to predict follow-up Chalder fatigue (n=43) without multiple comparison corrections.

5.3.3 Structural and PASAT connectivity fatigue predictors

There were no predictors of future Chalder fatigue using grey matter volumes or resting-state connectivity in patients (n=48) with active treatments (PEP, CBA). For SC, more fibers from the precuneus (a major region of the default mode network) to the anterior cingulate cortex (a major region in the salience network) were predictive of lower fatigue levels after six months in patients with active treatments (**Table 5-3**), visualised as arrow connections in **Figure 5-4**. For PASAT FC, strong predictors of future fatigue included connectivity between two visual areas (left cuneus-right pericalcarine cortex) and between the anterior cingulate cortex and the frontal pole (**Table 5-4**), drawn as dashed connections in **Figure 5-4**.

Structural connectivity

Seed region	Target region	t statistic	p value	Effect size
Left Precuneus	Left Caudal Anterior Cingulate	-4.16	p=0.01	0.29 (medium)
	Right Rostral Anterior Cingulate	-3.97	p=0.01	0.27 (medium)
Right Precuneus	Right Caudal Anterior Cingulate	-4.05	p=0.01	0.28 (medium)
	Right Rostral Anterior Cingulate	-3.99	p=0.01	0.27 (medium)
Right Superior Temporal Gyrus	Right Lateral Occipital Gyrus	-4.01	p=0.02	0.28 (medium)

Table 5-3: Agnostic structural connectivity predictors of future fatigue. Significant results ($p < 0.025$) using baseline single structural connectivity variables to predict future Chalder fatigue in patients with active treatments ($n=48$). Seed regions indicate the brain regions where the structural connection begins and target regions – where it ends. The results are after controlling for confounds (age, gender, imaging site, baseline fatigue) and correction for multiple comparisons using false discovery rate (FDR). Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).

PASAT Functional connectivity

Seed region	Target region	t statistic	p value	Effect size
Left Rostral Anterior Cingulate	Right Frontal Pole	-4.37	p=0.008	0.34 (medium)
Left Cuneus	Right Pericalcarine	4.21	p=0.013	0.32 (medium)

Table 5-4: Agnostic PASAT functional connectivity predictors of future fatigue. Significant results ($p < 0.05$) using baseline single PASAT functional connectivity variables to predict future Chalder fatigue in patients with active treatments ($n=43$). Seed regions indicates brain regions whose connections were tested for predicting fatigue while controlling for multiple comparisons using false discovery rate (FDR). Target regions indicates which connections with the seed regions were significant. The results are after controlling for confounds (age, gender, imaging site, baseline fatigue). Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 .

Baseline brain connectivity predictors of follow-up fatigue after PEP and CBA treatments in RA patients

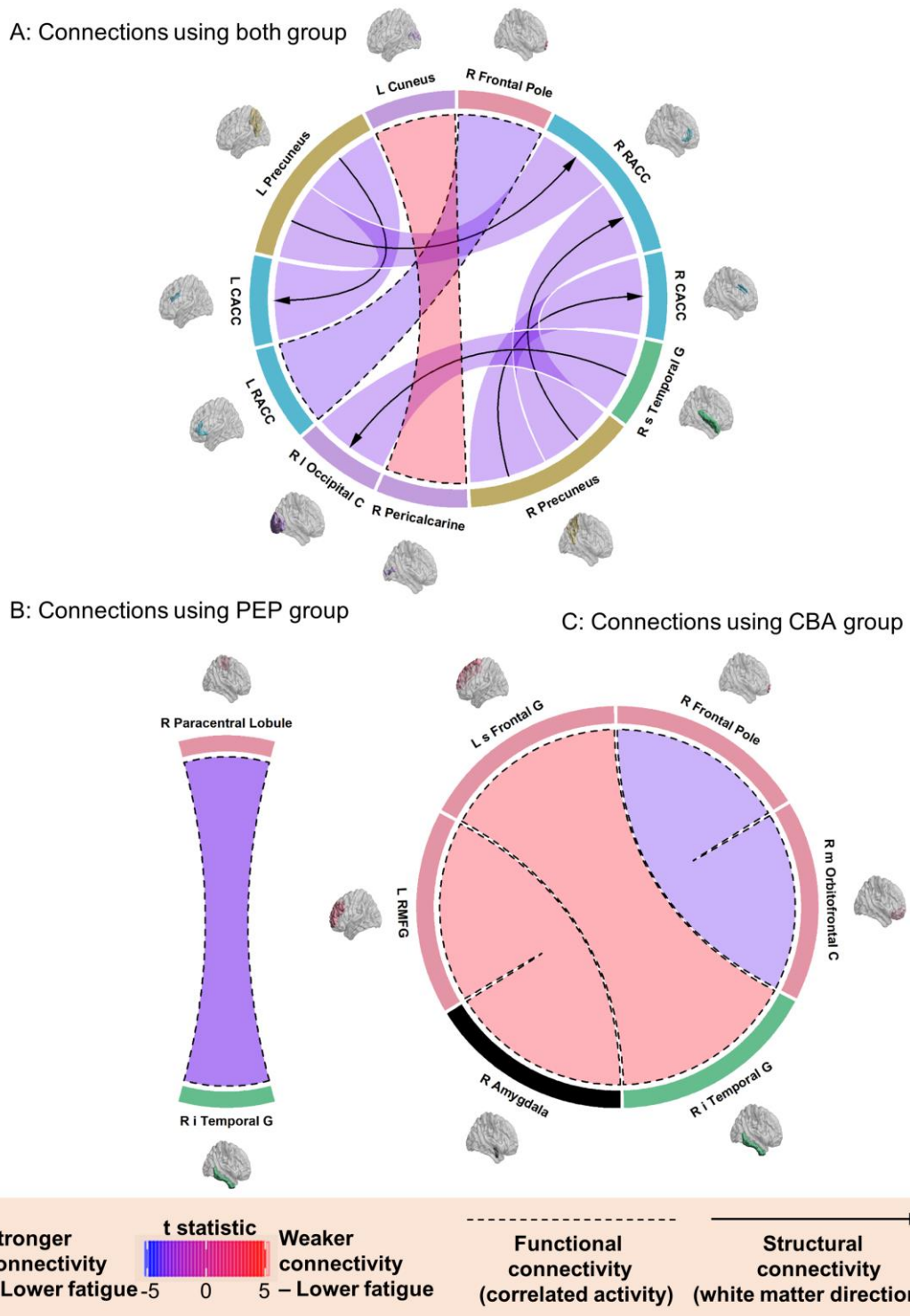


Figure 5-4: Fatigue associations within and across treatment groups. Panel A displays the five structural connectivity and two PASAT functional connectivity associations with fatigue follow-up found across the PEP and CBA groups. Panel B and C show the resting-state connectivity associations with follow-up fatigue in the PEP and CBA groups, respectively. All analyses used a general linear model, adjusted for age, gender, imaging site, and baseline fatigue. Abbreviations: L/R, left/right; i/s, inferior/superior; m/l, medial/lateral; C, cortex; G, gyrus; RACC, rostral anterior cingulate cortex; CACC, caudal anterior cingulate cortex; RMFG, rostral middle frontal gyrus. The brain images were created using BrainNet software v1.7 (Xia et al., 2013), and the diagram was constructed using circlize R package (Gu et al., 2014).

5.3.4 Results across treatment groups did not overlap between groups

After applying the agnostic approach to the treatment groups individually, the PEP group had one association between follow-up fatigue and resting-state connectivity of two regions (**Table 5-5**). These were different to those found in the CBA group which had predictors in the other modalities (**Appendix D**). Overall, the associations found when using both groups were different from those using single groups (**Figure 5-4**).

Resting-state associations in PEP group				
Seed region	Target region	t statistic	p value	Effect size
Right Inferior Temporal Gyrus	Right Paracentral	-4.14	p=0.046	0.47 (large)
Resting-state associations in CBA group				
Left Superior Frontal Gyrus	Right Inferior Temporal Gyrus	4.62	p=0.02	0.56 (large)
Right Amygdala	Left Rostral Middle Frontal Gyrus	4.53	p=0.025	0.55 (large)
Right Medial Orbitofrontal Cortex	Right Frontal Pole	-4.46	p=0.029	0.54 (large)

Table 5-5: Analysis in single treatment groups. Significant results ($p < 0.05$) using general linear model on baseline single resting-state functional connectivity variables to predict future Chalder fatigue in PEP ($n=25$) and CBA ($n=23$) groups individually. Seed region indicates brain region whose connections were tested for predicting fatigue while controlling for multiple comparisons using false discovery rate (FDR). Target region indicates which connection with the seed region was significant. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).

5.4 Discussion

This is the first study to use multimodal neuroimaging in RA patients who underwent fatigue interventions within a clinical trial. It aimed to identify metrics at baseline that link with fatigue after 6 months, which was lowered by both treatments compared to UC. Previously found markers of current fatigue did not associate with future fatigue in the trial. An agnostic approach, which searched through multiple brain regions, found five SC and two PASAT FC correlates but none using resting-state FC or grey matter volumes. In patients who underwent PEP or CBA, lower fatigue at follow-up was related to stronger physical precuneus-anterior cingulate cortex connections at baseline. Lower future fatigue was also associated with weaker (left cuneus and right pericalcarine cortex) and stronger (anterior cingulate cortex and frontal pole) correlated activity during the PASAT task. Finally, the two treatment groups differed on resting-state connectivity predictors of fatigue.

Neuroimaging modalities expectedly produced distinct results as SC and FC only moderately ($r \approx 0.3$) correlate with each other (Sadaghiani and Wirsich, 2020). This relationship strengthens with longer scanning times and/or higher integrated FC states such as during rest compared to when performing a task. Despite this stronger SC-FC relationship during rest, PASAT rather than resting-state functional data identified predictors in this study. This discrepancy may indicate the differences between the two treatment groups on predictors of fatigue, present in the resting-state data. While results from the RA study pinpoint FC as reflective of current fatigue, SC may be better equipped to predict future fatigue levels. Genova et al. (2013) observed something similar in MS as fMRI brain activity discerned task-based fatigue while DTI was associated with long-term fatigue. Within the context of RA and LIFT, white matter changes may have preceded functional ones and better reflected how susceptible patients are to clinical interventions.

In previous literature, white matter changes have preceded functional changes in MS (Koubiyr et al., 2019), major depressive disorder (Yao et al., 2019), and typical neurodevelopment (Wendelken et al., 2017) while grey matter changes have occurred last (Forouzannezhad et al., 2019). These studies observed that baseline SC rather than FC either correlated or predicted future clinical variables/cognitive abilities. This distinction may reflect differences between state and trait fatigue. State fatigue measures the current but likely temporary level of fatigue that can fluctuate during tasks, whereas trait fatigue represents the stable, overall fatigue status of patients that minimally changes over time

(Genova et al., 2013). Correlates of these two types of fatigue, both in location and modality, have been shown to differ in MS (Jameen et al., 2019) while also being the targets of brain stimulation (Linnhoff et al., 2019). Furthermore, some views have implied that functional brain changes would better represent state fatigue while psychological factors like depression and anxiety may better explain state fatigue in MS (Schreiber et al., 2015). Alternatively, FC preceded SC changes in other disorders such as attention deficit hyperactivity disorder (Bos et al., 2017) and epilepsy (Vaessen et al., 2014), however, both sequences may appear like in Alzheimer's disease (Balachandar et al., 2015, Shu et al., 2016). These inconsistencies may derive from both disease-specific mechanisms and differences in acquiring and analysing neuroimaging data (Straathof et al., 2019). Overall, although SC appeared as the best correlate of future fatigue in LIFT, previous literature limits these conclusions to the RA cohort and methodology.

Structural connectivity in LIFT revealed that low follow-up fatigue was associated with higher baseline integrity of white matter tracts from the precuneus to the anterior cingulate cortex. This negative fatigue-fibre integrity relationship complies with many studies in MS (Jameen et al., 2019). Baseline FC had a similar negative relationship between fatigue and the coherence of the frontal pole and anterior cingulate cortex but a positive one using connectivity between the left cuneus and right pericalcarine cortex. Such mixed results appear in MS and concern differences in methodology. However, these varied responses also involve how FC changes with different stages of the disease (Capone et al., 2020), which can apply to how RA fatigue progresses as well. Inadequate questionnaires/tasks may further confound results as they fail to distinguish transient perceptions from consistent feelings of fatigue (Baran et al., 2020). Overall, these technical and biological complexities hinder discoveries of consistent fatigue biomarkers not only in RA but in all chronic disorders that suffer from fatigue (Goni et al., 2018).

The design and multimodal imaging of this study addressed some of the hindrances It identified correlates of future fatigue while controlling for confounder effects because patients were followed in time. Longitudinal methods produce more reliable results but only 7 of 26 studies used this design within a systematic review of research on neural indicators of fatigue in chronic diseases (Goni et al., 2018). None of these studies integrated structural MRI, fMRI, and DTI modalities that broaden our insight into detrimental brain processes, as demonstrated in other diseases (Liu et al., 2015). Additionally, having a control group and two interventions introduced fatigue changes that better reflect the clinical setting compared to observational methods. However, the study did not manipulate brain activity to determine

if the identified neural changes can alleviate fatigue. Finally, conclusions are limited on whether we can reliably predict fatigue due to the lack of non-neuroimaging variables and the use of univariate approaches.

The results of this study were discussed in terms of which brain features predict low fatigue outcomes. This can be reversed; higher or lower baseline SC between the precuneus and the anterior cingulate cortex predicts lower or higher follow-up fatigue, respectively. Corresponding brain metrics of fatigue change through the use of change scores are in **Appendix D**. Clinically, it would be useful to identify which brain region characteristics at baseline can predict whether a patient after a fatigue-specific PEP or CBA intervention will have low fatigue. However, to create a comprehensive understanding of fatigue, future studies should incorporate additional variables. These could include biochemical, whose changes relate to fatigue in MS (Jameen et al., 2019), and genetic, as previous research established both hereditary factors (MacGregor et al., 2000) and epigenetic modifications (Liu et al., 2013) in RA. Furthermore, techniques like repetitive TMS or direct current stimulation could influence brain activity and thus elucidate pathways of fatigue induction or suppression, as applied in MS (Snow et al., 2019, Chalah et al., 2015). Finally, the use of multivariate instead of univariate analysis could construct models based on multiple variables to make reliable predictions of how fatigue will change in individual patients and what treatment would best suit them. Testing the plausibility of such a tool that combines imaging and clinical data to estimate the chances of a treatment response is subsequently the aim of **Chapter 7**.

Despite the limitations of the analysis, the results implicate SC and thus white matter integrity, as a reliable metric for predicting future fatigue, measured as a trait characteristic rather than the current state of the RA patient. FC at rest could also inform how fatigue will persist potentially because of stronger coherence with SC. Based on previous findings, FC during a task may be better suited for understanding current perceptions of fatigue. Therefore, to improve both our predictions and understanding of fatigue, clear distinctions should be made between trait and state fatigue and how they relate to each neuroimaging modality. In conclusion, although the existence of single biomarkers of fatigue is unlikely, this study demonstrated the potential of multimodal imaging to inform how a patient may react to a fatigue intervention.

Chapter 6 Predictors of fatigue in RA and PsA after pharmacological interventions

6.1 Introduction

The pathophysiologies of PsA and RA differ, yet patients with both diseases struggle with chronic fatigue (**Chapter 1.2**). Although PsA treatment involves synthetic and biological DMARDs with discrete modes of action from RA (**Chapter 1.2.7**), targeted therapies yield comparable response rates across these IA conditions. This universal response limits therapeutic efficacy as approximately 40% of IA patients do not respond to treatment and 75% cannot maintain remission (Pitzalis et al., 2020). Non-responders often experience chronic fatigue, alongside other clinical manifestations of CNS dysfunction such as depression and centralised pain. These CNS symptoms contribute similarly to poor response in both inflammatory (e.g., swollen joint counts and acute phase reactants) and subjective dimensions (e.g., patient global assessments) of IA disease activity measures (Matcham et al., 2018). Thus, the brain may regulate the response to inflammatory and subjective IA dimensions, and thereby explain the ceiling effect of existing therapies in both RA and PsA.

While brain imaging research has explored fatigue in RA (**Chapter 1.5.4**), less is known in PsA. Also, PsA and RA patients respond similarly to DMARDs, and CNS dysfunction symptoms predict poor response to treatment (Michelsen et al., 2017), but it remains unknown whether brain imaging measures can predict fatigue follow-up after such pharmacological interventions. Looking at both RA and PsA adopts a transdiagnostic approach, which seeks to assimilate the heterogeneity of symptoms within and across different disorders to answer clinically challenging questions. Transdiagnostic analyses in neuroimaging aim to find generalisable brain features rather than those that fit a specific disorder. Patients exhibit different patterns of symptoms and comorbidities rather than a single distinct pattern, which makes treating spectrum behaviours like fatigue difficult as patients frequently do not meet any set exemplar patterns. To address the challenge of fatigue in IA, clinical research would need to identify a “transdiagnostic” network of fatigue that predicts symptoms and generalises across a spectrum of clinical categories.

The transdiagnostic approach informs if symptoms rely on the same brain networks even if they range from average to subclinical to clinical and dynamically increase and decrease as the disease state changes. In the context of fatigue in IA, this would be shared predictors of

fatigue between RA and PsA. The agnostic approach used in **Chapter 5** was applied to an RA cohort and PsA cohort, which were part of “Characterising the Centralised Pain Phenotype in Chronic Rheumatic Disease” (CENTAUR) study. The analyses included finding predictors of fatigue after starting a new DMARD by agnostically searching through every neuroimaging variable within resting-state connectivity and grey matter properties. I hypothesised that:

- There will be statistically significant predictors of fatigue after starting a new DMARD in both the RA and PsA cohorts.
- Some brain regions will be shared between results in RA and PsA due to a common transdiagnostic network of fatigue, but others will be distinct due to differences between diseases in which brain metrics have the most influence on fatigue outcomes.

6.2 Methods

Patient recruitment and both clinical and MRI data descriptions for the RA and PsA cohorts have already been described in **Chapter 2.1**. Clinical characteristics of patients with PsA were recorded at baseline and 6 months after starting a new DMARD, a similar time window to the non-pharmacological approaches in RA of **Chapter 5**. Follow-up data of patients with RA was used for 3 months (n=17) after starting a new DMARD due to the low number of patients (n=7) who have completed a 6-month follow-up. However, treatment response to DMARDS generally occurs within 3 months thus allowing enough time for any changes to be comparable with a 6-month follow-up change (Smolen et al., 2023). Fatigue was measured using the PROMIS Fatigue-FM Profile (Kratz et al., 2016). This is a 16-item measure that represents fatigue experience in three subdomains—social, cognitive, and motivation (**Appendix E**). Pairwise t-tests or Wilcoxon signed rank tests (if non-normal data) determined whether treatment changed the clinical characteristics between baseline and follow-up.

T1-weighted scans quantified grey matter properties of brain regions. Functional scans from resting-state depicted FC between brain regions. The Desikan-Killiany anatomical atlas (Desikan et al., 2006) defined the brain regions for both modalities. The pre-processing of the resulting images included surface-based segmentation of grey and white matter (Dale et al., 1999) using the Free Surfer 6 software. This was followed by the default surface-based CONN pipeline for fMRI. The outputs were measures of grey matter volume, thickness, and surface area and a FC matrix, made of values for each pair of regions.

To analyse the data in a univariate way, I employed a general linear model that controls for variables of no interest (age, gender) and baseline PROMIS fatigue. The univariate model tested whether the outcome and a single feature are predictive of one another. Because each variable is modelled individually, all tests are controlled for multiple comparisons. Similarly to **Chapter 4** and **Chapter 5**, this method was implemented using a custom MATLAB code based on the “conn_glm” function of the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012) which can be seen here: <https://github.com/krisbg95/CORT.git>.

Baseline neuroimaging variables from the same brain regions were individually tested whether they predict future PROMIS fatigue. To achieve this, the covariates (mean effect, gender, age, baseline, raw outcome) were used in a design matrix within a general linear model. For grey matter properties only, TIV was also used as an additional covariate.

Because this analysis was an agnostic approach, all p values were adjusted for multiple comparisons using FDR. The significance level for the morphometric metrics and FC was set at $p < 0.05$. Additionally, predictors of baseline fatigue are also described in **Appendix F**.

6.3 Results

There were 47 patients with PsA who completed baseline neuroimaging, 24 of whom were female, with a mean (SD) age of 48.79 (11.1). The new DMARD treatments significantly reduced the disease activity of patients after six months in PsA (**Table 6-1**). Other clinical characteristics did not statistically significantly change but fatigue displayed trends of decreasing ($t(37) = 1.684, p = 0.101$), with 5 points lower score being in excess of recognised minimally important differences (Sedaghat, 2019). In comparison, there were 25 RA patients who completed baseline neuroimaging and the fatigue questionnaire, 23 of whom were female, with a mean (SD) age of 49.36 (12.51). While not enough patients completed disease activity examination at follow-up ($n=3$) to test for change, fatigue again displayed trends of lowering ($t(16) = 2.008, p = 0.062$) after the RA patients started their new pharmacological treatment. Patients who did not complete follow-up assessment in both cohorts did not differ from those who had a follow-up visit (**Table 6-2**).

Factor	Baseline RA	Three months RA	Baseline PsA	Six months PsA
Disease activity** a	5.77 (0.86)	NA	35.7 (23.45)	21.8 (24.7)**
CRP (mg/L)	0.9 (1.5)	NA	0.4 (0.57)	0.3 (0.3)
Fatigue ^b	51 (16)	44 (23)	53 (26.5)	48 (24.5)
Sleep ^c	24 (9)	19 (7)	24 (10)	24 (10.75)
Fibromyalgianess ^d	10 (6)	8 (4)	12 (8.5)	11 (7.5)
Depression ^e	17 (11.25)	16 (15)	18 (13.5)	16 (12.75)

Table 6-1: Clinical characteristics of patients with rheumatoid and psoriatic arthritis as Median (IQR). The disease activity of psoriatic patients was significantly lower six months after starting a new DMARD treatment ($W=381, p<0.001$, rank biserial effect size $r_B = 0.75$). ^aDAPSA (disease activity in psoriatic arthritis), DAS28 (disease activity in rheumatoid arthritis); ^bPROMIS FatigueFM Profile 16-item; ^cPROMIS Sleep Related Impairment Short form 8a; ^d2011 Fibromyalgia survey criteria; ^ePROMIS Emotional Distress – Depression Short form 8a; ** $p < 0.001$.

Factor	RA Complete	RA Missing	PsA Complete	PsA Missing
Sample size	17	8	39	8
Gender (female)	94%	75%	49%	62%
Age	51 (19)	51 (21)	49 (14)	40.5 (13.5)
Disease activity ^a	5.96 (0.92)	5.53 (0.53)	35.7 (25.75)	31.65 (11.65)
CRP (mg/L)	0.9 (1.4)	0.6 (1.15)	0.3 (0.45)	0.67 (0.38)
PROMIS Fatigue ^b	51 (9)	48 (23.75)	53 (27)	61 (14.25)

Table 6-2: Comparison between patients with completed and missing follow-up data in cohorts starting a new DMARD treatment as Median (IQR). The 8 RA patients with missing follow-up did not significantly differ (Mann-Whitney U test) from the 17 patients who completed follow-up on age, gender, disease activity, PROMIS Fatigue, and inflammation in the form of C reactive protein (CRP). ^aDAPSA (disease activity in psoriatic arthritis), DAS28 (disease activity in rheumatoid arthritis); ^bPROMIS FatigueFM Profile 16-item.

Both structural and functional brain metrics predicted PROMIS fatigue 6 months after starting a new DMARD treatment in PsA (**Table 6-3**), visualised in **Figure 6-1** but failed to do so in RA three months after starting a new DMARD treatment. Specifically, higher grey matter thickness of the visual right and left pericalcarine predicted higher fatigue at follow-up. Higher connectivity between the visual lingual gyrus and the right caudal anterior cingulate cortex also predicted higher fatigue at follow-up as well as higher connectivity between the left inferior parietal lobule and the left bank of the superior temporal sulcus.

Grey Matter Thickness

ROI Seed	t statistic	p value	Effect size
Right Pericalcarine	4.01	0.022	0.33 (medium)
Left Pericalcarine	3.6	0.035	0.28 (medium)

Resting-state Functional Connectivity

ROI Seed	ROI Target	t statistic	p value	Effect size
Left Bank of the Superior Temporal Sulcus	Left Inferior Parietal Lobule	3.87	0.039	0.31 (medium)
Left Lingual	Right Caudal Anterior Cingulate	3.65	0.024	0.28 (medium)

Table 6-3: Neuroimaging predictors of fatigue after pharmacological interventions in psoriatic arthritis. Significant results ($p < 0.05$) using general linear models on baseline grey matter volume, thickness, surface area and resting-state single functional connectivity variables to predict future PROMIS fatigue in patients after six months of starting a new disease-modifying antirheumatic drug ($n=39$). Seed region indicates brain region whose grey matter properties or connections were tested for predicting fatigue while controlling for multiple comparisons using false discovery rate. Target region indicates which connection with the seed region was significant. The results are after controlling for confounds (age, gender, baseline fatigue) for connectivity and total intracranial volume for grey matter measures. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019). Abbreviations: ROI, region of interest.

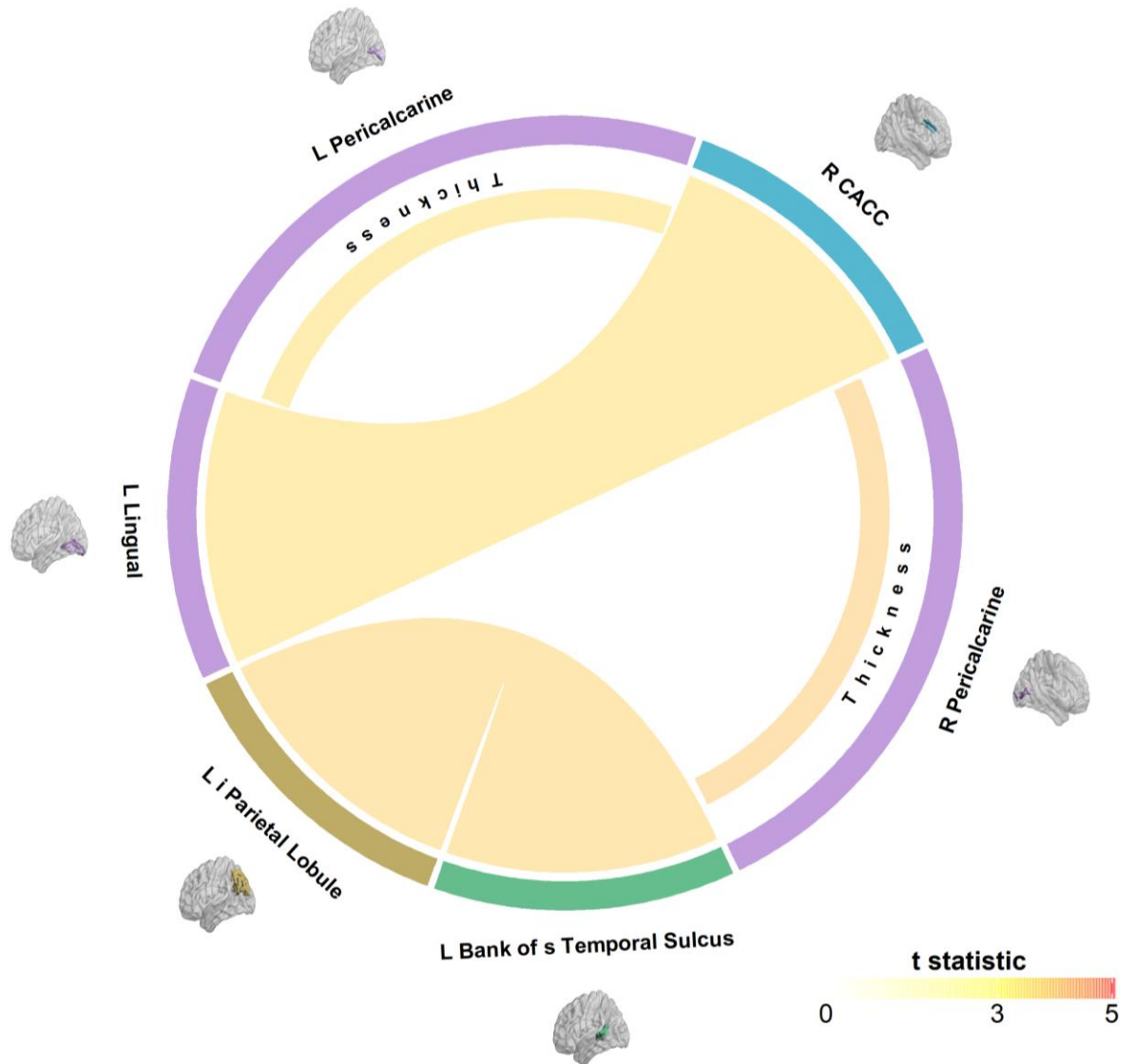


Figure 6-1: Neuroimaging predictors of fatigue after pharmacological interventions in psoriatic arthritis. The figure displays the resting-state connectivity and grey matter thickness predictors of PROMIS fatigue 6 months after starting a new disease-modifying antirheumatic drug. All predictors were positively associated with fatigue, while the strength of the association (t statistic) is reflected in the colouring of the connection or semicircle of the thickness measure. All analyses used a general linear model, adjusted for age, gender, and baseline fatigue while the structural measures were additionally corrected for intracranial volume. Abbreviations: L/R, left/right; CACC, caudal anterior cingulate cortex.

6.4 Discussion

In this study, higher thickness of the bilateral pericalcarine cortex (primary visual cortex) predicted greater fatigue 6 months after starting a new DMARD in PsA. Higher pericalcarine thickness distinguishes individuals with fibromyalgia, a centrally driven pain disorder, from healthy controls (Nhu et al., 2023, Aster et al., 2022). Altered structural pericalcarine properties may indicate dysfunctional CNS processing, potentially explaining why conventional DMARDs did not mitigate fatigue in PsA patients within this study. Both PsA and fibromyalgia patients endure chronic fatigue, and fibromyalgia's hallmark of CNS involvement—widespread pain—alongside comorbid fibromyalgia are predictors of severe fatigue in IA (Overman et al., 2016). Compared to healthy controls, fibromyalgia patients exhibit reduced resting-state fMRI connectivity of the primary visual cortex, correlating with self-reported pain (Pujol et al., 2014) and resiliency towards pain (Flodin et al., 2014). Other aspects of CNS involvement, such as low mood and cognitive deficits, relate to fatigue in PsA (Sumpton et al., 2020). These symptoms also associate with the pericalcarine cortex. Depression severity negatively correlates with the thickness of the right pericalcarine cortex in fibromyalgia (Jensen et al., 2013), and reduced microstructural integrity of the pericalcarine cortex differentiates cognitively impaired from cognitively normal MS patients (Pitteri et al., 2021). Cognitive impairment may arise from the role of the pericalcarine cortex in attention, as evidenced by a study that negatively correlated SC between the right pericalcarine cortex and the left pallidum with the visual attention task performance in MS (Llufriu et al., 2017). In summary, changes in the pericalcarine cortex may predict fatigue by serving as an indicator of CNS involvement and its associated symptoms.

Elevated fatigue at follow-up in PsA was associated with increased FC between the right caudal anterior cingulate cortex and the left lingual gyrus, and between the left inferior parietal lobule and the left bank of the superior temporal sulcus. The anterior cingulate cortex, in conjunction with the insula, constitutes the salience network responsible for evaluating the relevance of internal and external stimuli (Menon and Uddin, 2010). Lopez-Sola et al. (2014) showed augmented activity of the insula and the lingual gyrus following non-nociceptive sensory stimulation (auditory, visual, and tactile) among 35 fibromyalgia patients, who also displayed multisensory hypersensitivity compared to 25 healthy controls. In the current study, enhanced connectivity between the anterior cingulate cortex and lingual gyrus predicted fatigue potentially because it acted as a biomarker of multisensory hypersensitivity. Fibromyalgia pain implicates central sensitization, amplified responses to

sensory stimuli due to altered processing and/or rewiring of the CNS (Harte et al., 2018). However, hypersensitivity to nonpainful stimuli circumvents peripheral and spinal pain pathways, signifying a form of top-down central sensitization that, unlike bottom-up mechanisms, persists independently of nociceptive input (Harte et al., 2018). As evidence exists of central sensitization in chronic pain and chronic fatigue conditions (Meeus and Nijs, 2007), top-down mechanisms of central sensitization may have predisposed PsA patients to fatigue in this study, independent of peripheral nociception. The other key finding that future fatigue was predicted by hyperconnectivity between the inferior parietal lobule and the superior temporal sulcus, both part of the DMN (Uddin et al., 2019), suggests that DMARDs may have failed to address aberrant connectivity within this network. Enhanced DMN FC has been shown to drive fatigue in MS (Bisecco et al., 2018). FC analyses have also demonstrated that symptoms of fatigue and depression reflect altered DMN connectivity and that higher DMN activity is observed in MS patients with fatigue, even with low depression scores (Hogestol et al., 2019). Finally, the DMN may encapsulate general mechanisms of fatigue, as subjective mental fatigue in a healthy cohort correlated with increased on-task activity and resting-state FC within the DMN (Gergelyfi et al., 2021).

The neuroimaging variables did not predict fatigue post-DMARD treatment in the RA cohort. This could stem from the limited number of RA patients compared to those with PsA in the study, especially those who completed follow-up. Small samples (≤ 20) struggle to show brain-behaviour associations due to their effect sizes, ranging from very small (Cohen's $d \approx 0.15$) in healthy subjects (Marek et al., 2022) to double in psychiatric and neurological patients (Cohen's $d \approx 0.32$), yet still considered small effects (Libedinsky et al., 2022). These limited effects result from factors such as the brain's complex structure and function, the inverse relationship between spatial and temporal resolution in neuroimaging, the high dimensionality of even extracted features like connectivity matrices, and the variability across subjects and studies (Zhu et al., 2023). The brain is affected not only by disease but also by age, genes, race, and environmental factors. Additionally, scanner effects, acquisition parameters, and motion or thermal noise impact the neuroimaging signal. Moreover, different analysis pipelines can yield diverse findings, exemplified by an article where 70 independent neuroimaging teams varied in half of their results when testing the same hypotheses with identical neuroimaging data (Botvinik-Nezer et al., 2020). Nevertheless, standard GLM analyses, as applied here, have shown high test-retest reliability of subject-specific results (Brown et al., 2011, Friedman et al., 2008). Despite the small sample, this study aimed to optimize reliability and statistical power through subject-space processing and employing the Desikan-Killiany atlas. Subject-level analyses preserve fine-

grained anatomical details often lost in spatial normalization, thus better maintaining individual variability (Magalhaes et al., 2015). The Desikan-Killiany atlas, a relatively coarse brain parcellation, reduces parallel tests and minimizes variance due to methodological error (Helweggen et al., 2023).

Overall, predictors of PsA fatigue point to CNS involvement akin to conditions like fibromyalgia and MS, along with centralised symptoms like depression and cognitive impairment. Structural and functional findings focus on default mode and salience networks, implicated in hypersensitivity to non-painful stimuli and interoception. These associations suggest a potential role of top-down mechanisms, typically linked with centralised pain, in predisposing and sustaining fatigue in PsA. Although these results were discussed in terms of which brain features predict high fatigue outcomes, the reverse can also be applied such as having higher or lower baseline thickness of the pericalcarine cortex predicted higher or lower fatigue follow-up, respectively. Additionally, identical analyses that identify brain metrics of fatigue change by using change scores are in **Appendix F**. In the context of starting a new DMARD to reduce disease activity, identifying which brain region characteristics at baseline predict persistent high fatigue can guide early implementation of non-pharmacological fatigue-specific therapies in conjunction with pharmacological treatments. In comparison to PsA, imaging variables did not predict fatigue after starting a new DMARD in the RA cohort and thus failed to support the hypothesis of a common transdiagnostic network of fatigue. However, the lower number of RA patients, particularly those with follow-up fatigue data, limited the ability to delineate a transdiagnostic fatigue network within the RA and PsA cohorts. Future transdiagnostic studies would need to record longitudinal neuroimaging data to pinpoint specific CNS mechanisms of IA fatigue, probe pharmacological bottom-up and cognitive top-down interventions and assess the impact of factors such as centralised pain, low mood, and cognitive deficits that contribute to fatigue.

Chapter 7 Prognostic models of fatigue of non-pharmacological interventions in RA

7.1 Introduction

In the previous chapters, GLMs identified univariate predictors of fatigue that reflect associations between neuroimaging variables and treatment outcomes in RA and PsA. While univariate models aid in predicting treatment response, they cannot generalise to individual patients. Their main benefit lies in the interpretability of their features. For example, univariate analyses have highlighted connectivity changes in depression and linked these changes to different symptoms. Anhedonia demonstrates stronger associations with the reward network, whereas rumination is more tied with the DMN, suggesting how patients may differentially respond to treatments (Chahal et al., 2020, Li et al., 2018). Univariate findings can thus guide treatment decisions as well as inform the development of new treatments by uncovering why certain treatments may prove ineffective for specific individuals. Despite advancing our understanding of neural mechanisms, univariate brain activation and connectivity markers on their own fail to predict individual treatment outcomes accurately and consistently in complex symptoms like fatigue and depression. These limitations arise because, unlike univariate studies that build and evaluate their models within the same sample, studies that aim to predict individual outcomes incorporate validation techniques (e.g., cross-validation or external validation), where model evaluation uses a different sample from model building. Validation techniques enhance generalisability, but prediction models also require high accuracy. One way to bolster accuracy is to increase the number of predictors, an aspect that ML offers (**Chapter 1.7**). In the context of RA, this chapter will aim to predict individual fatigue treatment responses using ML to incorporate multiple prognostic factors, including immune, neuroimaging, and symptom-based information.

Often studies predict treatment response by categorising the outcome, such as defining responders as those with a $\geq 50\%$ reduction in symptoms. However, this dichotomisation of continuous variables strips out information and subsequently reduces the power to detect predictor-outcome associations (Altman and Royston, 2006). MacCallum et al. (2002) demonstrated this limitation by splitting observations using the median of a variable to “high” and “low” groups, which resulted in reducing power by the same amount as discarding 36% of observations. In addition to diminishing statistical power, categorisation

neglects the nuanced variability within groups. For instance, a patient experiencing a 49% symptom reduction would be in the same bracket of “non-responders” as one with no reduction, despite being much closer to a “responder” with a 50% reduction. Dichotomisation also conceals any non-linearity between the predictor and outcome, diminishing a potential strength of ML to learn non-linear relationships. Collectively, these limitations of categorisation can misinform which features are predictive (Austin and Brunner, 2004) and reduce the accuracy of predictions but can be addressed by regression models. A study compared classification to regression using fMRI data from 82 young adults seeking treatment for psychological distress and 72 matched healthy controls (Portugal et al., 2019). Specifically, they demonstrated that using a similar ML approach, regression models predicted depression and anxiety scores significantly above chance unlike classification models, which failed to differentiate healthy from distressed individuals. The study demonstrated that regression algorithms exhibit greater sensitivity than classification approaches in identifying relationships between continuous symptoms and neural measures within heterogeneous samples, also applicable in the context of RA fatigue.

Studies typically construct models that predict response to a single treatment. In clinical practice, a more valuable prediction would involve differential treatment response, determining which among several interventions is most likely to improve a patient's condition rather than only whether an individual should receive an intervention (Perlman et al., 2019). Devising a differential prediction could comprise developing individual models for each treatment and then aggregating their outputs. However, this process can become significantly more time-consuming and resource-intensive when different models rely on distinct data modalities and require separate validations. A far more practical solution would encompass a single model that can consistently predict outcomes across multiple treatment options. To provide differential predictions, studies have used ML models with neuroimaging biomarkers to select CBT or an antidepressant (McGrath et al., 2013) or the optimal combination of antidepressants (Williams et al., 2015). Ideally, studies of RA fatigue would establish predictions of differential treatment response, encompassing a broader range of treatments, including exercise, different DMARDs, centrally acting drugs like pregabalin, as well as CBT. Finally, it would be more advantageous when studying more heterogeneous clinical populations to compare different types of treatments than comparing very few treatments with similar mechanisms of action.

Based on the discussed points above I plan to: (1) predict treatment response with high accuracy that generalises to individual patients by using ML and validation techniques rather

than GLMs; (2) address the dimensionality of fatigue in heterogeneous populations like RA, by using regression instead of classification; and (3) develop a more clinically practical model by building a single model that can consistently predict outcomes across different types of interventions. Specifically, I will predict individual fatigue scores in RA patients six months and one year after receiving either usual care (UC), a personalised exercise programme (PEP), or a cognitive-behavioural approach (CBA) from the LIFT trial, using ML regression models. Single modality models will use brain morphometric, SC or FC data separately as well as clinical data. A multimodal model will also integrate all data modalities, including treatment allocation. I will then externally validate the best-performing model in the RA study cohort that had only usual care to test the reproducibility of the model. I hypothesise that:

- The multivariate model will perform better than chance through permutation testing.
- The neuroimaging modalities will perform differently from each other.
- The treatment allocation will have high discriminative power on performance based on feature importance.
- The external validation in the first RA cohort will show similar performance to the patients who had usual care in the internal validation RA cohort.

7.2 Methods

7.2.1 Datasets and modalities

The datasets used in this analysis included the LIFT trial and the RA study, described in **Chapter 2.1**. The model was built and internally evaluated using the LIFT dataset, while it was externally evaluated on the RA study dataset. The LIFT data included clinical variables along with the neuroimaging modalities of morphometric features, SC, and FC from the resting-state or the PASAT task. The neuroimaging variables were extracted using the same preprocessing procedure described in **Chapter 2** and **Chapter 4**. The same preprocessing steps were conducted in the RA study dataset. Overall, the features that were used as input to feature selection included:

Clinical (18 features): treatment allocation, age, gender, MRI site, physical activity, CRP, ESR, haemoglobin levels, DAS28, disease duration, Charlson comorbidity index, ACR Fibromyalgia scale, HADS depression, HADS anxiety, current pain NRS, sleep disturbance, baseline Chalder fatigue. These variables were selected based on previous predictive factors of fatigue (**Chapter 1.3.4**) like sickness behaviours (depression, anxiety, pain, sleep, fibromyalgia), inflammation (CRP, ESR), disease activity and confounding factors (DAS28, disease duration, comorbidity) and sociodemographic factors (age, gender, physical activity).

Morphometric (220 features): grey matter volume (84 ROIs), grey matter thickness (68 ROIs), grey matter surface area (68 ROIs). Thickness and surface area could be extracted only from 68 cortical regions while volumetric measurements additionally included 16 subcortical structures. The TIV was regressed out of the volumetric features within each cross-validation fold to control for the effects of brain volume, as suggested when running ML models on structural MRI data (Dhamala et al., 2023b).

Structural connectivity (6972 features): asymmetrical matrix between 84x84 ROIs, excluding self-connections. Structural connectivity is asymmetrical as connections between two regions can be different across the two directions.

Functional connectivity (3486): upper triangle of symmetrical matrix between 84x84 ROIs, excluding self-connections. There is no directionality in functional connections so only the upper triangle of the matrix is used. The LIFT dataset has functional MRI recorded during

both rest and the PASAT task, but the RA study only has functional MRI data during the PASAT task. Analysis was alternatively conducted with either PASAT or resting-state FC in the LIFT data set instead of combining both modalities.

7.2.2 Feature selection

Feature selection was implemented using two filter methods of a Relief algorithm and a canonical correlation. Before these steps, a filter was applied to discard any features with near zero variance and over 25% missing data. The first feature selection method included the regression ReliefF algorithm, which was run using the “attrEval” function of the CORElearn R package (Robnik-Sikonja et al., 2013). It was setup using a set number of neighbours using the total number of observations (m): $k = 0.154 \times (m - 1)$. The 0.154 pre-factor yields an approximation to a fixed radius that contains neighbours within a half standard deviation of a sample’s radius in the attribute space (Le et al., 2019). This value has been shown to provide a good balance for detecting main effects and interaction effects. Because the ReliefF algorithm was used as a screener, the selected number of features was separately tested as all positively ranked features as well as the top half or top quarter of positively ranked features rather than a sparser number of features (Windle, 2016). Since features in this study like FC are typically collinear, not addressing redundancies can be a potential advantage. If reducing redundancy is a concern in a specific problem, there are various methods that can be applied before, after, or integrated with Relief feature selection. This was achieved by first running Relief and then running sparse generalized canonical correlation analysis (SGCCA).

CCA maximizes the correlation between variables, as shown in **Chapter 3.2.3**. However, it cannot be applied in cases where features outnumber samples, and it only relates to two sets of variables. SGCCA addresses these limitations and acts as a component-based method for exploring data organised in variable blocks and has already been shown to identify brain-behaviour associations in high-dimensional data (Ing et al., 2019). Here, the neuroimaging and clinical data naturally fall into distinct blocks, each recorder from the LIFT RA patients. Rather than operating sequentially on parts of the measurements, SGCCA aims to encapsulate complementary information within and between these blocks (Garali et al., 2018). It leverages a priori knowledge to establish connections, aligning with biological hypotheses. SGCCA condenses the variable blocks by finding, for each block, a weighted composite of variables (called block component) that explains well their own block and/or block components assumed to be connected. The process involves a design matrix that

defines which blocks are supposed to be linked to one another, thus reflecting hypotheses about the biology underlying the data blocks. Here, the design was set so that instead of optimizing the associations between different neuroimaging measures, only the associations between the outcome and the neuroimaging measures were optimized. SGCCA standardizes the block weights vectors and utilizes the covariance between blocks, which stabilises the component (large variance) while simultaneously accounting for the correlations between blocks. Its sparsity component functions as a variable selection mechanism, identifying pertinent features. It detects significant variables within blocks that actively participate in inter-block relationships, using a penalty akin to lasso regression's regularization (**Chapter 2.4.3**).

While SGCCA has advantages over classical CCA in handling multiple variable sets with more features than observations, it may face instabilities due to the lasso penalty's introduction of sparsity. Stability selection addresses this concern through the repeated application of a model over multiple resamples, where the features with actual effects are consistently chosen, distinguishing them from mere noise (Meinshausen and Bühlmann, 2010). Here, stability selection divided the data into random subsamples, each half the size of the total training set, and performed SGCCA a hundred times on each resample. Features that emerged more frequently were regarded as more stable, similar to the approach taken by Ing et al. (2019). This process was parallelized by concurrently re-applying SGCCA to multiple resamples of the same data. The SGCCA analysis was run using the “sgcca” function in the RGCCA package in R (Girka et al., 2023). Notably, stability selection is insensitive to tuning parameters, such as the lasso penalty in SGCCA. The parameter was therefore set based on the number of features for each set of variables using the “tau.estimate” function also implemented in the RGCCA package, which uses a formula for estimating the optimal shrinkage parameter (Schafer and Strimmer, 2005). Determining feature stability necessitates deciding what proportion of resamples variables must be present in to be deemed stable. In this study, variables needed to be present in either 50% or 70% of resamples to be considered stable. While the specific threshold for retaining variables can vary, stability selection's effectiveness has been demonstrated across a reasonable range of thresholds (Meinshausen and Bühlmann, 2010), like the 50-70% range utilized in this study.

Combining all modalities into a single feature set may seem intuitive, but it's problematic due to their varying feature counts. For example, SC data has 6972 features, while morphometric and clinical data have only 220 and 18 features respectively. Merging modalities would thus result in SC overpowering the others. Therefore, the Relief algorithm

was applied separately to each modality while the SGCCA benefited from having separate blocks of clinical, morphometric, SC and FC features. Only after completing Relief and SGCCA, were the remaining features collapsed into a unified set. Before the next step of parameter tuning, these features were normalised by subtracting the mean and then dividing the standard deviation of the values of every feature column. Applying this transformation prevents inconsistencies when establishing feature associations by removing data redundancy (Sree and Bindu, 2018). To prevent overfitting, feature selection and parameter tuning occurred within the inner loop of a nested 10-fold cross-validation. This process was then repeated 100 times through random splits of the data (**Figure 2-15**). The repeated cross-validation procedure was conducted using alternative options for Relief and stability selection, PASAT or resting-state modality for FC, and clinical outcomes of either six months or one-year follow-up (**Table 7-1**).

Run	Relief proportion	Stability selection	Functional modality	Clinical outcome
1	0.25	50	PASAT	6 months
2	0.5	50	PASAT	6 months
3	1	50	PASAT	6 months
4	0.25	70	PASAT	6 months
5	0.5	70	PASAT	6 months
6	1	70	PASAT	6 months
7	0.25	50	Resting-state	6 months
8	0.5	50	Resting-state	6 months
9	1	50	Resting-state	6 months
10	0.25	70	Resting-state	6 months
11	0.5	70	Resting-state	6 months
12	1	70	Resting-state	6 months
13	0.25	50	PASAT	1 year
14	0.5	50	PASAT	1 year
15	1	50	PASAT	1 year
16	0.25	70	PASAT	1 year
17	0.5	70	PASAT	1 year
18	1	70	PASAT	1 year
19	0.25	50	Resting-state	1 year
20	0.5	50	Resting-state	1 year
21	1	50	Resting-state	1 year
22	0.25	70	Resting-state	1 year
23	0.5	70	Resting-state	1 year
24	1	70	Resting-state	1 year

Table 7-1: Alternative runs of cross-validation. The table displays the 24 different runs of repeated 10-fold cross-validation for the multivariate (all modalities) and functional connectivity models. This

included using all positively ranked features from running Relief and using the top half (0.5) or top quarter (0.25) of those features. The stability selection retained features that were in at least 50% or 70% of all resamples while the clinical outcomes could be either six months or one-year follow-up. For single modality models using separately the clinical, morphometric, or structural connectivity measures, only 12 runs were run since the functional modality settings were not applicable.

7.2.3 Algorithm and parameter tuning

After feature selection, the unified set of features was then used to build models using a number of algorithms with different hyperparameters. To fulfil the research aim of predicting individual fatigue scores in RA patients after receiving usual care or fatigue-specific treatments, I opted for regression models that: (1) leveraged fatigue-associated brain features, (2) integrated treatment allocation and clinical data, and (3) ranked the variables according to importance to the predictions. These models had distinct capacities to discern both linear and non-linear relationships (**Chapter 2.4.3**), each with its advantages and disadvantages (**Table 2-4**). Furthermore, similar algorithms have demonstrated efficacy in using brain imaging to predict age (Sone and Beheshti, 2022) and depression (Lee et al., 2022), as well as using clinical data to predict outcomes in IA (Madrid-Garcia et al., 2023). The models included the following:

- Elastic net (linear): a linear regression with an imposed penalty to reduce model complexity and decrease the risk of overfitting, making it more interpretable and suitable for small datasets.
- Support vector regression: a model that projects data in multidimensional space to find the best hyperplane to predict outcomes which makes it robust to outliers and produce low generalisation error.
 - Linear kernel
 - Radial (non-linear) kernel
- Gaussian process regression: a model that updates the probability of an outcome given new data, which makes it interpretable and suitable for small datasets
 - Linear kernel
 - Radial (non-linear) kernel
- Random forest (linear and non-linear): a model that combines the outputs of many “if-then-else” tree models that can handle collinearity between features.
- Gradient Boosting regression (linear and non-linear)
 - gbm: similar to random forest but where each subsequent tree model focuses on minimizing the errors made by the previous models.

- xgbDART: incorporates dropout regularization to gbm, in which it randomly drops a subset of trees during each iteration, which helps to prevent overfitting and can lead to better generalization but requires more tuning due to additional hyperparameters

All of the model algorithms were implemented using the caret R package (Kuhn, 2008) and structured within a single framework using the modelgrid R package (Kjeldgaard, 2018). The models are applied to each training set and then evaluated on the subsequent test set. This is done for each possible value within a grid of hyperparameters, displayed in **Table 7-2**. The selected features and parameters with the the minimum difference between train and test performance, are then conveyed to the outer loop of the cross-validation.

Algorithm	Hyperparameters	Rationale and sources
Elastic net	alpha = weighting between ridge (0) and lasso (1) penalties: 0.1-1 (0.1 increments) lambda = weighting of the sum of both penalties: 0.002 0.005 0.011 0.026 0.06 0.139 0.322 0.744 1.72	Based on a reasonable parameter length of 10, using the default values from Kuhn (2008).
Support vector regression	C = penalisation of samples within the margins: 1e-07, 1e-06, 1e-05, 1e-04, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 2000 gamma = range of influence of kernel over observations (radial kernel): a sequence from 1e-04 to 0.015 with a total length of 20	The default range for C and gamma from caret was a single value, so a range was chosen based on Thakur (2020)
Gaussian process regression	gamma = range of influence of kernel over observations (radial kernel): a sequence from 1e-04 to 0.015 with a total length of 20	The default range for gamma from caret was a single value, so a range was chosen based on Thakur (2020)
Random forest	Randomly Selected Predictors: a sequence from 1 to one third of the total number of features in the training set with a total length of 10	A suggested value for regression problems by Kuhn (2008)
Gradient boosting	Boosting iterations: 50-500 (50 increments) Maximum tree depth: 1:10 (1 increments) Shrinkage: 0.1 Minimum terminal node size (only gbm): 10 <u>Only in xgbDART</u> Minimum loss reduction: 0 Subsample percentage: 0.5 Subsample ratio of columns: 0.6, 0.8 Fraction of trees dropped: 0.01, 0.5 Probability of skipping drop-out: 0.05, 0.95 Minimum sum of instance weight: 1	Based on a reasonable parameter length of 10 for gbm and 1 for xgbDART using the default values from Kuhn (2008).

Table 7-2: Machine learning algorithms and hyperparameter search. The table notes the specific hyperparameter range for each algorithm used in this study. The last column describes the reason for choosing that range and the supporting source.

7.2.4 Imputation of missing values

Missing data inflates standard errors due to a smaller sample size but doesn't necessarily introduce bias (Sterne et al., 2009). An incomplete dataset can still represent the entire dataset but assuming no bias of missing data can be unrealistic in clinical datasets. Most ML algorithms struggle with training data containing missing values as they cannot deal with them directly (Lo Vercio et al., 2020). Instead, imputation methods can fill in the missing values. A common approach is replacing missing values with the feature's average or median. This is a form of single imputation, where missing values are replaced by a specific rule. Variants include 'last observation carried forward' and 'worst observation carried forward,' as well as simple mean imputation. Single imputations can also be model-based, such as the K-Nearest Neighbour method, which similar to Relief uses a distance function. The method randomly selects values from the k nearest similar cases, using the closest one to replace the missing value (Zhang, 2012). When the data distribution is unknown, it is suitable to use K-Nearest Neighbour (Joel et al., 2022). However, single imputation methods often underestimate variability because unobserved values weigh equally in the analysis (Dziura et al., 2013). Single imputations assume specific conditions, like assuming missing values are identical to the last observed value. These assumptions are often unrealistic, making single imputation a potentially biased method.

A second option involves determining the range or actual distribution of available values for a missing data point, and then replacing it with a randomly sampled value from this distribution. A more advanced method for this is multiple imputation, which produces several imputed values for each missing data point, incorporating uncertainty into the estimates (Azur et al., 2011). It detects missing values and substitutes them with a randomly selected set of plausible values using all of the other variables as predictors. This process generates multiple completed datasets using a chosen imputation model. Subsequently, the prediction model is separately constructed and executed for each dataset produced during the imputation phase. The results obtained from each completed data analysis are agglomerated into multiple outputs that can better represent the uncertainty of the imputations. ML models can also be trained on the complete samples to predict single missing feature values, such as random forest (Stekhoven and Buehlmann, 2012). Lastly, ML algorithms can serve as imputation models within a multiple imputation framework,

particularly beneficial for handling higher-dimensional data with non-linearities (Feng et al., 2021). Multiple imputation methods require more computational time than single imputation methods, especially using multiple predictor variables and ML algorithms as imputation models.

If possible, removing incomplete observations offers a straightforward approach to handling missing data. However, this can severely diminish the sample size and statistical power as well as introduce bias if the values are not missing at random (Demissie et al., 2003). This complete case analysis is effective when missing data is negligible such as no more than 5%, according to general guidelines. It's also applicable when only the dependent variable has missing values and we haven't identified auxiliary variables—features not considered in the analysis but correlated with a variable exhibiting missing values and/or linked to its absence (Jakobsen et al., 2017). In such cases, techniques like multiple imputation may not provide significant advantages but only inflate standard errors due to the uncertainty introduced by the imputation method. If the outcomes are missing within a cross-validation framework, you can train an imputation model exclusively on the training set and then impute the missing features in both the training and test sets (Mertens et al., 2020). Specifically, for each fold you perform multiple imputations separately for the predictor variables and the outcome variable, but the imputation process remains blind to the testing set. This ensures that information from the training set does not affect the test set and thus avoids double dipping but also maintains the sample size and statistical power of the full dataset (Jaeger et al., 2020).

In this study, the K-Nearest Neighbour method imputed missing clinical predictors before feature selection and hyperparameter tuning, using only cases with complete outcome data within the inner loop of the cross-validation. A single imputation method was used to avoid unnecessary computation time as the inner loop prioritized a solution that minimizes overfitting over one that seeks the best predictive performance. For missing values in both predictors and outcomes in the outer loop of the cross-validation, multiple imputation was used. Imputation models were constructed solely with the training set observations and then applied to impute missing values in both the training and testing sets, as illustrated in Figure. Predictive models were then built on the training set and evaluated against ten different imputed testing sets to account for imputation uncertainty. This process was conducted through the mice R package (Van Buuren and Groothuis-Oudshoorn, 2011) using random forest as its underlying imputation method to accommodate the high-dimensional input data. Each out of ten imputed datasets was completed using 20 iterations to attain stable results

(Van Buuren, 2018), while the range of outcome values was restricted to the range of the Chalder fatigue scale of 0 to 33.

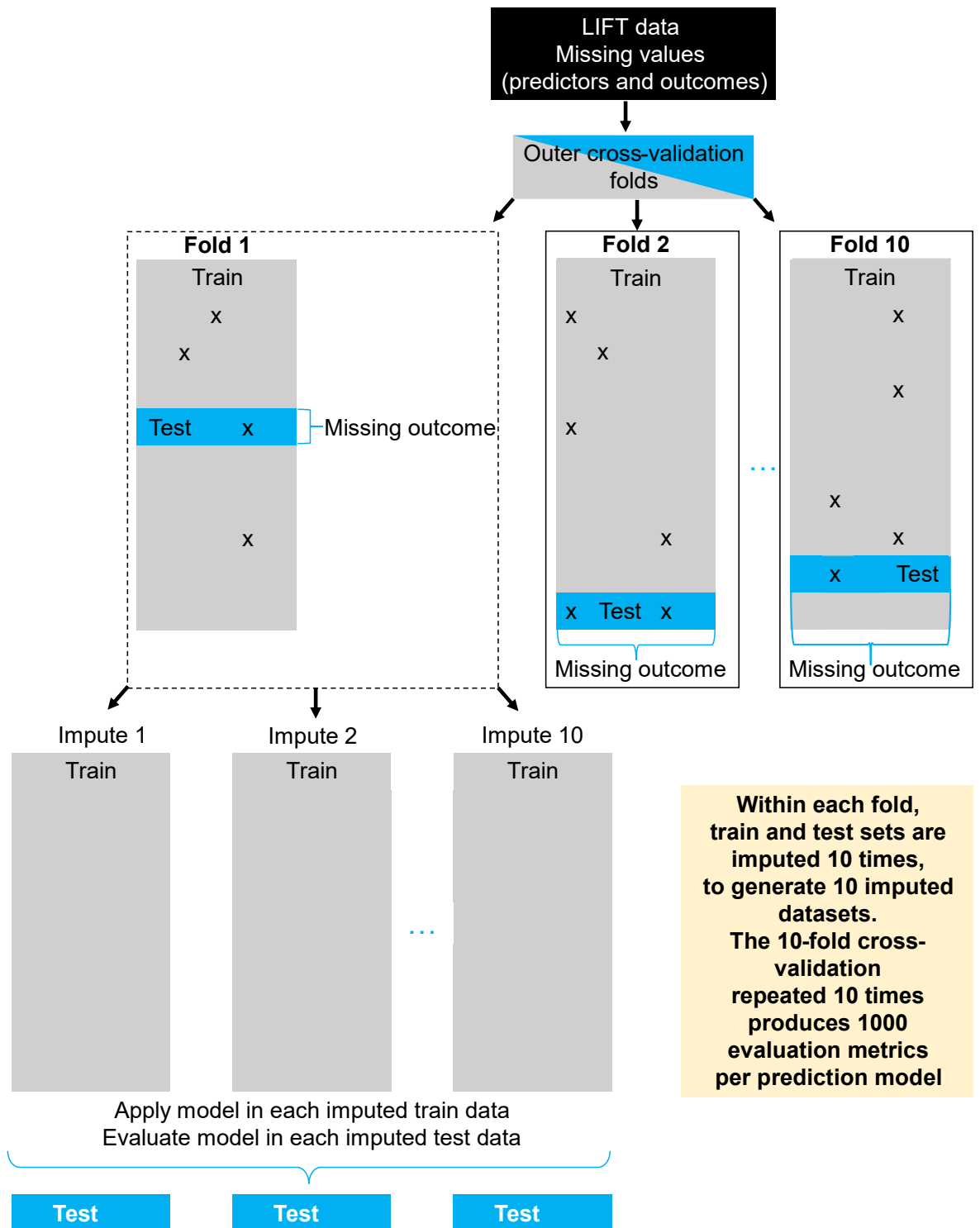


Figure 7-1: Framework of multiple imputation within a repeated 10-fold cross-validation.

7.2.5 Model evaluation, comparisons, and significance

Model development and assessment commenced only after the inner loop determined the optimal features and hyperparameters for each ML algorithm and the imputation model generated 10 separate training and testing sets for that fold. These steps involved fitting the ML algorithms to each imputed train set using the selected features and evaluating the algorithms on the imputed test sets based on the RMSE metric. The error measurement RMSE was chosen as the main performance metric over relational accuracy measures like predicted R^2 because it is a more appropriate metric for evaluating predictions in individuals (Nielsen et al., 2020). As an error measurement, RMSE offers advantages in terms of interpretability and its ability to penalise extreme errors (**Chapter 2.4.5**). This process yielded ten performance metrics for each outer fold. Over ten repeats of the 10-fold cross-validation, a total of 1000 performance measures were collected for all 8 ML algorithms. These performance measures were recorded for 24 distinct feature selection and outcome measures in the FC models and the multimodal models, which incorporated all modalities (**Table 7-1**). Additionally, 12 different feature selection and outcome measures were applied to models using only clinical, morphometric, or SC data.

The best performance models were selected using the lowest median RMSE for the multimodal model and each of the single modality models. Their performance was compared against baseline models, which predicted the testing set outcomes solely based on the median baseline Chalder fatigue in the training sets. Non-parametric Wilcoxon signed-rank tests quantified these comparisons as both the baseline and ML models evaluated the same testing sets. Subsequently, a non-parametric Kruskal-Wallis test assessed if the performance differed significantly between the best-performing models, which were evaluated on different train and test sets. The models were then compared pairwise using post-hoc Dunn tests. Here, the Dunn tests were preferred over Mann-Whitney U tests because they preserved the rank sums of the Kruskal-Wallis (Dinno, 2015). I also compared these models with their opposite clinical outcome to investigate differences when using outcomes at six months and one year, employing Mann-Whitney U tests. A similar comparison was made between the PASAT and resting-state functional feature variants of the best-performing functional and multimodal models. To determine if combining clinical and neuroimaging data can significantly predict fatigue, I tested whether the best-performing multimodal model predicted the outcome significantly better than chance. This was implemented using permutation testing that repeated the fitting process using shuffled outcome values (**Chapter**

2.4.5). Finally, the best-performing multimodal model that used FC from PASAT data to predict fatigue after six months was externally validated using the data from RA study 1, where all patients were coded as usual care for treatment allocation.

To elucidate the decision-making process of the best multimodal model, several interpretability methods were applied (**Chapter 2.5**). Specifically, permutation feature importance (Altmann et al., 2010) was used to provide straightforward explanations about the significance of selected features; ALE plots (Apley and Zhu, 2020) were employed to potentially illustrate non-linear relationships with fatigue outcomes; and LIME models (Ribeiro et al., 2016) were constructed to depict model decisions at the single case level. For each feature, permutation importance substituted its values with random values from the same data distribution, ranking the features based on their impact on prediction accuracy when applied to the permuted data. ALE plots were subsequently constructed to depict the average effect of input features on the output. These plots included the four most influential features from the feature importance method, along with the features for treatment allocation. The ALE plots for the remaining features were categorized into three groups: SC, FC, and a combined group of clinical and morphometric features. To represent single-case predictions, I selected the five patients with the greatest actual fatigue change and built interpretable LIME models that emulate the original model. The LIME method adjusts feature values and uses the resulting impact on the output to evaluate the contribution of each feature to the prediction of the outcome. LIME plots were constructed for the five most influential features in the patient with the greatest fatigue change, as well as the two most influential features across the five patients. For these explanatory analyses, the `iml` R package (Molnar et al., 2018) was used for feature importance and ALE plots, applying the default 1000 permutations. Additionally, the `lime` R package (Hvitfeldt et al., 2022) was used to build the LIME models, utilising 5000 permutations according to the default function settings.

All of the data prepared for ML, the scripts implemented to run the analyses, and the final models can be found at:

<https://github.com/krisbg95/LIFT/tree/41095c2292bdb45cc59e01170f232452cbe8e588/M>
L

7.3 Results

7.3.1 Patient characteristics

At baseline, 81 RA patients completed all four modalities in LIFT: T1, DTI, PASAT and resting-state fMRI. They were randomised to the three treatment groups: PEP ($n = 28$), UC ($n = 28$), or CBA ($n = 25$) group. Some patients did not complete follow-up Chalder Fatigue assessments: PEP (six months, $n = 6$; one year, $n = 6$), UC (six months, $n = 4$; one year, $n = 1$), and CBA (six months, $n = 4$; one year, $n = 1$). For the validation dataset (RA study 1), 50 patients completed baseline neuroimaging and clinical data, along with a six-month Chalder fatigue follow-up. The trial patients exhibited higher baseline compared to the RA study 1 patients (**Figure 7-2A**) while fatigue was significantly reduced in the PEP and CBA groups using both six-month and one-year outcomes (**Figure 7-2B**). The fatigue change in the UC group in the trial did not significantly differ from that of the RA study 1 patients (**Figure 7-2C**), who effectively received only standard care.

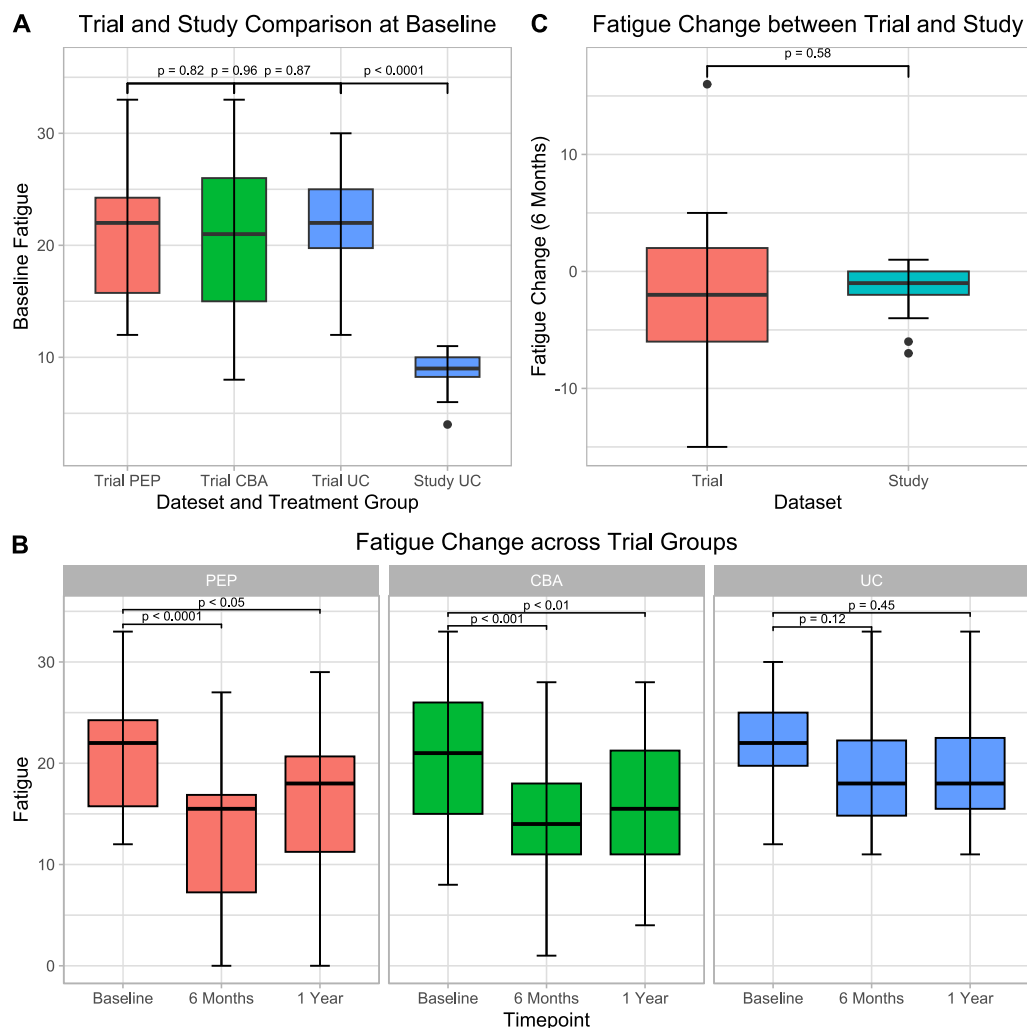


Figure 7-2: Fatigue characteristics in the trial and study RA cohorts. Panel A displays boxplots of baseline Chalder Fatigue for the three groups in the LIFT trial prior to treatment initiation,

alongside the patients from RA study 1, whose data served for final model validation. There were no significant differences between the three trial groups, but the study patients had significantly lower fatigue, as compared using independent sample t-tests. **Panel B** visualises significant changes in fatigue at both six months and one year after starting treatment for the personalised exercise programme (PEP) and cognitive behavioural approach (CBA) groups but not for the usual care (UC) group), as confirmed by paired samples t-tests. **Panel C** illustrates that fatigue differences at six months were not significantly different between the UC trial group and the RA study patients, as assessed by independent sample t-tests.

7.3.2 All model descriptions

Models employing different modalities and ML algorithms exhibited very similar performance in terms of RMSE (**Figure 7-3**). All of the algorithms performed similarly when predicting six-month or one-year outcomes across feature selection and modality options (**Figure 7-4**). The best-performing model for each modality used a different configuration of feature selection and/or clinical outcome (**Table 7-3**). Notably, the Gaussian process regression with a radial kernel emerged as the algorithm of choice for all modalities except the morphometric feature model, which employed a random forest algorithm instead. The top-performing models based on MAE or predicted R^2 are depicted in **Appendix G**.

Modality	RMSE	Algorithm	Outcome	FC	Feature selection	
					Relief	SGCCA
Clinical	6.25	gaussprRadial	1 year	N/A	1:1	50%
FC	6.60	gaussprRadial	6 months	Resting-state	1:1	50%
Multimodal	6.74	gaussprRadial	6 months	Resting-state	1:4	70%
SC	6.57	gaussprRadial	1 year	N/A	1:1	70%
Morphometric	6.64	rf	1 year	N/A	1:4	50%

Table 7-3: Configuration of the best model for each modality. Each row represents the configuration for the best-performing model in the respective modality, determined by the median root mean squared error (RMSE). The top-performing models employed either the Gaussian process regression with a radial kernel (gaussprRadial) or a random forest algorithm. The configuration includes the timing of the clinical outcome, which can be either six months or one year after treatment allocation. The feature selection column specifies whether all positive features (1:1) were included post RRelief F, or if it was narrowed down to the top half or quarter (1:4) of positive features. It also indicates whether 50% or 70% of resamples had to include the features during the stability selection for the sparse generalized canonical correlation analysis (SGCCA). In specific cases of functional connectivity (FC) and multimodal models, the selection of the source of FC was an additional consideration. The models could use either FC estimated from fMRI during resting-state or from the Paced Auditory Serial Addition Test (PASAT) task.

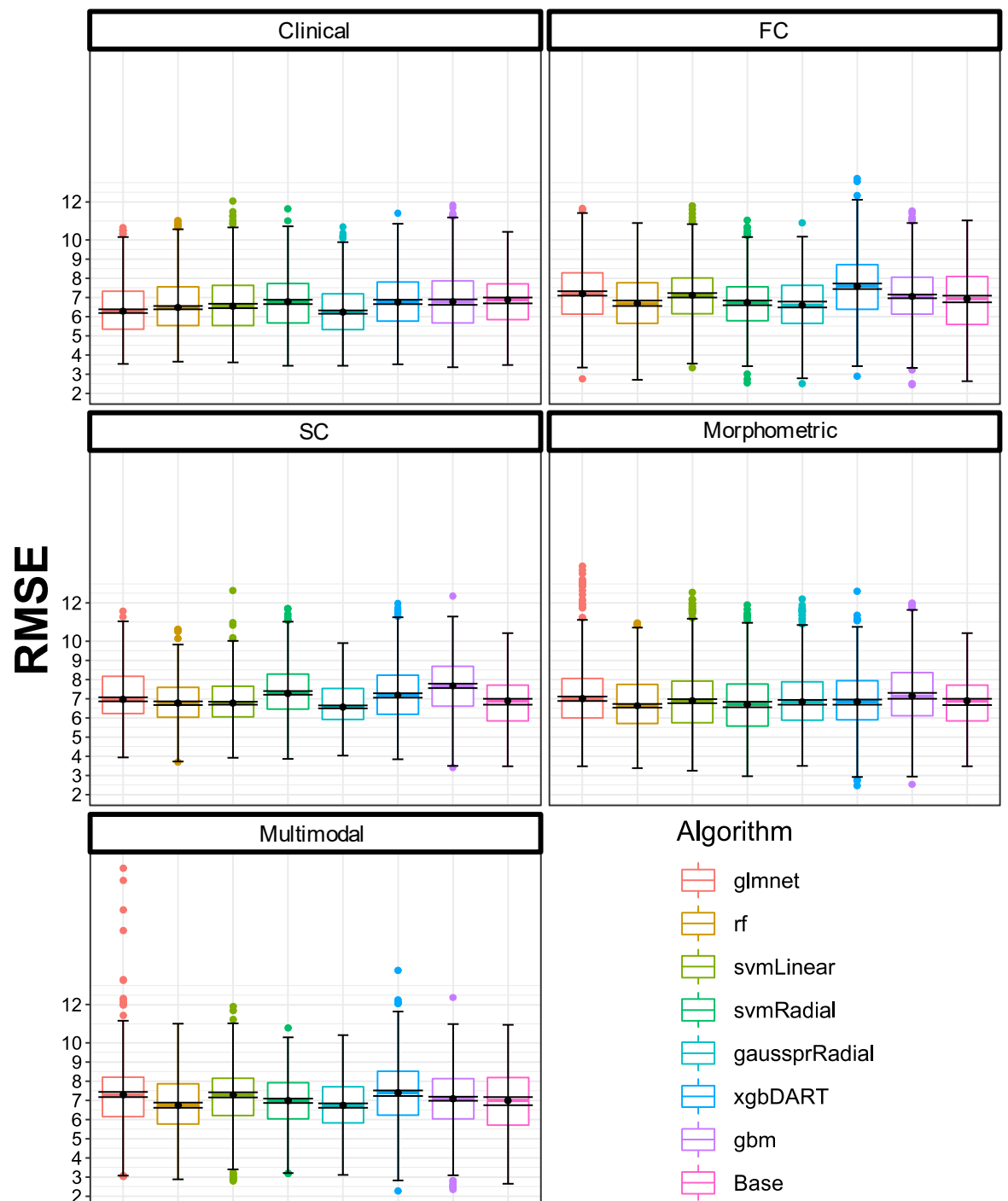


Figure 7-3: Cross-validation performance across modalities. Each panel within the figure showcases the performance of models within individual modalities, assessed by the root mean square error (RMSE). The modalities include clinical, functional connectivity (FC), structural connectivity (SC), morphometric features and a multimodal panel integrating all the features. The models are colour-coded based on the algorithm employed, which includes elastic net (glmnet), random forest (rf), support vector regression (svm), Gaussian process regression (gausspr) with linear and radial kernels, as well as gradient boosting (gbm) and extreme gradient boosting with dropout (xgbDART). Additionally, a 'Base' model relying solely on the median of the training set's baseline Chalder fatigue score is presented as a baseline reference.

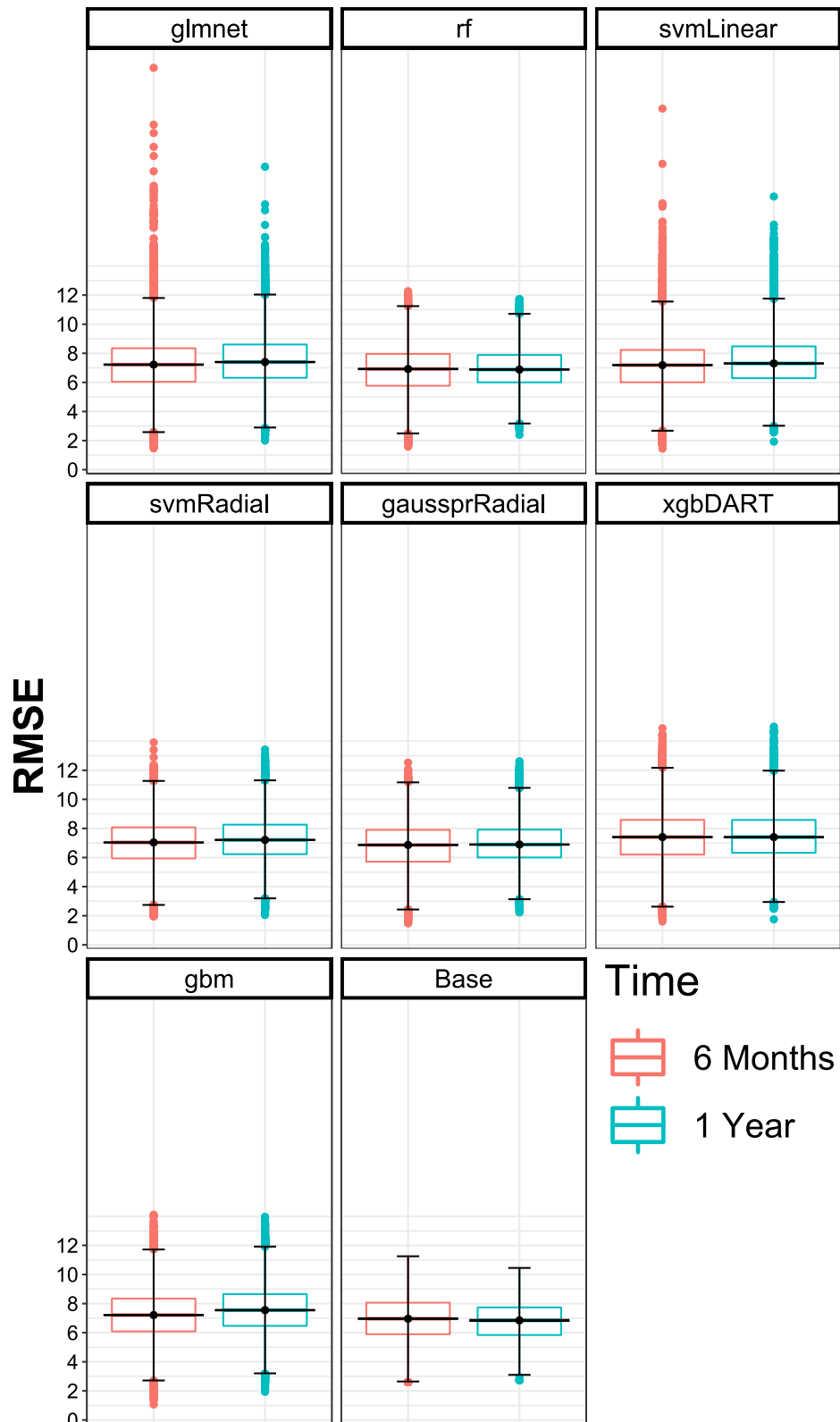


Figure 7-4: Cross-validation performance across time. In each panel, the performance of a single algorithm is presented as the root mean square error (RMSE), using Chalder fatigue outcomes at either six months or one year after treatment allocation. The results are agglomerated across different feature modalities. These algorithms include elastic net (glmnet), random forest (rf), support vector regression (svm) and Gaussian process regression (gausspr) using linear and radial kernels as well gradient boosting (gbm) and extreme gradient boosting using drop out (xgbDART). Additionally, a model (Base) solely using the median of the training set baseline Chalder fatigue score is also depicted as a baseline reference.

7.3.3 Best model descriptions

The best-performing model for each modality exhibited a significantly lower RMSE compared to the baseline model, which employed only the median (**Figure 7-5**). Specifically, the clinical model demonstrated a large effect size, the FC model displayed a medium effect size, whereas the multimodal, SC, and morphometric models demonstrated small effect sizes (Goss-Sampson, 2019). While these best-performing models were significantly different, that distinction primarily stemmed from differences with the clinical model (**Figure 7-6**). A trend level difference ($p = 0.08$) was also observed between the FC and SC models. The clinical model used as predictors treatment allocation, baseline Chalder Fatigue, ACR Fibromyalgia scale, and current pain when applied to the full dataset.

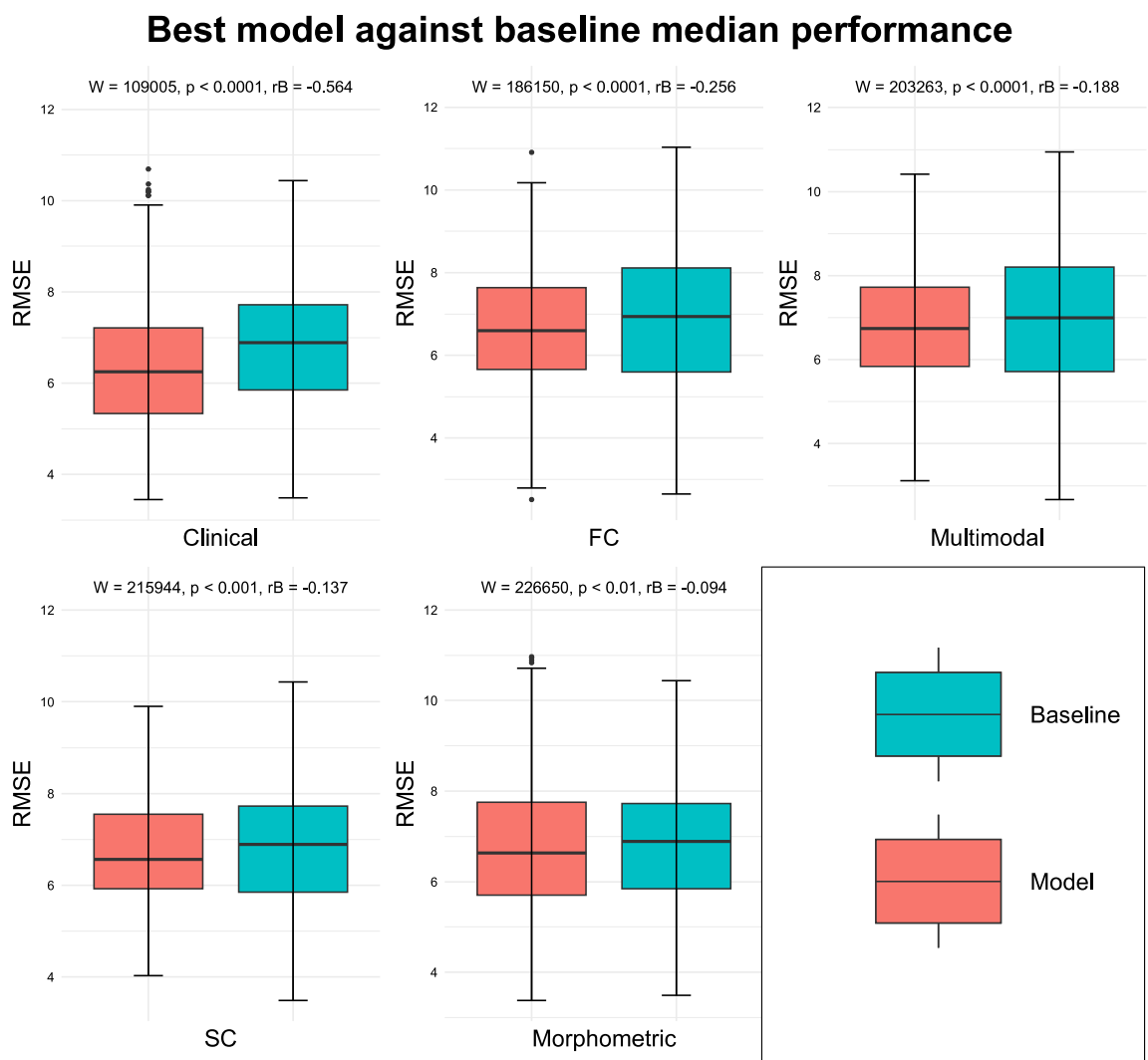


Figure 7-5: Best model against baseline median performance. For each modality, the figure displays the comparison between the best model and the corresponding median performance for the identical test sets. Above each plot, the results of a Wilcoxon signed-rank test are presented, comprising the test statistic (W), p-value, and effect size denoted by rank-biserial correlation (rB). Effect sizes can be categorised as follows: trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.3 , and large > 0.5 (Goss-Sampson, 2019).

Comparison of best model for each modality

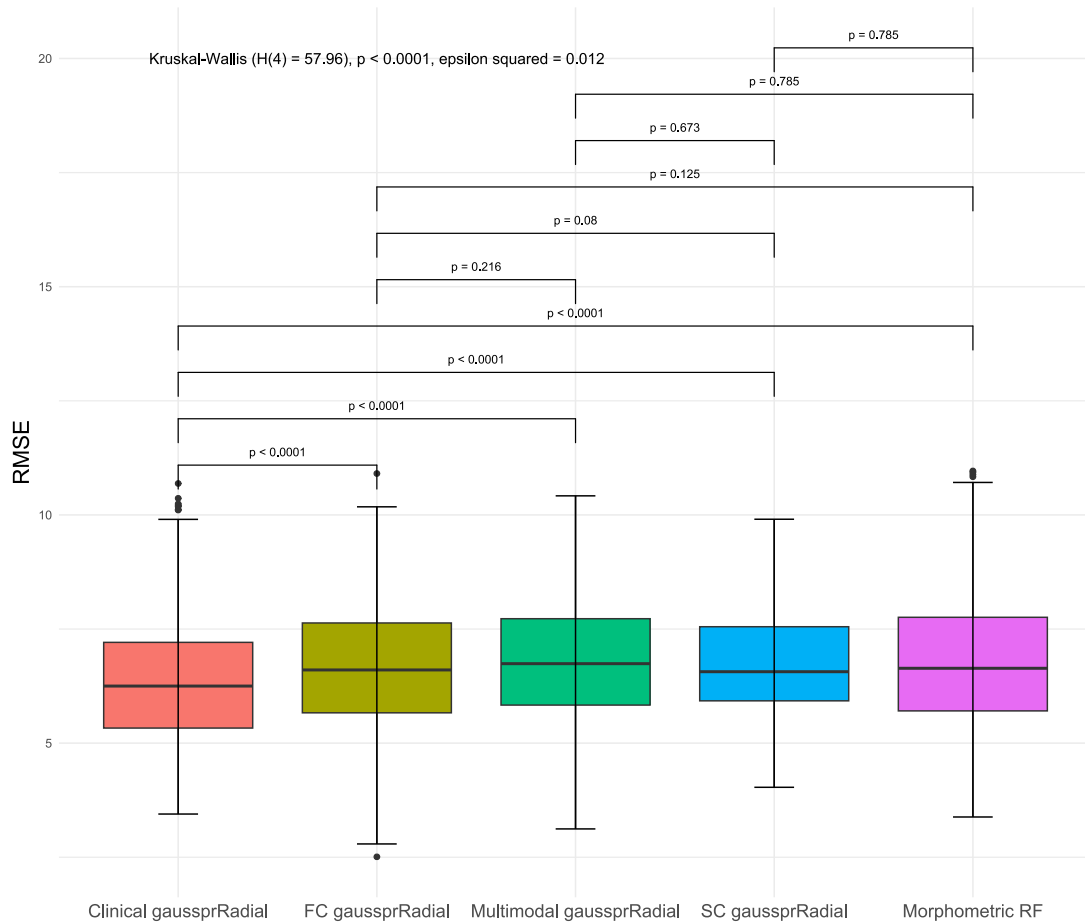


Figure 7-6: Best model comparison across all modalities. The plot displays the performance in root mean squared error (RMSE) of the best model for each modality. Along the x-axis, the modality and corresponding algorithm are displayed, including the Gaussian process regression (gaussprRadial) and the random forest (RF) algorithms. Atop the plot, a non-parametric, Kruskal-Wallis test demonstrates significant differences in model performances across modalities. This comprises of the test statistic (H), p-value, and a medium epsilon squared effect size (Valladares-Neto, 2018). The p-values of post hoc comparisons using Dunn tests are also presented.

The best models varied in the clinical outcome they predicted. The FC and multimodal models performed best in predicting the six-month Chalder Fatigue outcome, while the others were more adept at predicting the one-year outcome (**Table 7-3**). To evaluate the effect of outcome choice, I compared these models to ones with opposite outcomes (**Figure 7-7**). For example, the best-performing FC model that predicted six-month outcomes was compared to the same configuration model that predicted one-year outcomes. Although the clinical and morphometric modalities did not show significant differences, the remaining modalities were significantly different, albeit with small effect sizes. Similar comparisons were made for FC and multimodal models using FC, extracted from fMRI either during resting-state or the PASAT task (**Figure 7-8**). In both cases, the top-performing models utilized resting-state FC, and they demonstrated significant differences with small effect sizes when compared to identical models employing PASAT FC.

Best model against opposite outcome model

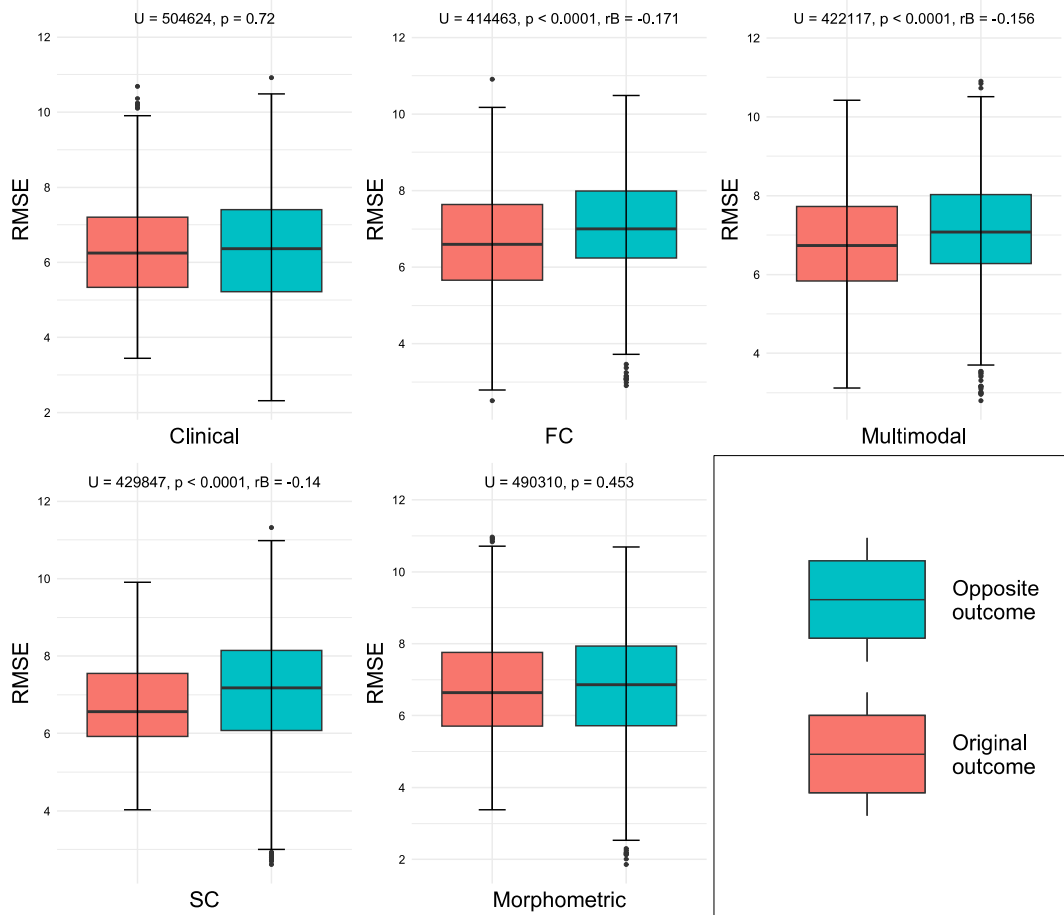


Figure 7-7: Best model outcome comparison. For each modality, the figure displays the comparison between the best model and an identical model that predicts the opposite outcome. Above each plot, the results of a Mann-Whitney U test are presented, comprising the test statistic (U), p-value, and effect size denoted by rank-biserial correlation (rB). Effect sizes can be categorised as follows: trivial < 0.1, small ≥ 0.1 , medium ≥ 0.3 , and large > 0.5 (Goss-Sampson, 2019).

Best model against opposite FC modality

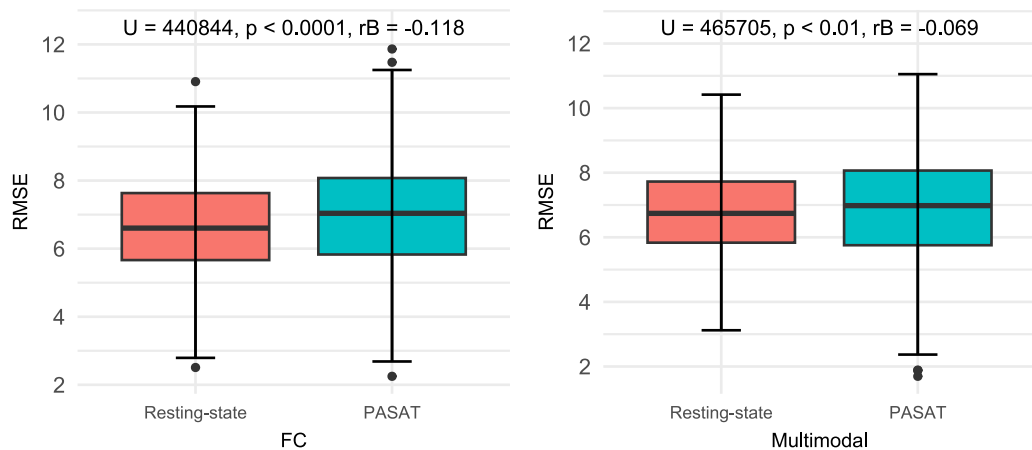


Figure 7-8: Functional connectivity modality effect. For the functional connectivity and multimodal models, the figure displays the comparison between the best model and an identical model that uses a different FC modality (fMRI during resting-state or PASAT). Above each plot, the results of a Mann-Whitney U test are presented, comprising the test statistic (U), p-value, and effect size denoted by rank-biserial correlation (rB). Effect sizes can be categorised as follows: trivial < 0.1, small ≥ 0.1 , medium ≥ 0.3 , and large > 0.5 (Goss-Sampson, 2019).

7.3.4 Best model significance and validation

To assess whether the best multimodal model performed better than chance, the model was compared to a null distribution made from shuffling (permuting) the outcome value (**Figure 7-9**). Each set of permuted values was correlated with the actual values, ensuring none exceeded a set correlation value ($r = 0.1$), thus confirming the effectiveness of the permutation. However, the actual model performance was inferior to what would be expected by chance alone.

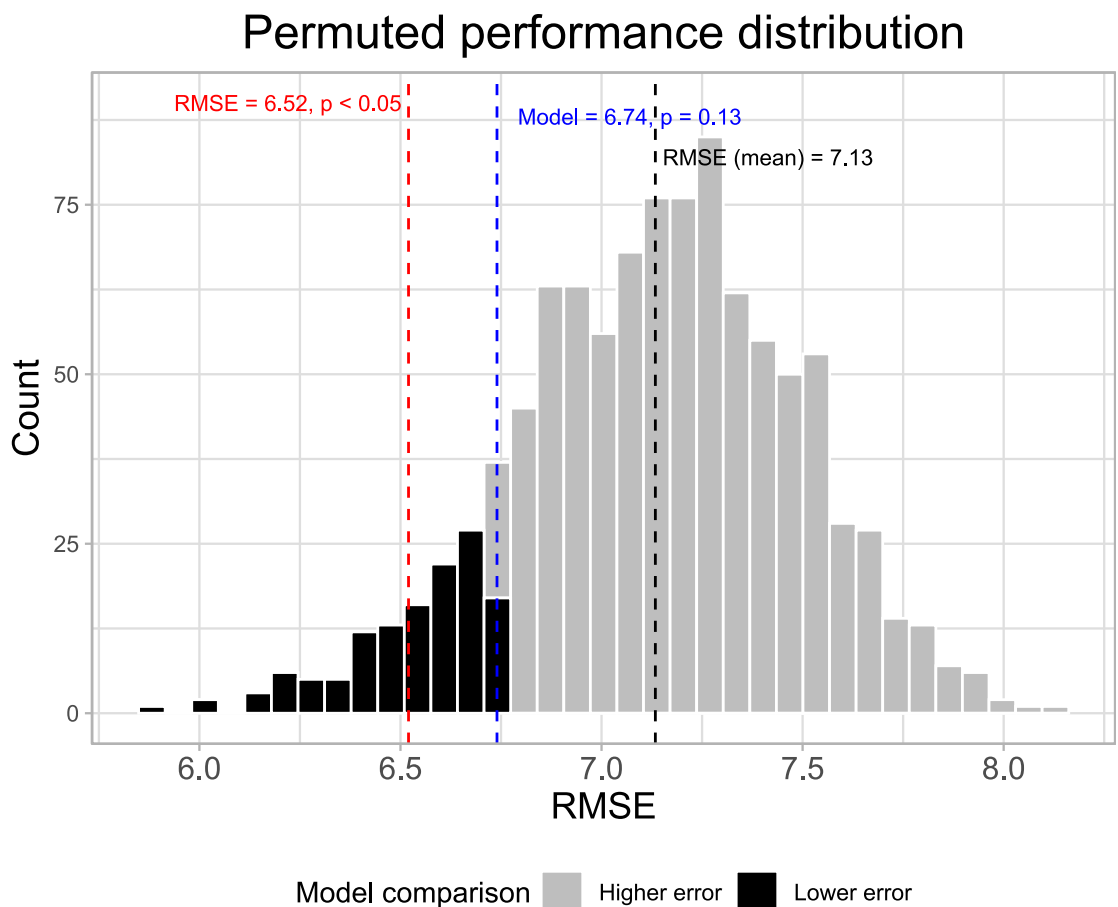


Figure 7-9: Best model significance. The figure depicts the sampling distribution of the root mean squared error (RMSE) for the best-performing multimodal model under the null hypothesis. This distribution is generated by applying the model to 1000 permutations of the outcomes and assessing its performance through 10 repeated cross-validations. Every value in the histogram represents the median performance across the validation procedure. The black dotted line indicates the mean of these values. The blue dotted line corresponds to the actual model's performance (RMSE = 6.74). Values in grey indicate higher error rates compared to the actual model, while those in black represent lower errors. The red dotted line, at an RMSE of 6.52, signifies the threshold below which performance would be deemed better than chance at a significance level of 0.05. This value represents the top 50 out of 1000 permuted performances. The actual model performed worse than what would be expected by chance ($p = 0.13$).

The top-performing model used resting-state FC, but the validation set only had FC from PASAT fMRI. Therefore, the best-performing multimodal model using PASAT FC was used for validation in the RA study 1 cohort. The two models also had differences in the configuration (**Table 7-4**) and selected features (**Appendix H**). While the SC and morphometric features overlapped, the clinical features did not. The functional connection between the left frontal pole and left insula appeared in both the resting-state and PASAT FC features of both models. When the multimodal PASAT FC model was used to predict Chalder Fatigue after six months in the RA study cohort 1, it resulted in a higher error rate (RMSE = 8.21) than in the trial cohort (RMSE = 6.87).

Model	RMSE	Algorithm	Outcome	Total features	Feature selection	
					Relief	SGCCA
Multimodal rsFC	6.74	gaussprRadial	6 months	44	1:4	70%
Multimodal PASAT FC	6.87	gaussprRadial	6 months	237	1:1	70%

Table 7-4: Best-performing multimodal models using resting-state and PASAT functional connectivity. The table lists the performance in median root mean square error (RMSE) along with the algorithm and its feature selection configuration. It also depicts the total number of features selected after applying the feature selection to the whole trial data.

7.3.5 Model interpretation

After evaluating its performance, permutation feature importance and ALE plots were applied to clarify the global behaviour of the best-performing multimodal model and uncover the reasons behind its performance. The permutation feature importance analysis highlighted the HADS depression score as the most influential feature among all 44 features. (**Figure 7-10**). Subsequently, SC from the medial orbitofrontal cortex to the frontal pole, FC between the pars triangularis and the pars orbitalis, and FC between the bank of the superior temporal sulcus and the middle temporal gyrus were identified as significant contributors (**Figure 7-12**). Permuting the remaining features resulted in similar increases in error, emphasizing the model's reliance on most of the features rather than a few selected variables. In contrast to feature importance, ALE plots can illustrate non-linear relationships. These plots focused on the top four variables identified by feature importance, along with treatment allocation (**Figure 7-11**). The plots illustrated how changes in these variables while keeping others constant, influenced the average prediction of the outcome. The ALE plots for the remaining features are in **Appendix I**. The plots displayed that being allocated to the CBA group, along with higher SC from the left medial orbitofrontal cortex to the left frontal pole, led to a decrease in the predicted fatigue levels. Conversely, elevated values for the remaining four

features were indicative of higher fatigue score predictions. Notably, all these relationships exhibited a linear nature.

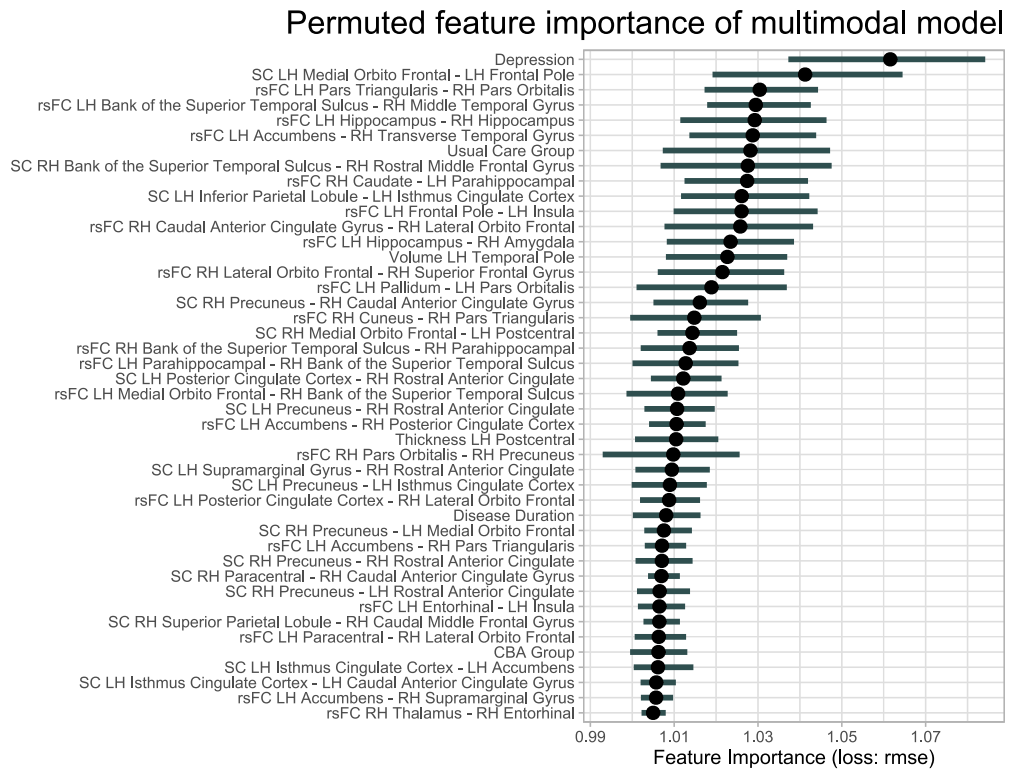


Figure 7-10: Feature importance. The figure displays the permutation feature importance of all 44 features included in the best-performing multimodal model, using a Gaussian process regression to predict Chalder Fatigue after six months in the trial data. The HADS depression score emerged as the most important feature, which when permuted resulted in an increase in RMSE by 1.06. The bars represent permutation variability, reflecting the median error across all repetitions, and 95% confidence intervals. Abbreviations: CBA, cognitive behavioural approach; LH, left hemisphere; rsFC, resting-state functional connectivity; RH, right hemisphere; SC, structural connectivity.

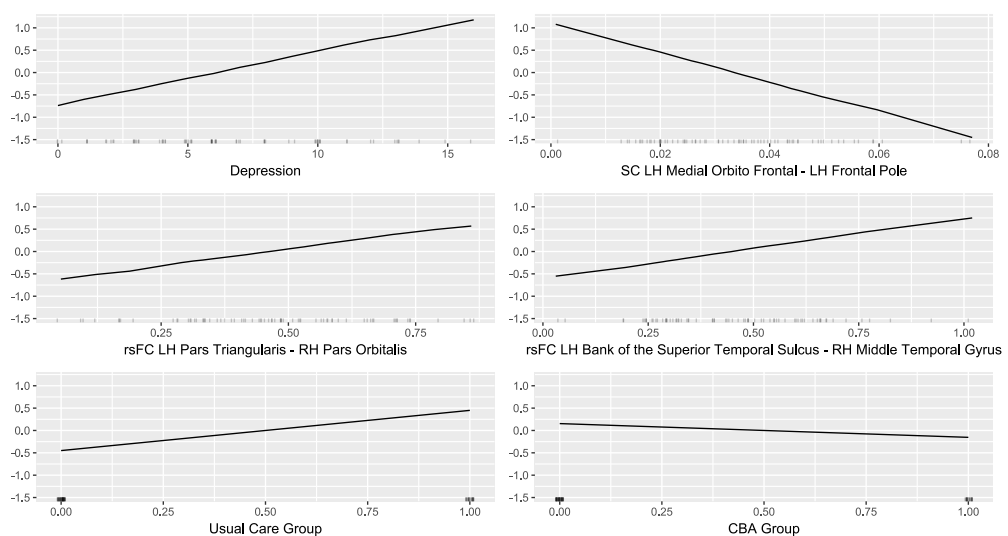


Figure 7-11: Accumulated local effects (ALE) plots of the four most influential features and treatment allocation. The plots display the main effects of the six features on the prediction of Chalder Fatigue after six months. The y-axes represent the size of the mean effect each feature has on the fatigue prediction relative to the overall model prediction. Abbreviations: LH, left hemisphere; RH, right hemisphere; rsFC, resting-state functional connectivity; SC, structural connectivity.

Top Important Imaging Features

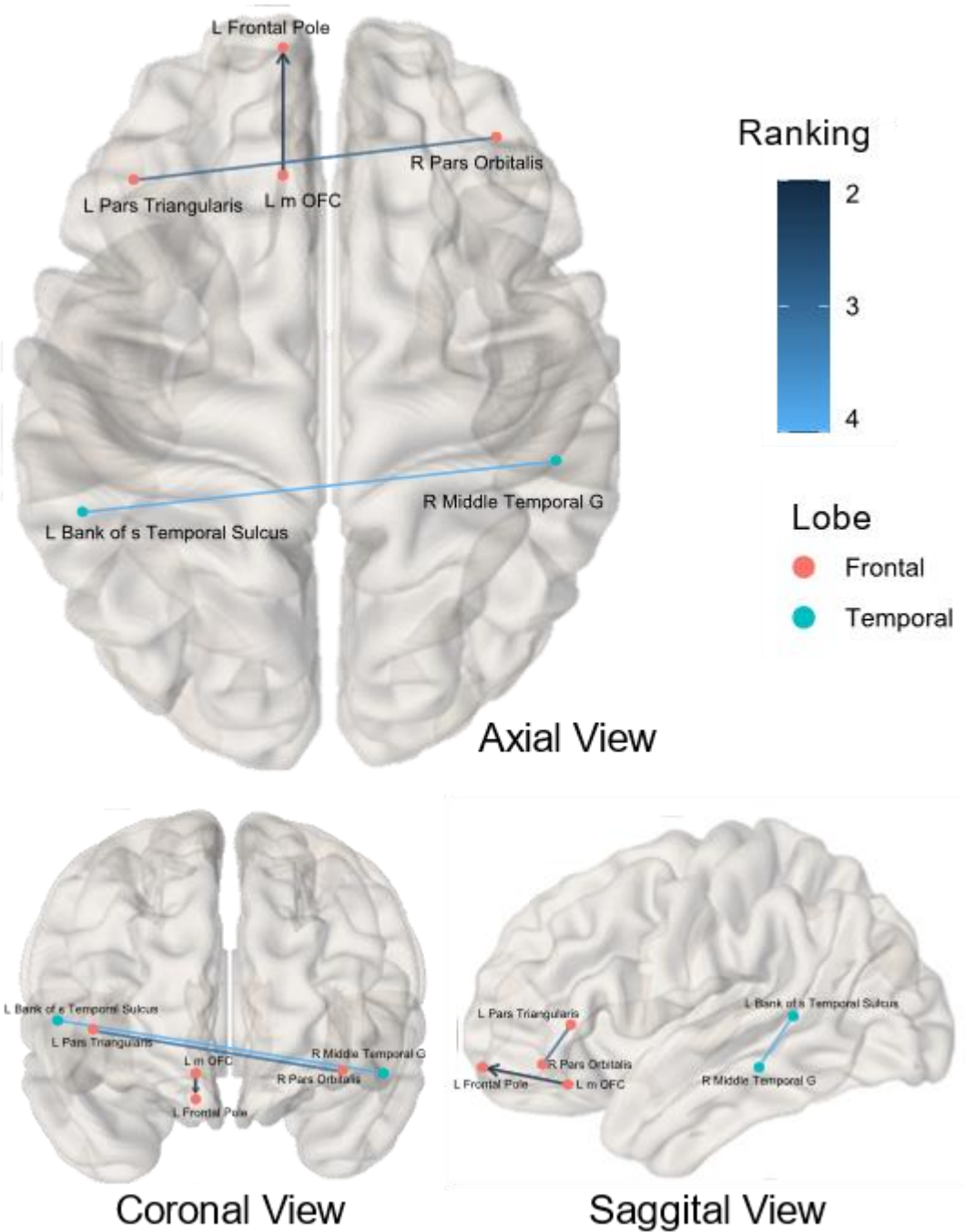


Figure 7-12: Visualisation of top imaging features. The plot displays the top three imaging features according to permutation feature importance, with HADS depression scores being the first. Structural connectivity is outlined with an arrow while functional connectivity is depicted as a single line. The colouring of the connections is based on their ranking. Abbreviations: L, left; R, right; m, medial; S, superior; G, gyrus; OFC, orbitofrontal cortex.

LIME models next offered insight into individual predictions of the multimodal model and identified key variables for specific observations. For the patient with the largest actual fatigue change (change score = -24), I constructed a LIME model to highlight the most influential features contributing to a low fatigue score. This LIME model was based on the top five contributing features (**Figure 7-13**) and comprehensively captured the actual model's behaviour ($R^2 = 0.99$). Four of the five features led to a lower predicted fatigue level prediction, with allocation to the CBA group exerting the most significant effect. Additionally, SC from the precuneus and supramarginal gyrus to the anterior cingulate cortex played a substantial role in this patient's outcome prediction. LIME models were then generated for this patient and four others with the greatest fatigue change, using the top two contributing features for each. Heat maps of these models revealed common features influencing predictions for these patients (**Figure 7-14**). The sole feature shared among two or more patients was allocation to the CBA group. It played a significant role in two patients yet reduced the predicted fatigue score in the patient with the most substantial fatigue change (case 60), while conversely increasing it in another patient (case 7). Furthermore, allocation to the usual care group predicted a lower fatigue outcome in a third patient (case 62). Finally, SC to the cingulate cortex from different regions emerged as one of the two most important features in two patients.

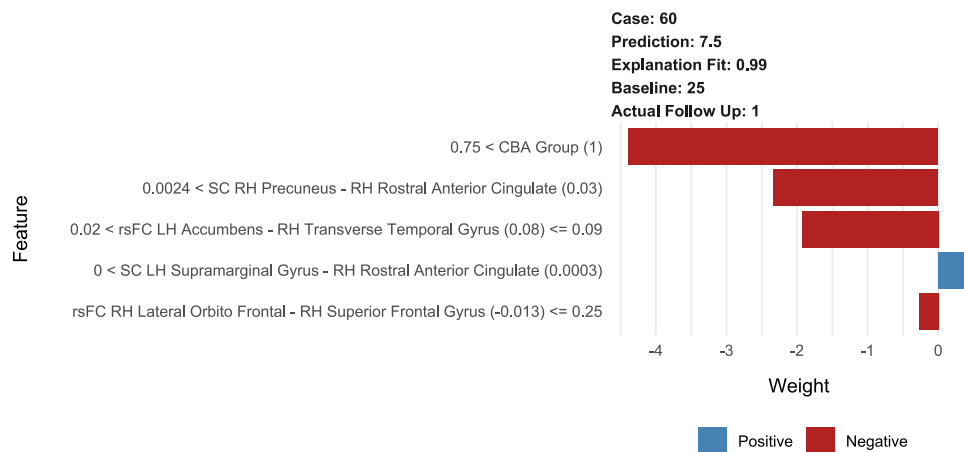


Figure 7-13: Model behaviour over the case with the greatest fatigue change. The plot delineates the 5 most influential features that best explain the linear model for the specific case. The colour of the bars depicts whether the feature brings an increase (positive) or decrease (negative) in the average predicted value for Chalder Fatigue at six months. The x-axis visualises the weighted importance of each variable to the case prediction. The plot also displays how well the linear interpretable model explains the actual model in R^2 (Explanation Fit = 0.99), the baseline fatigue score (25), the predicted fatigue score (7.5), and the actual follow-up fatigue score (1). Each feature is dichotomised for interpretation, such as structural connectivity from the right precuneus to the right rostral anterior cingulate cortex being either below or above 0.0024. Next to each feature is the actual value for that case in brackets. Abbreviations: LH, left hemisphere; RH, right hemisphere, rsFC, resting-state functional connectivity; SC, structural connectivity.

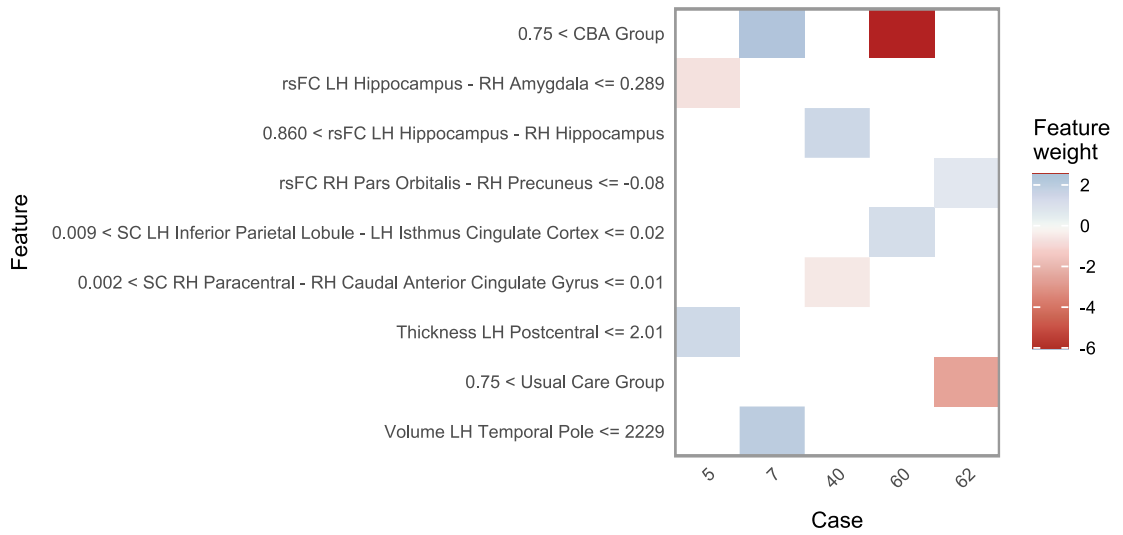


Figure 7-14: Model behaviour over the five cases with the greatest fatigue change. This figure is a heatmap showing how the top 2 features selected across all the observations influence each case. The interpretability models that used the two most influential features explained the actual model in R^2 differently for each case (case 5 = 0.46, case 7 = 0.95, case 40 = 0.51, case 60 = 0.95, case 62 = 0.85). Abbreviations: LH, left hemisphere; RH, right hemisphere; rsFC, resting-state functional connectivity; SC, structural connectivity.

7.4 Discussion

This study uniquely applied ML to predict fatigue in a trial of non-pharmacological interventions in RA, yielding several findings. Fatigue significantly reduced in the PEP and CBA groups at both six months and one year. Models using diverse neuroimaging and clinical modalities along with different ML algorithms, demonstrated comparable performance. The best-performing model in each modality showed significantly lower error rates than the baseline model, which relied solely on the median. While these top models exhibited statistical differences, this primarily stemmed from differences between the clinical and the other modality models. Moreover, the performance of the top-performing multimodal model fell below what would be expected by chance alone. Fatigue changes in the trial UC group did not significantly deviate from those of the RA study 1 patients, whose data served as validation. The highest-performing model utilized resting-state FC, yet the validation set only had FC from PASAT fMRI. Consequently, the best-performing multimodal model using PASAT FC was employed for validation in the RA study 1 cohort. When this multimodal PASAT FC model predicted Chalder Fatigue after six months in the RA study cohort 1, it exhibited a much higher error rate compared to the trial cohort. Interpretability methods highlighted depression scores, along with specific structural and functional connections, as the four most influential factors, with most indicating higher fatigue predictions. They also emphasized treatment allocation. While allocation to an active treatment group suggested a lower fatigue score prediction at a global model behaviour level, an examination at the individual case level revealed a more ambiguous relationship. CBA group allocation was associated with both a decrease and increase in fatigue level predictions in different patients. Finally, interpreting local predictions also revealed that usual care could lead to a lower fatigue level prediction in some patients compared to the global model predictions of higher fatigue.

Despite the multimodal model incorporating clinical variables alongside neuroimaging variables, it was the clinical model that performed best. The clinical model demonstrated a large effect size when compared to a baseline reference, whereas the multimodal model showed a small effect. When applied to the full dataset, the clinical model selected treatment allocation, baseline Chalder Fatigue, ACR Fibromyalgia scale, and current pain as predictors. In comparison, the multimodal model also utilized treatment allocation but included only depression and disease duration as additional clinical variables. In the full trial, both treatment groups significantly reduced fatigue outcomes, and the PEP group notably

lowered depression scores at one year (Bachmair et al., 2022). However, neither intervention substantially reduced pain compared to the UC group. Pain and depression are commonly recognized as predictors of fatigue across chronic fatigue conditions (Davies et al., 2021) and so might be predictive of fatigue outcomes irrespective of the administered treatment. The difference between the clinical and multimodal performance may be partially attributed to the clinical model's inclusion of baseline fatigue as a predictor, a factor known to be longitudinally linked to fatigue outcomes (Holdren et al., 2019).

Why did models perform similarly across modalities? In each modality, the leading model adopted a distinct configuration for feature selection and clinical outcome options. While these models varied in FC modality and predicted clinical outcomes, these variations exerted a significant but small effect on performance. What remained consistent across best-performing models was the use of Gaussian process regression with a radial kernel, except for the morphometric feature model, which opted for a random forest algorithm. Gaussian process regression thrives with smaller datasets (Li et al., 2021). This was exemplified by a study encompassing multiple ML algorithms, including elastic net and random forest, demonstrating the superiority of the Gaussian process regression with fewer observations ($n = 50-400$) and more features (400-1000) (Jollans et al., 2019). These results were demonstrated by using both simulated data, designed to have large effect sizes, and real data predicting participant age with established moderate-to-large effects. Notably, the study also indicated that the random forest algorithm excelled in scenarios with small effect sizes, potentially explaining the superior performance of the morphometric model in predicting fatigue in this study. Gaussian process regression models have found extensive application in predicting brain age (Cole et al., 2017). They have also been successful in using fMRI data to predict anxiety (Portugal et al., 2019) and pain scores (Marquand et al., 2010), both of which are multifaceted behaviours akin to fatigue. Furthermore, when Portugal et al. (2019) predicted anxiety scores, they depicted that predictions were based on the overall pattern rather than on a small combination of features. This aligns with the findings of feature importance in this study, where all features contributed similarly to error increases, emphasizing the model's reliance on a comprehensive feature set. Finally, while Portugal et al. (2019) worked with a smaller sample, studies have also documented success in implementing Gaussian process regression in larger datasets (Chen et al., 2022a, Dunas et al., 2021, Niu et al., 2020).

The failure of the multimodal model to exceed chance performance or externally validate in the RA study cohort raises questions about its construction. First, differences emerged in the

configuration and selected features of the best resting-state multimodal and PASAT multimodal models. While the SC and morphometric features overlapped, the clinical features did not align. The fact that the model did not perform better than chance indicates its limited capacity to offer predictive value in this dataset. A previous study highlighted similar challenges, demonstrating that even previously successful ML and linear models struggled to outperform chance levels in predicting relapse for stimulant dependence using fMRI and clinical data (Gowin et al., 2019). The authors attributed this to both the small size of the training sample and unique sample characteristics. Second, the performances of the imaging models could also have been a result of the Desikan-Killiany atlas's limitations in capturing functionally maximally distinct brain regions. A study demonstrated that common anatomical parcellations, including the Desikan-Killiany atlas, failed to surpass chance levels in predicting functional boundaries for both task-based and resting-state fMRI datasets (Zhi et al., 2022). This was attributed to task-based functional boundaries not aligning well with sulcal landmarks. In contrast, parcellations based on resting-state fMRI data showed promising performance, in some cases rivalling parcellations defined on the evaluation data itself. Furthermore, multi-modal parcellations combining functional and anatomical criteria fared considerably worse than those based on functional data alone, suggesting that functionally homogeneous regions often span major anatomical landmarks, which may have also happened in this study.

The nature of fatigue also likely contributed to the lack of success of the multimodal model. Being an internalised state, fatigue may be more difficult to predict compared to other behaviours. Research has shown that internalising behaviours, such as withdrawal, are inherently more challenging to predict compared to externalising behaviours like aggression and attention, as evidenced by a study utilizing FC (Dhamala et al., 2023a). Moreover, combining resting FC and task FC significantly improves predictions related to externalised cognition and personality scores, but not internalised mental health scores (Chen et al., 2022b). Another study employing ML models failed to outperform chance levels when predicting antidepressant response using genetic and clinical characteristics (Maciukiewicz et al., 2018). While the authors partially attributed these results to unbalanced responder and non-responder groups, the low performance may have been rooted in the multi-causal nature of depression and treatment response, affirming the widely accepted notion that depression onset, presentation, and reduction involve a complex interplay of various factors (Scala et al., 2023). Such notions can similarly be applied to fatigue and explain the modest performance of the models predicting fatigue in this study.

Permutation feature importance and ALE plots provided further insight into the findings. Depression scores proved most influential, followed by SC from the medial orbitofrontal cortex to the frontal pole, FC between the pars triangularis and pars orbitalis, and FC between the bank of the superior temporal sulcus and middle temporal gyrus. CBA group allocation, along with higher SC from the left medial orbitofrontal cortex to the left frontal pole, predicted lower fatigue levels. Conversely, UC group allocation, along with elevated values for the two FC features, predicted higher fatigue scores. These observations centred on the frontal cortex, which along with the basal ganglia, exhibit grey matter pathology, glucose hypometabolism, hypoperfusion, and metabolic abnormalities, in the context of fatigue in neuroinflammatory and autoimmune diseases (Morris et al., 2015). These findings are also mirrored in depression, potentially suggesting why depression scores emerged as the prime predictor in this study. Dysfunction in the frontostriatal neurocircuitry has been suggested to lead to an increase in effort perception and consequently fatigue (Kuppuswamy, 2022). Frontal regions have subsequently become focal points for brain stimulation therapies for fatigue in both MS and Parkinson's disease (Zaehle, 2021). Characteristics of frontal regions have also been shown to predict fatigue. For example, baseline FC between the dorsolateral prefrontal cortex and the caudate predicted modafinil-associated decreases in poststroke fatigue within a randomized controlled trial (Visser et al., 2019). Apart from frontal regions, feature importance also highlighted the DMN in this study. The FC between the bank of the superior temporal sulcus and the middle temporal gyrus, both components of this network (Uddin et al., 2019), was the third most influential feature in the multimodal model. As discussed in the previous chapter on fatigue predictors in PsA, augmented DMN FC has been shown to drive fatigue in MS (Bisecco et al., 2018) and correlate with subjective mental fatigue in healthy cohorts (Gergelyfi et al., 2021).

The LIME model for the patient with the largest actual fatigue change identified four key features leading to lower predicted fatigue levels, with CBA group allocation being the most influential. Among the four other patients with substantial fatigue change, CBA group allocation was the sole shared feature. While it played a significant role in two patients, it reduced predicted fatigue scores in one and increased them in another. Allocation to the UC group predicted lower fatigue in a third patient. These contradicting feature effects may explain why the model ultimately failed to predict fatigue outcomes better than chance. While ALE plots suggested CBA group allocation generally predicted lower fatigue and UC group allocation predicted higher fatigue, individual cases did not consistently follow this pattern. In a similar vein, a study used comparable interpretability methods to identify why ML models occasionally made false-negative predictions regarding the severity of SARS-

CoV-2 infections (Wu et al., 2021). The authors developed models to predict infection severity based on baseline symptoms and inflammatory markers. They discovered that initial asymptomatic presentations led the models to incorrectly classify patients with severe outcomes as normal cases. Conversely, models that made the correct predictions were less influenced by symptom data, focusing instead on the elevated baseline inflammatory markers in these patients. Another study used interpretability methods on models that predicted mortality from perioperative cardiac arrest prior to hospital discharge (Shang et al., 2022). The authors identified that the values of either CPR duration or haemorrhage variables overpowered other factors in instances of false-positive or false-negative predictions.

This study had several strengths and limitations. This includes the inherent challenge of collecting and scanning patients for neuroimaging datasets, which tend to be relatively small due to the associated costs and time investments. Nevertheless, this study uses the largest neuroimaging database to date of RA patients who have undergone a trial of non-pharmacological fatigue-specific treatments. Also, the study used the Chalder Fatigue Scale as the main outcome, which is a self-assessment questionnaire that inevitably introduces some bias from the subjective individual perception of fatigue. Alternatively, this could be addressed by focusing on fatigue measures that consistently associate with neurobiological metrics so as to maximise the predictive potential of models using neuroimaging. The integration of multimodal data presented a challenge, requiring careful consideration of heterogeneities between different modalities. This encompassed differences in data distributions, varying feature counts, and distinct discrimination capacities for fatigue changes. Despite applying feature selection to each modality in isolation, the study might have still failed to overcome the multimodal heterogeneity. To address this issue, alternative integration strategies could be used such as multikernel or stacked learning that construct representations for each modality (Chen et al., 2022c). These representations could then serve as inputs for the final algorithm, a strategy previously suggested in psychiatry (Chen et al., 2023). A strength of the study was the regression strategy, which as opposed to a classification approach, offered more clinically pertinent insights into patient progression. Furthermore, while all patients were locally recruited (Scotland, UK), potentially limiting universal applicability due to environmental variations, the inclusion of data from two imaging sites, accounted for differences in scanning parameters, and subsequently enhanced ecological validity. Although the results did not support combining different MRI data, the integration of additional biomarker sources, such as genetic or metabolic information, holds promise for improving predictive accuracy and uncovering novel fatigue biomarkers.

Overall, this study built a ML model to predict fatigue in RA patients who underwent non-pharmacological interventions, which significantly reduced their fatigue at follow-up. Different models using neuroimaging and clinical data showed comparable performance, potentially because most used the same algorithm while the effects of imaging modality time of outcome were not large. The clinical model performed best probably because of its few but very influential features, including baseline fatigue which the multimodal model did not use. The multimodal model also failed to exceed chance performance or validate in a separate RA cohort potentially due to a small training sample size, the limitations of the brain region atlas used, and the inherent difficulty in predicting internalised behaviours like fatigue. Despite its low performance, the model highlighted the importance of frontal regions and the DMN in predicting fatigue in coherence with previous findings. Interpretability methods revealed contradictory feature effects and inconsistent patterns in individual cases, which may explain the failure of the multimodal model to predict fatigue outcomes accurately. Nevertheless, this study showcased the feasibility of predicting fatigue outcomes in the context of non-pharmacological interventions combining neuroimaging with clinical data and where further improvements can be made.

Chapter 8 Conclusions

8.1 New knowledge

This thesis presented several novel findings. It established a link between neuroimaging brain connectivity and distinct subgroups in RA (rheumatoid arthritis) related to fatigue subdimensions, albeit only within a specific cohort. Some brain regions overlapped in governing how patients were grouped, including the default mode network regions of the precuneus and parahippocampal gyrus. I also found associations between brain imaging metrics and baseline subdimensions of fatigue in RA, with different subcomponents correlating with different metrics. Emotional and cognitive fatigue did converge on the same resting-state FC (functional connectivity) between the frontal pars orbitalis and a part of the inferior parietal lobule called the supramarginal gyrus. **Chapter 5** identified baseline brain imaging predictors of fatigue in RA patients undergoing either an exercise or cognitive-behavioural intervention. In both fatigue-specific treatments, lower fatigue at follow-up was predicted by stronger SC (structural connectivity) from the precuneus to the anterior cingulate cortex as well as stronger PASAT (Paced Auditory Serial Addition Test) task FC between the anterior cingulate cortex and the frontal pole. In contrast, significant neuroimaging predictors of fatigue were not found in RA patients who started a new DMARD (disease-modifying antirheumatic drug) treatment. However, I did find such predictors in PsA (psoriatic arthritis) patients. These were comprised of cortical thickness of the visual pericalcarine cortex and FC of regions of the default mode and salience networks, including the inferior parietal lobule and the anterior cingulate cortex. Finally, models using diverse neuroimaging and clinical modalities along with different ML (machine learning) algorithms, demonstrated comparable performance in predicting individual fatigue follow-up scores in the same trial data from **Chapter 5**, with the inclusion of the control usual care group. Although the best multimodality model outperformed the model using solely the baseline median fatigue, it neither surpassed chance level nor did it replicate its utility in usual care patients in an independent RA cohort. Interpretation methods indicated that while this model heavily relied on treatment allocation, it showed contradictory feature effects and inconsistent patterns in individual cases.

The results discussed highlight multiple brain regions and their affiliated networks (**Table 8-1**). This could be attributed to the sporadic nature of findings inherent in human brain-behaviour neuroimaging research. However, it may also in part be due to the involvement

of multiple potential mechanisms. This may elucidate why even the default mode network (DMN), the one consistent brain network across all studies, is typically discussed in relation to various mechanisms. Firstly, the dynamic FC of the DMN played a crucial role in clustering patients alongside the SMN and insula. These components share a role in pain processing, with the sensorimotor network (SMN) contributing to peripheral pain pathways, while associations between the DMN and insula serve as the most frequently cited biomarkers of centralised pain (Kaplan et al., 2019). Consequently, these results may reflect that the impact of fatigue in RA patients hinges on the influence of each type of pain. Furthermore, the connectivity of the DMN, particularly the inferior parietal lobule, has been previously associated with systemic inflammation in the same RA cohort (Schrepf et al., 2018). Interestingly, hyperconnectivity of the inferior parietal lobule predicted high fatigue in the PsA cohort that underwent new immunological DMARD treatment. Also, as this was connectivity with another region of the DMN, it also implied the role of the DMN as an indicator of attentional resources (Wylie and Flashman, 2017). This role could be important in non-pharmacological fatigue treatments, as connectivity within the DMN also predicted fatigue in RA patients who underwent exercise or cognitive-behavioural interventions. Despite its implications in disease, the DMN may also be influential in capturing a general mechanism of tiredness, as observed in healthy cohorts (Gergelyfi et al., 2021).

Multiple mechanisms were also suggested as a result of different associations with subdimensions of fatigue. The impact of living with fatigue (BRAFF MDQ), which measures motivation, implicated both the nucleus accumbens, a prime region of the mesolimbic pathway, and limbic anterior cingulate areas as well as prefrontal regions which regulate this pathway. Due to its functions, the mesolimbic pathway may induce fatigue by heightening the perception of energy costs for actions and/or reducing expectations of rewards. This aligns with a central mechanism believed to underlie anhedonia in depression (Clery-Melin et al., 2019). Physical and cognitive fatigue, on the other hand, implicated regions within task networks that are related to their respective behaviours. Theoretically, both types of fatigue can emerge when connectivity within task networks is disrupted, coupled with amplified interoceptive and cognitive control signals from a general monitoring network (Muller and Apps, 2019). This network comprises of the dorsolateral prefrontal cortex, dorsal anterior cingulate cortex, and the anterior insula. Inflammation may serve as one pathway to disrupt this monitoring network, particularly through cytokine-induced interoceptive changes in the insula (Hanken et al., 2014). One potential avenue of alleviating fatigue may therefore be improving interoceptive processes, which have been shown to associate with treatment response in depression (Eggart and Valdes-Stauber, 2021).

In PsA, higher cortical thickness of the primary visual cortex (pericalcarine) predicted greater fatigue six months after starting a new DMARD. This region's structural characteristics may signal general centralised symptoms, as it is linked to fibromyalgia, low mood, and cognitive deficits (Nhu et al., 2023, Aster et al., 2022, Pitteri et al., 2021, Jensen et al., 2013). Fibromyalgia, known for being primarily driven by top-down mechanisms, shares chronic fatigue as a prominent feature. FC between the pericalcarine and the cuneus, both parts of the primary visual cortex, also predicted fatigue, but in RA patients after receiving non-pharmacological interventions. In PsA, FC between the anterior cingulate cortex (a region of the SLN) and the lingual gyrus served as a neuroimaging marker for multisensory hypersensitivity—an amplified response to non-painful sensory stimuli (Lopez-Sola et al., 2014). This suggests that top-down mechanisms might predispose PsA patients to fatigue, irrespective of nociceptive input. Conversely, higher baseline SC from the precuneus to the anterior cingulate cortex predicted lower fatigue after non-pharmacological interventions in RA, indicating that patients with such characteristics may be more receptive to these treatments. Overall, SC was more prevalent than FC in predicting fatigue after exercise or cognitive behavioural interventions, possibly due to its better reflection of trait fatigue (Genova et al., 2013). On the other hand, FC may be more adept at characterising current state fatigue. In MS, functional changes are postulated to signify different stages of fatigue progression in brain circuits (Capone et al., 2020). Finally, apart from differences between SC and FC, another noteworthy methodological variation lies in the use of multivariate analyses, which led to a more comprehensive, interdependent set of variables predicting fatigue, as opposed to a single hub in the brain.

Brain regions related to fatigue	Affiliated brain networks or pathways
Neuroimaging clusters of fatigue in Chapter 3 (associations)	
precuneus	DMN
parahippocampal gyrus	SMN
precentral gyrus	SLN
insula	
supramarginal gyrus (inferior parietal lobule)	
Correlates of fatigue subdimensions in Chapter 4 (associations)	
impact on living (motivational) fatigue	mesolimbic pathway
nucleus accumbens	SLN
caudal and rostral middle frontal gyrus	SMN
frontal pole	
pars opercularis and triangularis	
rostral anterior cingulate	
isthmus cingulate	
paracentral lobule	
physical fatigue	SMN
paracentral lobule	
precentral and postcentral gyrus	
cognitive fatigue	SLN
insula	DAN
lingual gyrus	DMN
pars orbitalis	
supramarginal gyrus (inferior parietal lobule)	
emotional fatigue	DMN
pars orbitalis	
supramarginal gyrus (inferior parietal lobule)	
Fatigue predictors in RA after non-pharmacological interventions in Chapter 5	
precuneus	DMN
caudal and rostral anterior cingulate	SLN
frontal pole	Visual network
cuneus	
pericalcarine cortex	
Fatigue predictors in PsA after pharmacological interventions in Chapter 6	
pericalcarine cortex	Visual network
caudal anterior cingulate	SLN
lingual gyrus	DAN
inferior parietal lobule	DMN
bank of the superior temporal sulcus	
Fatigue prognostic models in Chapter 7	
medial orbitofrontal cortex	DMN
frontal pole	
pars triangularis and orbitalis	
bank of the superior temporal sulcus	
middle temporal gyrus	

Table 8-1: Summary of brain regions found across thesis results and their affiliated brain networks. The table is segregated according to the results of each chapter. In the case Chapter 4, this

is further subdivided according to the sub-score of fatigue. The brain networks are derived from Uddin et al. (2019) and listed in Error! Reference source not found.. Abbreviations: DMN, default mode network; DAN, dorsal attention network; PsA, psoriatic arthritis; RA, rheumatoid arthritis; SLN, salience network; SMN, sensorimotor network.

8.2 Limitations

Limitations across the studies in this thesis relate first to the neuroimaging methodology. Neuroimaging encounters inherent challenges due to the intricate nature of the brain, the inverse spatial-temporal resolution relationship of these techniques, the high dimensionality of features like connectivity matrices, and subject and study variability (Zhu et al., 2023). The brain's state is influenced by disease, age, genes, race, and environmental factors. Scanner effects, acquisition parameters, and noise from motion or thermal factors affect the neuroimaging signal. Neuroimaging results can be sensitive to preprocessing and analysis pipelines (Botvinik-Nezer et al., 2020). Small sample sizes (≤ 20) in neuroimaging struggle to demonstrate brain-behaviour associations due to their small effect sizes (Libedinsky et al., 2022). This applies to the limited number of RA patients compared to PsA who underwent new DMARD treatment, especially those who completed follow-up. Although not small in univariate contexts, the LIFT dataset did not afford a good sample size to train ML algorithms. While not small in univariate contexts, the LIFT dataset did not provide a robust sample size for training ML algorithms. Despite no overfitting, such sample sizes have shown inflated error bars in ML prediction (Varoquaux, 2018). These sample size challenges arise from the difficulty in collecting and scanning patients for neuroimaging datasets, due to associated costs and time commitments. Although dataset sizes aligned well with established univariate analysis methodology, they primarily identify independent biomarkers. For subdimensions of fatigue, they might have pinpointed variables with the strongest effect on different dimensions of fatigue, but common patterns may emerge with multivariate testing. The Desikan-Killiany atlas was suitably paired with the univariate approach as a coarse anatomical parcellation capable of distinguishing large brain regions for future treatment targeting. However, in the ML context, it has limitations in capturing functionally distinct brain regions, as functionally homogeneous regions often span major anatomical landmarks (Zhi et al., 2022). Finally, integrating multimodal data in ML models posed a challenge due to heterogeneities between different modalities in their data distributions, varying feature counts, and distinct discrimination capacities for fatigue differences.

Limitations apply due to the characteristics of the cohorts and their alignment with the analysis pipelines as well as the concept of fatigue. Sub-dimensional fatigue findings are confined to the LIFT cohort, characterised by relatively low systemic inflammation, which may not represent the broader RA population. Thus, certain brain circuits may be more affected by fatigue than others. Moreover, correlation and clustering results did not hold when exclusively using baseline imaging and clinical data from LIFT. The clustering analysis in this thesis may not have been optimal for more heterogeneous data. Patients from RA study 1, where this pipeline was effective, exhibited more inflammation but less fatigue compared to those in the LIFT trial. While the adapted clustering pipeline may have reduced overfitting, it potentially resulted in insufficient brain metrics for extracting clinically relevant data, especially if underlying biological mechanisms differ between the cohorts. Findings of brain correlates of subdimensions of fatigue are confined to the BRAF measure, as no other multidimensional questionnaire was used. Additionally, while the living subscore in this questionnaire gauges how fatigue impacts planned behaviour and motivation, it is designed to measure impact rather than motivational fatigue specifically. The Chalder Fatigue Scale, though valid for assessing fatigue severity, aligns with the study's focus but remains a self-assessment questionnaire, introducing inherent bias from subjective individual perceptions of fatigue. Such questionnaires may still struggle to distinctly discern transient perceptions from persistent feelings of fatigue. The intrinsic nature of fatigue adds complexity to brain-based investigations. Being an internalised state, predicting fatigue is more challenging compared to more externally observable behaviours like aggression (Dhamala et al., 2023a). Mainly, the multi-causal nature of fatigue means that its onset, presentation, and reduction involve a complex interplay of various factors. Lastly, while these studies leveraged applicable pharmacological and non-pharmacological treatments to determine fatigue predictors, they did not manipulate brain activity to determine if the identified neural signatures can alleviate fatigue and thus inform directly on its mechanisms.

8.3 Alternative approaches

The rationales behind the pipelines of each study have been discussed, yet there are alternative approaches to address the same research questions with the existing data. Firstly, while clustering methods lack definitive standards, recent efforts have aimed to establish better guidelines, particularly in mental health research (Gao et al., 2023). In **Chapter 3**, considering the high-dimensional imaging data, although the employed dimensionality reduction technique was suitable, it is not specifically tailored for cluster identification,

potentially omitting important information. In this context, utilizing subspace clustering could have been more effective in identifying variables that best segregate potential clusters (Sim et al., 2013). Another option could have been to implement multi-view clustering (Chao et al., 2021), also replacing the canonical correlation analysis in the process. This method leverages naturally occurring views of data representations (such as brain connectivity and fatigue subdimensions), and the combined clustering information from individual views could then be consolidated to yield a final cluster solution. Notably, while the study adhered to an already published and statistically sound pipeline, I could have devised a unique workflow, entirely based on the research question and available data.

Brain imaging associations varied among fatigue subdimensions in **Chapter 4**, but employing a multivariate approach might have revealed common structural or functional brain patterns. This could be particularly impactful for FC, which reflects state fatigue. One alternative would be to use functional connectivity Multi-Variate Pattern Analyses (fc-MVPA) in the CONN toolbox, which makes brain-wide connectome inferences (Nieto-Castanon, 2022). This method evaluates, for each voxel separately, the complete multivariate pattern of functional connections between that voxel and the rest of the brain and then repeats the analysis for each voxel and repeats these analyses across all voxels. This creates an FC map for each subject, which is then characterised by lower dimensional scores that best capture the observed voxel-specific variability across subjects. Unlike other multivariate approaches, this method uses a forward model of the data which enables to make statistical inferences about individual voxels in the brain and a behaviour of interest like fatigue. This offers insights into whether the configuration of FC patterns alters with varying levels of fatigue. Consequently, it enables the identification of specific regions where FC plays a pivotal role in fatigue and which part of the brain drives this connectivity pattern.

Finally, as the multimodal ML model failed to predict fatigue, there are potential solutions that could address its limitations. First, adopting alternative integration strategies could address potential heterogeneities across different modalities. Such strategies include multikernel or stacked learning that construct representations for each modality (Chen et al., 2022c). These representations could then serve as inputs for the final algorithm, a strategy previously proposed in psychiatry (Chen et al., 2023). Second, the Desikan-Killiany atlas may have limitations in capturing functionally distinct brain regions. Parcellations based on resting-state fMRI data may offer superior performance and could be on par with parcellations defined on the evaluation data itself. To enhance predictive accuracy, selecting

a fatigue measure based on the predictive performance of neuroimaging variables in the other datasets could maximize the potential of ML models employing neuroimaging in LIFT.

8.4 Future directions

The primary aim of this thesis was to address the question of how brain imaging can deconstruct the heterogeneity of fatigue and provide both pathophysiological insights and inform treatment stratification in IA. Although none of the findings can currently translate to ready-to-use applications, it does devise innovative patient solutions.

Solutions for patients could entail integrating insights from individual neuroimaging models with direct human brain modulation methods. Neuromodulation, like TMS, offers an expanded toolkit for combating chronic fatigue. In TMS, electromagnetic coils directly modulate brain activity (Hallett, 2007). In clinical settings, a repetitive TMS protocol is frequently utilized, involving the administration of repetitive pulses over a variable timeframe to either excite (high-frequency) or inhibit (low-frequency) brain activity that has been previously implicated in the targeted behaviour or symptom. Repetitive TMS targets enduring biological substrates, including neural oscillations or neurochemistry, impacting not only the target but also its connectivity with other brain regions (Beynel et al., 2020). Integration with neuroimaging insights can be accomplished in ways previously proposed in psychiatric research (Oliver et al., 2022). For instance, ML models that seek to identify predictive features using neuroimaging data, such as the one in **Chapter 7**, can subsequently map the associated neural circuitry and identify the stimulation target. Alternatively, patients can also be stratified for treatment based on biotypes, such as in **Chapter 3**, which share signatures of brain circuit dysfunction. However, TMS targets derived from individual connectivity to a region of interest lack reproducibility to offer an advantage over group-based connectivity due to low signal-to-noise ratio (Fox et al., 2013). Recent trials in depression have addressed this by targeting connectivity to a network, rather than a single heterogeneous region, resulting in robust individualised targets (Cash et al., 2021). Given the application of TMS treatments in reducing fatigue in MS (Korzhova et al., 2019, Gaede et al., 2018), such treatments hold promises of fatigue improvement in IA as well.

The outcomes of this thesis offer insights for guiding animal experiments. Studies could delve into how manipulating equivalent dysfunctional brain circuits in animals, like the mesolimbic pathway in motivational fatigue, impacts animal behaviour using molecular techniques such as optogenetics and chemogenetics. Similar attempts at reverse translation

research have already been undertaken in psychiatric conditions to refine neuromodulation research and elucidate fundamental questions about their mechanisms of action (Rudebeck et al., 2019). Furthermore, methods like resting-state FC and DTI in humans can be similarly employed in animals to estimate and compare connections. This approach was used to reverse-translate a satiation network from humans to mice, which then unveiled the molecular foundations of the network (Low et al., 2021). While considerably more challenging, this approach could also be extended to investigate a complex, internal behaviour like fatigue. Interoceptive models of fatigue in animals are already in use, achieved by employing immunological stimulation to induce sickness-like behaviours (Lasselin, 2021). The findings from such pre-clinical work can then pinpoint pertinent molecular mediators and examine their temporal interrelationships. These mediators may well offer valuable insights for rational drug development.

Finally, although the work here failed to find a transdiagnostic fatigue network, future studies would need to adopt such an approach to address the multi-causal nature of fatigue. Heterogenous findings have confounded neuroimaging studies of fatigue in IA and chronic fatigue conditions overall, but this can be counteracted via a long-term strategy in design and applications. Longitudinal neuroimaging data would need to pinpoint specific CNS mechanisms of IA fatigue, probe pharmacological bottom-up and cognitive top-down interventions and assess the impact of factors such as centralised pain, low mood, and cognitive deficits that contribute to fatigue. Such steps can hopefully then make way for personalised management of fatigue in IA.

Contributions

- RA study dataset 1: conceived aims and hypotheses (Chapter 3), constructed methodology based on previous studies, conducted MRI preprocessing and analysis.
- LIFT dataset: conceived aims and hypotheses (Chapters 3, 4, 5, 7), constructed methodologies to answer the different research questions, conducted different MRI preprocessing and analysis according to aim.
- PsA study dataset: MRI data collection, clinical data cleaning, constructed methodology (Chapter 6), conducted preprocessing and analysis.
- RA study dataset 2: constructed methodology (Chapter 6), conducted MRI preprocessing and analysis.

Contributions to journals

- STEFANOV, K., AL-WASITY, S., PARKINSON, J. T., WAITER, G. D., CAVANAGH, J. & BASU, N. 2023. Brain mapping inflammatory-arthritis-related fatigue in the pursuit of novel therapeutics. *Lancet Rheumatology*, 5, E99-E109.
- DEHSARVI, A., AL-WASITY, S., STEFANOV, K., WISEMAN, S. J., RALSTON, S. H., WARDLAW, J. M., EMSLEY, R., BACHMAIR, E.-M., CAVANAGH, J., WAITER, G. D. & BASU, N. 2023. Characterising the neurobiological mechanisms of action of exercise and cognitive behavioural interventions for rheumatoid arthritis fatigue: an MRI brain study. *Arthritis & rheumatology* (Hoboken, N.J.).

Conference proceedings/abstracts

- STEFANOV, K., SUNZINI, F., AL-WASITY, S., HARTE, S., HARRIS, R., CLAUW, D., WAITER, G., CAVANAGH, J. & BASU, N. Fibromyalgia associates with pain-promoting and inhibitory functional connectivity of the default mode network in psoriatic arthritis. *2023 ISMRM & ISMRT Annual Meeting & Exhibition*, 2023 Toronto (Canada).
- STEFANOV, K., AL-WASITY, S., WAITER, G., DEHSARVI, A., BACHMAIR, E.-M., CAVANAGH, J. & BASU, N. Multimodal brain correlates of treatment

response for fatigue interventions in rheumatoid arthritis. *OHBM 2022 Annual Meeting*, 2022 Glasgow (United Kingdom).

- STEFANOV, K., AL-WASITY, S., WAITER, G., CAVANAGH, J. & BASU, N. Dynamic connectivity and current fatigue levels stratify rheumatoid arthritis patients into distinct subgroups. *2020 SINAPSE ASM*, 2020 Virtual Meeting.
- STEFANOV, K., AL-WASITY, S., WAITER, G., CAVANAGH, J. & BASU, N. Dynamic connectivity and current fatigue levels stratify rheumatoid arthritis patients into distinct subgroups. *FENS 2020 Virtual Forum*, 2020.

Appendices

Appendix A CONN toolbox default atlas regions

NO.	REGION
CORTICAL REGIONS	
1	Frontal Pole Right/Left
2	Insular Cortex Right/Left
3	Superior Frontal Gyrus Right/Left
4	Middle Frontal Gyrus Right/Left
5	Inferior Frontal Gyrus, pars triangularis Right/Left
6	Inferior Frontal Gyrus, pars opercularis Right/Left
7	Precentral Gyrus Right/Left
8	Temporal Pole Right/Left
9	Superior Temporal Gyrus, anterior division Right/Left
10	Superior Temporal Gyrus, posterior division Right/Left
11	Middle Temporal Gyrus, anterior division Right/Left
12	Middle Temporal Gyrus, posterior division Right/Left
13	Middle Temporal Gyrus, temporooccipital part Right/Left
14	Inferior Temporal Gyrus, anterior division Right/Left
15	Inferior Temporal Gyrus, posterior division Right/Left
16	Inferior Temporal Gyrus, temporooccipital part Right/Left
17	Postcentral Gyrus Right/Left
18	Superior Parietal Lobule Right/Left
19	Supramarginal Gyrus, anterior division Right/Left
20	Supramarginal Gyrus, posterior division Right/Left
21	Angular Gyrus Right/Left
22	Lateral Occipital Cortex, superior division Right/Left
23	Lateral Occipital Cortex, inferior division Right/Left
24	Intracalcarine Cortex Right/Left
25	Frontal Medial Cortex
26	Juxtapositional Lobule Cortex -formerly Supplementary Motor Cortex- Right/Left
27	Subcallosal Cortex
28	Paracingulate Gyrus Right/Left
29	Cingulate Gyrus, anterior division
30	Cingulate Gyrus, posterior division
31	Precuneous Cortex
32	Cuneal Cortex Right/Left
33	Frontal Orbital Cortex Right/Left
34	Parahippocampal Gyrus, anterior division Right/Left
35	Parahippocampal Gyrus, posterior division Right/Left
36	Lingual Gyrus Right/Left
37	Temporal Fusiform Cortex, anterior division Right/Left
38	Temporal Fusiform Cortex, posterior division Right/Left
39	Temporal Occipital Fusiform Cortex Right/Left
40	Occipital Fusiform Gyrus Right/Left
41	Frontal Operculum Cortex Right/Left
42	Central Opercular Cortex Right/Left
43	Parietal Operculum Cortex Right/Left
44	Planum Polare Right/Left
45	Heschl's Gyrus Right/Left
46	Planum Temporale Right/Left
47	Supracalcarine Cortex Right/Left
48	Occipital Pole Right/Left
SUBCORTICAL REGIONS	
1	Thalamus Right/Left
2	Caudate Right/Left
3	Putamen Right/Left
4	Pallidum Right/Left
5	Hippocampus Right/Left
6	Amygdala Right/Left
7	Accumbens Right/Left

8	Brain-Stem
CEREBELLAR REGIONS	
1	Cerebellum Crus1 Right/Left
2	Cerebellum Crus2 Right/Left
3	Cerebellum 3 Right/Left
4	Cerebellum 4 5 Right/Left
5	Cerebellum 6 Right/Left
6	Cerebellum 7b Right/Left
7	Cerebellum 8 Right/Left
8	Cerebellum 9 Right/Left
9	Cerebellum 10 Right/Left
10	Vermis 1 2
11	Vermis 3
12	Vermis 4 5
13	Vermis 6
14	Vermis 7
15	Vermis 8
16	Vermis 9
17	Vermis 10

Appendix B BRAF-MDQ

Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire V2 05/09/12

Dimension	Questions	Range	Score	
Physical	1	NRS (numerical rating scale) fatigue	0-10	
	2	How many days?	0-7	
	3	How long on average has each episode of fatigue lasted?	0-2	
	4	Have you lacked physical energy because of fatigue?	0-3	
	Physical total (0-22)			
Living	5	Has fatigue made it difficult to bath or shower?	0-3	
	6	Has fatigue made it difficult to dress yourself?	0-3	
	7	Has fatigue made it difficult to do your work or other daily activities?	0-3	
	8	Have you avoided making plans because of fatigue?	0-3	
	9	Has fatigue affected your social life?	0-3	
	10	Have you cancelled plans because of fatigue?	0-3	
	11	Have you refused invitations because of fatigue?	0-3	
Living total (0-21)				
Cognition	12	Have you lacked mental energy because of fatigue?	0-3	
	13	Have you forgotten things because of fatigue?	0-3	
	14	Has fatigue made it difficult to think clearly?	0-3	
	15	Has fatigue made it difficult to concentrate?	0-3	
	16	Have you made mistakes because of fatigue?	0-3	
Cognition total (0-15)				
Emotion	17	Have you felt you have less control because of fatigue?	0-3	
	18	Have you felt embarrassed because of fatigue?	0-3	
	19	Has being fatigued upset you?	0-3	
	20	Have you felt down or depressed because of fatigue?	0-3	
Emotion total (0-12)				
BRAF-MDQ Total score (0-70)				

- Questions 1 and 2 are compulsory.
- Only 1 question may be missing from each dimension (maximum of 3 overall). Replace the missing question score with the average score for that dimension.
- For the Physical Fatigue dimension, a weighted average score is used to account for the varying item score ranges: Total the 3 completed scores, divide by the total max possible score for those 3 questions, then multiply by the maximum score possible for all 4 questions (22).

Appendix C Chalder Fatigue Severity

We would like to know more about any problems you have had with feeling tired, weak or lacking in energy in the last month. Please answer ALL the questions by ticking the answer which applies to you most closely. If you have been feeling tired for a long while, then compare yourself to how you felt when you were last well. Please tick only one box per line.

	less than usual	no more than usual	more than usual	much more than usual
do you have problems with tiredness?				
do you need to rest more?				
do you feel sleepy or drowsy?				
do you have problems starting things?				
do you lack energy?				
do you have less strength in your muscles?				
do you feel weak?				
do you have difficulties concentrating?				
do you make slips of the tongue when speaking?				
do you find it more difficult to find the right word?				
	better than usual	no worse than usual	worse than usual	much worse than usual
how is your memory?				

- The questionnaire is scored in “Likert” style 0, 1, 2 & 3 with a range from 0 to 33.
- The first 10 questions are scored as less than usual (0), no more than usual (1), more than usual (2), much more than usual (3).
- The last question is scored as better than usual (0), no worse than usual (1), worse than usual (2), much worse than usual (3).
- The total score (0-33) is the sum of every question.

Appendix D Non-pharmacological predictors

Predictors in both treatment groups (PEP and CBA) using change scores

Corrected for multiple comparisons using false discovery rate (FDR) at $p < 0.05$ except structural connectivity, which is corrected at the $p < 0.025$ level. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 . The design and contrast matrix are: covName= ['subj' 'gender' 'age' 'site' 'change score']; C= [0 0 0 0 1]

Baseline Structural connectivity to predict future fatigue (Chalder) in 48 patients with either CBA or PEP in the LIFT dataset:

Seed region	Target region	t statistic	p value	Effect size
Left Precuneus	Left Caudal Anterior Cingulate	-4.1	p=0.015	0.28 (medium)
	Left Superior Frontal Gyrus	-3.83	p=0.017	0.26 (medium)
Right Frontal Pole	Left Posterior Cingulate	-3.95	p=0.024	0.27 (medium)
Left Parahippocampal	Left Temporal Pole	-3.94	p=0.024	0.27 (medium)
Right Rostral Middle Frontal Gyrus	Right Supramarginal Gyrus	-3.94	p=0.025	0.27 (medium)

Baseline Resting-state functional connectivity to predict future fatigue (Chalder) in 48 patients with either CBA or PEP treatments in the LIFT dataset: Nothing significant ($p < 0.05$ level)

Baseline PASAT functional connectivity to predict future fatigue (Chalder) in 43 patients with either CBA or PEP in the LIFT dataset:

Seed region	Target region	t statistic	p value	Effect size
Left Cuneus	Right Pericalcarine	3.91	p=0.03	0.28 (medium)

Baseline grey matter volumes to predict future fatigue (Chalder) in 48 patients with either CBA or PEP treatments in the LIFT dataset: Nothing significant ($p < 0.05$ level)

Correlates of baseline fatigue in all patients

Corrected for multiple comparisons using false discovery rate (FDR) at $p < 0.05$ except structural connectivity, which is corrected at the $p < 0.025$ level. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 . The design and contrast matrix are: covName= ['subj' 'gender' 'age' 'site' 'baseline']; C= [0 0 0 0 1]

Baseline Structural connectivity to associate with baseline fatigue (Chalder) in 87 patients in the LIFT dataset: Nothing significant ($p < 0.025$ level)

Baseline resting-state functional connectivity to associate with baseline fatigue (Chalder) in 87 patients in the LIFT dataset:

Seed region	Target region	t statistic	p value	Effect size
Right Hippocampus	Right Pars Orbitalis	4.07	$p = 0.009$	0.17 (small)
	Left Pars Orbitalis	3.78	$p = 0.012$	0.15 (small)
	Left Pars Triangularis	3.55	$p = 0.018$	0.13 (small)
Left Precuneus	Right Cuneus	3.65	$p = 0.023$	0.14 (small)
	Left Cuneus	3.59	$p = 0.023$	0.14 (small)

Baseline PASAT functional connectivity to associate with baseline fatigue (Chalder) in 82 patients in the LIFT dataset: Nothing significant ($p < 0.05$ level)

Baseline grey matter volumes to associate with baseline fatigue (Chalder) in 88 patients in the LIFT dataset: Nothing significant ($p < 0.05$ level)

Sensitivity analysis (individual group predictors)

Corrected for multiple comparisons using false discovery rate (FDR) at $p < 0.05$ except structural connectivity, which is corrected at the $p < 0.025$ level. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 . The design and contrast matrix for adjusting for baseline fatigue are: covName= ['subj' 'gender' 'age' 'site' 'baseline' 'follow-up score']; C= [0 0 0 0 0 1]. The design and contrast matrix for change score are: covName= ['subj' 'gender' 'age' 'site' 'change score']; Change score: C= [0 0 0 0 1].

Group 1 (Personalised exercise programme [PEP])

Baseline structural connectivity to predict future fatigue (Chalder) in 25 patients (PEP group) in the LIFT dataset:

- Adjusted for baseline fatigue: Nothing significant ($p < 0.025$ level)
- Using change scores: Nothing significant ($p < 0.025$ level)

Baseline resting-state functional connectivity to predict future fatigue (Chalder) in 25 patients (PEP group) in the LIFT dataset:

- Adjusted for baseline fatigue:

Seed region	Target region	t statistic	p value	Effect size
Right Inferior Temporal Gyrus	Right Paracentral Lobule	-4.14	$p = 0.046$	0.47 (large)

- Using change scores:

Seed region	Target region	t statistic	p value	Effect size
Left Supramarginal Gyrus	Right Paracentral Lobule	-4.15	$p = 0.041$	0.48 (large)

Baseline PASAT functional connectivity to predict future fatigue (Chalder) in 22 patients (PEP group) in the LIFT dataset:

- Adjusted for baseline fatigue: Nothing significant ($p < 0.05$ level)
- Using change scores: Nothing significant ($p < 0.05$ level)

Baseline grey matter volumes to predict future fatigue (Chalder) in 22 patients (PEP group) in the LIFT dataset:

- Adjusted for baseline fatigue: Nothing significant ($p < 0.05$ level)
- Using change scores: Nothing significant ($p < 0.05$ level)

Group 3 (Cognitive-behavioural approach [CBA])

Baseline structural connectivity to predict future fatigue (Chalder) in 23 patients (CBA group) in the LIFT dataset:

- Adjusted for baseline fatigue:

Seed region	Target region	t statistic	p value	Effect size
Left Entorhinal	Left Lateral Occipital Gyrus	-5.19	$p=0.006$	0.61 (large)
	Left Caudate	-4.43	$p=0.015$	0.54 (large)
Left Pars Triangularis	Left Pericalcarine	-5.08	$p=0.008$	0.6 (large)
Right Pars Triangularis	Right Fusiform	-4.73	$p=0.016$	0.57 (large)
Left Caudal Anterior Cingulate	Left Precuneus	-4.37	$p=0.021$	0.53 (large)
	Left Lingual	-4.28	$p=0.021$	0.52 (large)

- Using change scores:

Seed region	Target region	t statistic	p value	Effect size
Left Pars Triangularis	Left Pericalcarine	-4.69	$p=0.015$	0.57 (large)
Left Entorhinal	Left Lateral Occipital Gyrus	-4.64	$p=0.017$	0.57 (large)

Baseline resting-state functional connectivity to predict future fatigue (Chalder) in 23 patients (CBA group) in the LIFT dataset:

- Adjusted for baseline fatigue:

Seed region	Target region	t statistic	p value	Effect size
Left Superior Frontal Gyrus	Right Inferior Temporal Gyrus	4.62	$p=0.02$	0.56 (large)
Right Amygdala	Left Rostral Middle Frontal Gyrus	4.53	$p=0.025$	0.55 (large)
Right Medial Orbitofrontal Cortex	Right Frontal Pole	-4.46	$p=0.029$	0.54 (large)

- Using change scores:

Seed region	Target region	t statistic	p value	Effect size
Right Amygdala	Left Rostral Middle Frontal Gyrus	4.36	p=0.03	0.52 (large)

Baseline PASAT functional connectivity to predict future fatigue (Chalder) in 21 patients (CBA group) in the LIFT dataset:

- Adjusted for baseline fatigue:

Seed region	Target region	t statistic	p value	Effect size
Left Paracentral Lobule	Right Pars Opercularis	5.47	p=0.005	0.67 (large)

- Using change scores: Nothing significant (p<0.05 level)

Baseline grey matter volumes to predict future fatigue (Chalder) in 23 patients (CBA group) in the LIFT dataset:

- Adjusted for baseline fatigue: Nothing significant (p<0.05 level)
- Using change scores: Nothing significant (p<0.05 level)

Appendix E PROMIS Fatigue-FM Profile

No.	In the past 7 days...	Not at all	A little bit	Somewhat	Quite a bit	Very much
E1	How tired did you feel on average?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
E2	How fatigued were you on average?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
E3	How exhausted were you on average?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
S1	To what degree did fatigue interfere with your social activities?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
S2	To what degree did fatigue interfere with your recreational activities?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
M1	To what degree did you have trouble starting things because of fatigue?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
M2	To what degree did you have trouble finishing things because of fatigue?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
C1	To what degree did fatigue make it difficult to make decisions?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
C2	To what degree did fatigue make you feel slowed down in your thinking?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
	In the past 7 days...	Never	Rarely	Sometimes	Often	Always
M3	How often were you less effective at home due to fatigue?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
M4	How often did you have to push yourself to get things done because of your fatigue?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
S3	How often did you have to limit your social activities because of fatigue?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
S4	How often were you too tired to socialize with your friends?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
C3	How often were you too tired to think clearly?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
C4	How often did fatigue make you more forgetful?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
	In the past 7 days...	None	Mild	Moderate	Severe	Very Severe
E4	What was the level of your fatigue on most days?	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

- The No notes the short form dimensions for Experience (E), Social Impact (S), Motivational Impact (M), and Cognitive Impact (C)
- The total score is the sum of all the individual scores on each question

Appendix F Pharmacological predictors

Predictors of fatigue after DMARD treatment using change scores

Corrected for multiple comparisons using false discovery rate (FDR) at $p < 0.05$. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 . The design and contrast matrix are: covName= ['subj' 'gender' 'age' 'site' 'change score']; C= [0 0 0 0 1]

- No significant predictors of PROMIS fatigue 3 months after starting DMARD treatment in PsA.
- Predictors of PROMIS fatigue 6 months after starting DMARD treatment in PsA

Grey Matter Thickness

ROI Seed	t statistic	p value	Effect size
Right Pericalcarine	4.29	0.009	0.35 (medium)
Left Pericalcarine	3.66	0.028	0.28 (medium)

Grey Matter Volume

Left Pericalcarine	3.98	0.028	0.32 (medium)
Right Pericalcarine	3.69	0.032	0.28 (medium)

Resting-state Functional Connectivity

ROI Seed	ROI Target	t statistic	p value	Effect size
Right Bank of the Superior Temporal Sulcus	Left Caudate	-4.49	0.006	0.36 (medium)
	Right Cuneus	4.33	0.005	0.35 (medium)
	Right Lingual Gyrus	4.21	0.005	0.34 (medium)
Right Caudal Anterior Cingulate	Left Lingual Gyrus	4.16	0.016	0.33 (medium)
	Left Cuneus	3.97	0.028	0.31 (medium)

- No significant predictors of PROMIS fatigue 3 months after starting DMARD treatment in RA.

Correlates of baseline fatigue in PsA and RA

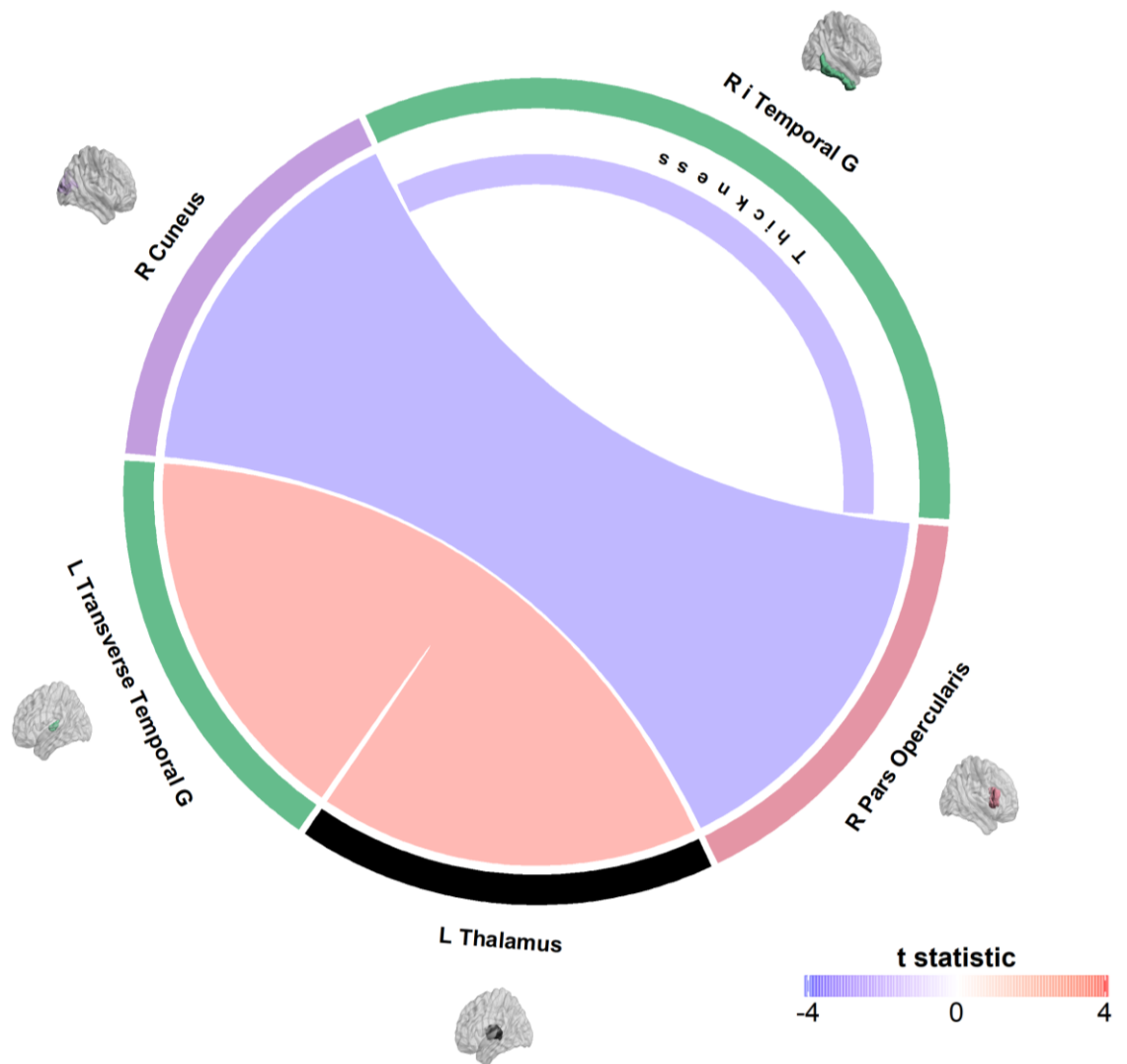
Correlates of baseline fatigue in PsA patients are depicted in the table and figure below. Fatigue was negative associated with grey matter thickness of the right inferior temporal gyrus as well as connectivity between the visual right cuneus and the frontal right pars opercularis. Higher fatigue at baseline corresponded to higher connectivity between the left thalamus and the left transverse temporal gyrus. In the 25 RA patients, higher baseline fatigue was associated with weaker functional connectivity between the left caudal middle frontal gyrus and the right inferior parietal lobule ($t(21) = -4.06$, FDR $p = 0.46$, eta squared effect size = 0.44 (large)).

Grey Matter Thickness

ROI Seed		t statistic	p value	Effect size
Right Inferior Temporal Gyrus		-3.67	0.046	0.24 (small)

Resting-state Functional Connectivity				
ROI Seed	ROI Target	t statistic	p value	Effect size
Left Thalamus	Left Transverse Temporal Gyrus	3.92	0.026	0.26 (medium)
Right Cuneus	Right Pars Opercularis	-3.85	0.033	0.26 (medium)

Significant results ($p < 0.05$) using general linear models on baseline grey matter volume, thickness, surface area and resting-state single functional connectivity variables to associate with baseline PROMIS fatigue in patients of the CENTAUR study ($n=47$). Seed region indicates brain region whose grey matter properties or connections were tested for predicting fatigue while controlling for multiple comparisons using false discovery rate. Target region indicates which connection with the seed region was significant. The results are after controlling for confounds (age, gender) for connectivity and total intracranial volume for grey matter measures. Effect sizes are eta squared labelled as trivial < 0.1 , small ≥ 0.1 , medium ≥ 0.25 , large > 0.37 (Goss-Sampson, 2019).



The figure displays the resting-state connectivity and grey matter thickness associations with baseline PROMIS fatigue. The strength (t statistic) and direction of the association is reflected in the colour of the connection or semicircle of the structural measure, with blue indicating the higher the metric the lower the fatigue while red displays metrics that are high when fatigue is high. All analyses used a general linear model, adjusted for age and gender while the structural measures were additionally corrected for intracranial volume. Abbreviations: L/R, left/right; G, gyrus.

Appendix G Machine learning configuration

Modality	Predicted R ²	Algorithm	Outcome	FC	Feature selection	
					Relief	SGCCA
Clinical	0.197	glmnet	6 months	N/A	1:1	50%
FC	0.119	svmLinear	1 year	PASAT	1:1	70%
Multimodal	0.104	svmRadial	6 months	Resting-state	1:4	70%
SC	0.101	gbm	1 year	N/A	1:4	50%
Morphometric	0.107	svmLinear	1 year	N/A	1:2	50%

Each row represents the configuration for the best-performing model in the respective modality, determined by the highest predicted R². The top-performing models employed an elastic net (glmnet), a support vector regression with a linear (svmLinear) or radial (svmRadial) kernel or a gradient boosting (gbm) algorithm. The configuration includes the timing of the clinical outcome, which can be either six months or one year after treatment allocation. The feature selection column specifies whether all positive features (1:1) were included post RRelief F, or if it was narrowed down to the top half (1:2) or quarter (1:4) of positive features. It also indicates whether 50% or 70% of resamples had to include the features during the stability selection for the sparse generalized canonical correlation analysis (SGCCA). In specific cases of functional connectivity (FC) and multimodal models, the selection of the source of FC was an additional consideration. The models could use either FC estimated from fMRI during resting-state or from the Paced Auditory Serial Addition Test (PASAT) task.

Modality	MAE	Algorithm	Outcome	FC	Feature selection	
					Relief	SGCCA
Clinical	5.01	glmnet	6 months	N/A	1:1	50%
FC	5.35	gaussprRadial	6 months	Resting-state	1:1	50%
Multimodal	5.45	gaussprRadial	6 months	Resting-state	1:4	70%
SC	5.51	gaussprRadial	1 year	N/A	1:4	50%
Morphometric	5.45	rf	6 months	N/A	1:1	50%

Each row represents the configuration for the best-performing model in the respective modality, determined by the lowest predicted mean absolute error (MAE). The top-performing models employed an elastic net (glmnet), a Gaussian process regression with a radial kernel (gaussprRadial) or a gradient boosting (gbm) algorithm. The configuration includes the timing of the clinical outcome, which can be either six months or one year after treatment allocation. The feature selection column specifies whether all positive features (1:1) were included post RRelief F, or if it was narrowed down to the top half or quarter

(1:4) of positive features. It also indicates whether 50% or 70% of resamples had to include the features during the stability selection for the sparse generalized canonical correlation analysis (SGCCA). In specific cases of functional connectivity (FC) and multimodal models, the selection of the source of FC was an additional consideration. The models could use either FC estimated from fMRI during resting-state or from the Paced Auditory Serial Addition Test (PASAT) task.

Appendix H Best performing features

- Best-performing multimodal model using resting-state functional connectivity (rsFC)

NO.	FEATURE
1	Usual Care Group
2	CBA Group
3	Depression
4	Disease duration
5	rsFC RH Caudate - LH Parahippocampal
6	rsFC LH Bank of the Superior Temporal Sulcus - RH Middle Temporal Gyrus
7	rsFC LH Entorhinal - LH Insula
8	rsFC LH Medial Orbitofrontal - RH Bank of the Superior Temporal Sulcus
9	rsFC LH Parahippocampal - RH Bank of the Superior Temporal Sulcus
10	rsFC LH Paracentral - RH Lateral Orbitofrontal
11	rsFC LH Pars Triangularis - RH Pars Orbitalis
12	rsFC LH Posterior Cingulate - RH Lateral Orbitofrontal
13	rsFC LH Pallidum - LH Pars Orbitalis
14	rsFC LH Frontal Pole - LH Insula
15	rsFC LH Hippocampus - RH Hippocampus
16	rsFC LH Hippocampus - RH Amygdala
17	rsFC RH Bank of the Superior Temporal Sulcus - RH Parahippocampal
18	rsFC RH Caudal Anterior Cingulate - RH Lateral Orbitofrontal
19	rsFC RH Cuneus - RH Pars Triangularis
20	rsFC RH Lateral Orbitofrontal - RH Superior Frontal Gyrus
21	rsFC RH Pars Orbitalis - RH Precuneus
22	rsFC LH Accumbens - RH Pars Triangularis
23	rsFC LH Accumbens - RH Posterior Cingulate
24	rsFC LH Accumbens - RH Supramarginal Gyrus
25	rsFC LH Accumbens - RH Transverse Temporal Gyrus
26	rsFC RH Thalamus - RH Entorhinal
27	SC LH Inferior Parietal Lobule - LH Isthmus Cingulate

28	SC LH Isthmus Cingulate - LH Caudal Anterior Cingulate
29	SC LH Isthmus Cingulate - LH Accumbens
30	SC LH Medial Orbitofrontal - LH Frontal Pole
31	SC LH Posterior Cingulate - RH Rostral Anterior Cingulate
32	SC LH Precuneus - LH Isthmus Cingulate
33	SC LH Precuneus - RH Rostral Anterior Cingulate
34	SC LH Supramarginal Gyrus - RH Rostral Anterior Cingulate
35	SC RH Bank of the Superior Temporal Sulcus - RH Rostral Middle Frontal Gyrus
36	SC RH Medial Orbitofrontal - LH Postcentral
37	SC RH Paracentral - RH Caudal Anterior Cingulate
38	SC RH Precuneus - LH Medial Orbitofrontal
39	SC RH Precuneus - LH Rostral Anterior Cingulate
40	SC RH Precuneus - RH Caudal Anterior Cingulate
41	SC RH Precuneus - RH Rostral Anterior Cingulate
42	SC RH Superior Parietal Lobule - RH Caudal Middle Frontal Gyrus
43	Thickness LH Postcentral
44	Volume LH Temporal Pole

- Best-performing multimodal model using PASAT functional connectivity (PFC)

NO.	FEATURE
1	Usual Care Group
2	CBA Group
3	Anxiety
4	PFC RH Caudate - LH Frontal Pole
5	PFC RH Caudate - RH Frontal Pole
6	PFC RH Pallidum - RH Ventral Diencephalon
7	PFC RH Pallidum - RH Rostral Anterior Cingulate
8	PFC RH Accumbens - RH Supramarginal Gyrus
9	PFC RH Accumbens - RH Frontal Pole
10	PFC LH Bank of the Superior Temporal Sulcus - RH Entorhinal
11	PFC LH Caudal Anterior Cingulate - LH Frontal Pole
12	PFC LH Caudate - LH Posterior Cingulate
13	PFC LH Cuneus - LH Precentral
14	PFC LH Cuneus - LH Rostral Middle Frontal Gyrus
15	PFC LH Cuneus - LH Temporal Pole
16	PFC LH Cuneus - RH Caudal Anterior Cingulate
17	PFC LH Cuneus - RH Transverse Temporal Gyrus
18	PFC LH Entorhinal - RH Transverse Temporal Gyrus
19	PFC LH Inferior Parietal Lobule - LH Paracentral
20	PFC LH Inferior Parietal Lobule - LH Rostral Anterior Cingulate
21	PFC LH Inferior Parietal Lobule - RH Caudal Anterior Cingulate

22	PFC LH Lateral Occipital Gyrus - LH Temporal Pole
23	PFC LH Lateral Occipital Gyrus - RH Temporal Pole
24	PFC LH Medial Orbitofrontal - RH Medial Orbitofrontal
25	PFC LH Medial Orbitofrontal - RH Frontal Pole
26	PFC LH Putamen - LH Bank of the Superior Temporal Sulcus
27	PFC LH Putamen - LH Superior Parietal Lobule
28	PFC LH Putamen - RH Entorhinal
29	PFC LH Putamen - RH Lateral Occipital Gyrus
30	PFC LH Paracentral - LH Frontal Pole
31	PFC LH Paracentral - RH Rostral Anterior Cingulate
32	PFC LH Paracentral - RH Rostral Middle Frontal Gyrus
33	PFC LH Pars Orbitalis - LH Rostral Middle Frontal Gyrus
34	PFC LH Pericalcarine - LH Rostral Middle Frontal Gyrus
35	PFC LH Pericalcarine - RH Entorhinal
36	PFC LH Postcentral - LH Insula
37	PFC LH Postcentral - RH Inferior Temporal Gyrus
38	PFC LH Pallidum - LH Rostral Anterior Cingulate
39	PFC LH Precuneus - RH Precuneus
40	PFC LH Rostral Anterior Cingulate - LH Frontal Pole
41	PFC LH Rostral Anterior Cingulate - RH Medial Orbitofrontal
42	PFC LH Rostral Anterior Cingulate - RH Transverse Temporal Gyrus
43	PFC LH Rostral Middle Frontal Gyrus - RH Pars Orbitalis
44	PFC LH Superior Temporal Gyrus - LH Temporal Pole
45	PFC LH Superior Temporal Gyrus - RH Entorhinal
46	PFC LH Superior Temporal Gyrus - RH Temporal Pole
47	PFC LH Frontal Pole - LH Insula
48	PFC LH Frontal Pole - RH Lateral Orbitofrontal
49	PFC LH Temporal Pole - RH Inferior Temporal Gyrus
50	PFC LH Temporal Pole - RH Precentral
51	PFC LH Temporal Pole - RH Frontal Pole
52	PFC LH Hippocampus - LH Superior Parietal Lobule
53	PFC LH Insula - RH Entorhinal
54	PFC RH Caudal Anterior Cingulate - RH Inferior Parietal Lobule
55	PFC RH Caudal Anterior Cingulate - RH Superior Parietal Lobule
56	PFC RH Cuneus - RH Rostral Anterior Cingulate
57	PFC RH Entorhinal - RH Middle Temporal Gyrus
58	PFC RH Entorhinal - RH Pericalcarine
59	PFC RH Entorhinal - RH Postcentral
60	PFC RH Entorhinal - RH Precentral
61	PFC LH Amygdala - LH Medial Orbitofrontal
62	PFC LH Amygdala - LH Rostral Anterior Cingulate
63	PFC LH Amygdala - RH Inferior Temporal Gyrus

64	PFC LH Amygdala - RH Medial Orbitofrontal
65	PFC RH Lateral Orbitofrontal - RH Rostral Anterior Cingulate
66	PFC RH Paracentral - RH Frontal Pole
67	PFC LH Accumbens - RH Lingual
68	PFC RH Postcentral - RH Frontal Pole
69	PFC RH Precentral - RH Frontal Pole
70	PFC RH Precentral - RH Temporal Pole
71	PFC RH Rostral Anterior Cingulate - RH Transverse Temporal Gyrus
72	PFC LH Ventral Diencephalon - RH Supramarginal Gyrus
73	PFC RH Frontal Pole - RH Transverse Temporal Gyrus
74	PFC RH Thalamus - RH Pericalcarine
75	SC LH Thalamus - RH Middle Temporal Gyrus
76	SC RH Caudate - RH Ventral Diencephalon
77	SC RH Caudate - LH Lateral Occipital Gyrus
78	SC RH Caudate - RH Pericalcarine
79	SC RH Hippocampus - LH Inferior Temporal Gyrus
80	SC RH Amygdala - LH Frontal Pole
81	SC RH Accumbens - LH Rostral Middle Frontal Gyrus
82	SC RH Accumbens - LH Superior Frontal Gyrus
83	SC RH Ventral Diencephalon - RH Temporal Pole
84	SC LH Caudal Anterior Cingulate - LH Precuneus
85	SC LH Caudal Anterior Cingulate - LH Rostral Anterior Cingulate
86	SC LH Caudal Anterior Cingulate - RH Precuneus
87	SC LH Caudal Anterior Cingulate - RH Insula
88	SC LH Caudal Middle Frontal Gyrus - RH Superior Temporal Gyrus
89	SC LH Caudate - LH Superior Frontal Gyrus
90	SC LH Cuneus - LH Temporal Pole
91	SC LH Cuneus - RH Inferior Temporal Gyrus
92	SC LH Entorhinal - LH Lateral Occipital Gyrus
93	SC LH Entorhinal - LH Parahippocampal
94	SC LH Fusiform - LH Entorhinal
95	SC LH Inferior Parietal Lobule - LH Caudate
96	SC LH Inferior Parietal Lobule - LH Isthmus Cingulate
97	SC LH Inferior Parietal Lobule - LH Putamen
98	SC LH Inferior Parietal Lobule - LH Pars Triangularis
99	SC LH Inferior Parietal Lobule - RH Middle Temporal Gyrus
100	SC LH Inferior Parietal Lobule - RH Pars Orbitalis
101	SC LH Inferior Parietal Lobule - RH Precuneus
102	SC LH Inferior Parietal Lobule - RH Thalamus
103	SC LH Isthmus Cingulate - LH Caudal Anterior Cingulate
104	SC LH Isthmus Cingulate - LH Precuneus
105	SC LH Isthmus Cingulate - RH Medial Orbitofrontal

106	SC LH Isthmus Cingulate - LH Accumbens
107	SC LH Isthmus Cingulate - RH Rostral Anterior Cingulate
108	SC LH Lateral Occipital Gyrus - LH Entorhinal
109	SC LH Lateral Occipital Gyrus - LH Amygdala
110	SC LH Lateral Orbitofrontal - LH Amygdala
111	SC LH Lateral Orbitofrontal - RH Lateral Occipital Gyrus
112	SC LH Lateral Orbitofrontal - RH Middle Temporal Gyrus
113	SC LH Lingual - LH Isthmus Cingulate
114	SC LH Lingual - LH Temporal Pole
115	SC LH Medial Orbitofrontal - LH Frontal Pole
116	SC LH Medial Orbitofrontal - LH Insula
117	SC LH Putamen - LH Caudate
118	SC LH Putamen - LH Rostral Middle Frontal Gyrus
119	SC LH Middle Temporal Gyrus - LH Posterior Cingulate
120	SC LH Middle Temporal Gyrus - LH Superior Frontal Gyrus
121	SC LH Middle Temporal Gyrus - RH Paracentral
122	SC LH Paracentral - LH Rostral Anterior Cingulate
123	SC LH Paracentral - LH Accumbens
124	SC LH Paracentral - RH Rostral Anterior Cingulate
125	SC LH Pars Orbitalis - RH Postcentral
126	SC LH Pars Triangularis - LH Middle Temporal Gyrus
127	SC LH Pars Triangularis - LH Pallidum
128	SC LH Pericalcarine - LH Lateral Orbitofrontal
129	SC LH Pericalcarine - LH Temporal Pole
130	SC LH Postcentral - RH Insula
131	SC LH Posterior Cingulate - LH Precuneus
132	SC LH Posterior Cingulate - RH Rostral Anterior Cingulate
133	SC LH Precuneus - LH Caudal Anterior Cingulate
134	SC LH Precuneus - LH Isthmus Cingulate
135	SC LH Precuneus - RH Caudal Anterior Cingulate
136	SC LH Precuneus - RH Medial Orbitofrontal
137	SC LH Precuneus - RH Rostral Anterior Cingulate
138	SC LH Rostral Anterior Cingulate - RH Precuneus
139	SC LH Rostral Middle Frontal Gyrus - LH Putamen
140	SC LH Superior Frontal Gyrus - RH Lateral Orbitofrontal
141	SC LH Superior Parietal Lobule - LH Supramarginal Gyrus
142	SC LH Supramarginal Gyrus - LH Isthmus Cingulate
143	SC LH Supramarginal Gyrus - LH Rostral Anterior Cingulate
144	SC LH Supramarginal Gyrus - RH Rostral Anterior Cingulate
145	SC LH Frontal Pole - LH Pars Orbitalis
146	SC LH Frontal Pole - RH Pars Triangularis
147	SC LH Frontal Pole - RH Pericalcarine

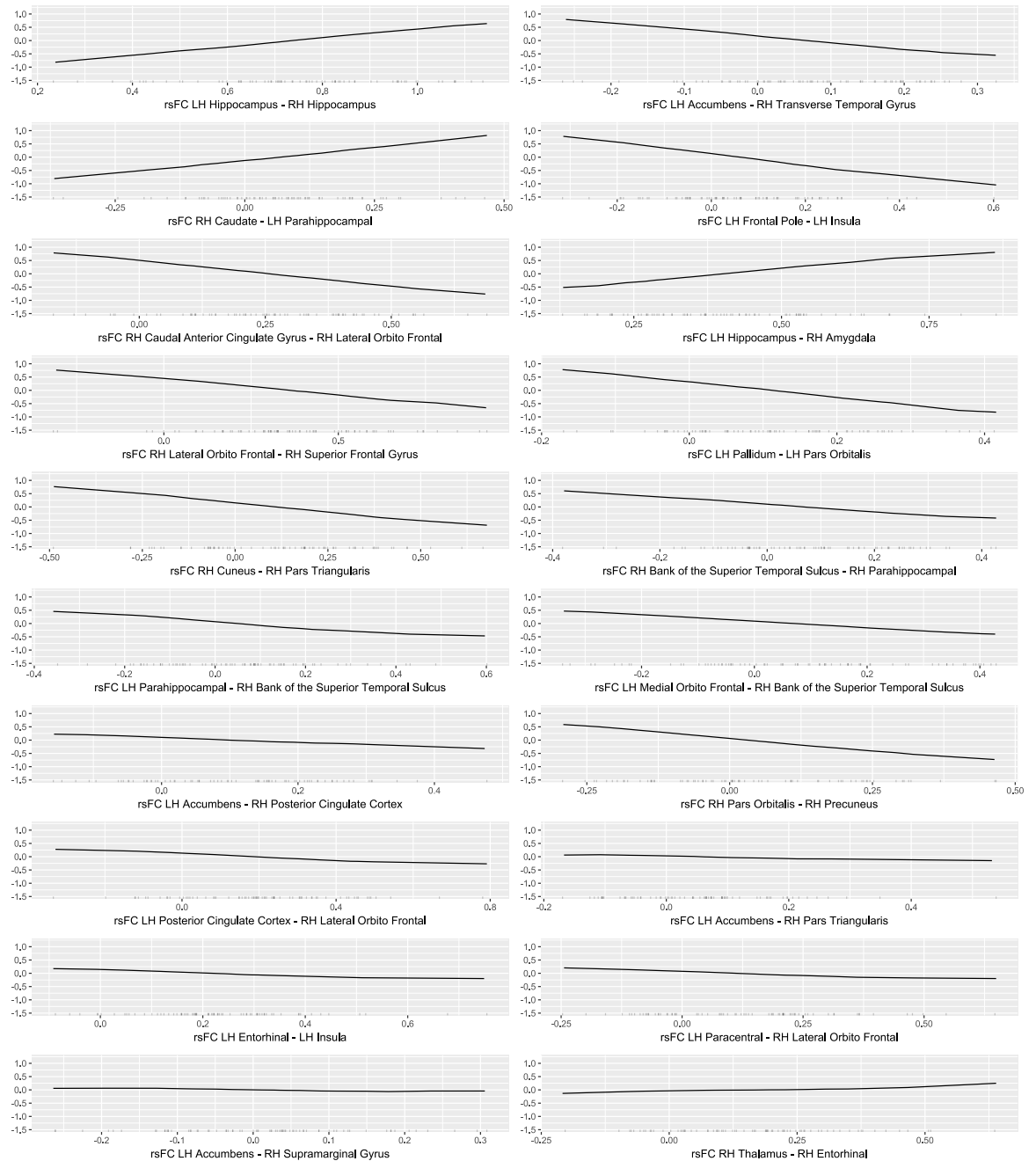
148	SC LH Frontal Pole - RH Postcentral
149	SC LH Temporal Pole - LH Lingual
150	SC LH Insula - LH Paracentral
151	SC RH Bank of the Superior Temporal Sulcus - RH Pallidum
152	SC RH Bank of the Superior Temporal Sulcus - RH Ventral Diencephalon
153	SC RH Bank of the Superior Temporal Sulcus - RH Caudal Middle Frontal Gyrus
154	SC RH Bank of the Superior Temporal Sulcus - RH Inferior Parietal Lobule
155	SC RH Bank of the Superior Temporal Sulcus - RH Rostral Middle Frontal Gyrus
156	SC RH Caudal Anterior Cingulate - LH Fusiform
157	SC RH Caudal Anterior Cingulate - LH Isthmus Cingulate
158	SC RH Caudal Anterior Cingulate - RH Precuneus
159	SC RH Caudal Middle Frontal Gyrus - RH Bank of the Superior Temporal Sulcus
160	SC RH Cuneus - RH Pars Orbitalis
161	SC RH Entorhinal - LH Isthmus Cingulate
162	SC RH Entorhinal - LH Lateral Orbitofrontal
163	SC RH Entorhinal - LH Accumbens
164	SC RH Fusiform - RH Pallidum
165	SC RH Inferior Parietal Lobule - LH Rostral Anterior Cingulate
166	SC RH Inferior Parietal Lobule - RH Inferior Temporal Gyrus
167	SC RH Inferior Temporal Gyrus - RH Pallidum
168	SC RH Inferior Temporal Gyrus - LH Caudate
169	SC RH Inferior Temporal Gyrus - LH Fusiform
170	SC RH Inferior Temporal Gyrus - LH Transverse Temporal Gyrus
171	SC RH Isthmus Cingulate - RH Medial Orbitofrontal
172	SC RH Isthmus Cingulate - LH Accumbens
173	SC RH Lateral Occipital Gyrus - LH Inferior Temporal Gyrus
174	SC RH Lateral Occipital Gyrus - LH Precuneus
175	SC RH Lateral Orbitofrontal - LH Inferior Temporal Gyrus
176	SC RH Lingual - RH Inferior Temporal Gyrus
177	SC RH Medial Orbitofrontal - RH Amygdala
178	SC RH Medial Orbitofrontal - LH Isthmus Cingulate
179	SC RH Medial Orbitofrontal - LH Paracentral
180	SC RH Medial Orbitofrontal - LH Postcentral
181	SC RH Medial Orbitofrontal - LH Precentral
182	SC RH Medial Orbitofrontal - LH Temporal Pole
183	SC RH Middle Temporal Gyrus - LH Pericalcarine
184	SC RH Middle Temporal Gyrus - RH Cuneus
185	SC RH Middle Temporal Gyrus - RH Isthmus Cingulate
186	SC RH Middle Temporal Gyrus - RH Thalamus
187	SC RH Paracentral - LH Rostral Anterior Cingulate
188	SC RH Paracentral - LH Superior Parietal Lobule
189	SC RH Paracentral - RH Caudal Anterior Cingulate

190	SC RH Paracentral - RH Lateral Orbitofrontal
191	SC RH Paracentral - RH Posterior Cingulate
192	SC RH Pars Triangularis - RH Superior Temporal Gyrus
193	SC LH Accumbens - LH Superior Parietal Lobule
194	SC LH Accumbens - RH Pericalcarine
195	SC LH Accumbens - RH Rostral Middle Frontal Gyrus
196	SC RH Pericalcarine - LH Frontal Pole
197	SC RH Pericalcarine - LH Temporal Pole
198	SC RH Postcentral - LH Rostral Anterior Cingulate
199	SC RH Postcentral - LH Frontal Pole
200	SC RH Posterior Cingulate - LH Middle Temporal Gyrus
201	SC RH Posterior Cingulate - RH Precuneus
202	SC RH Precuneus - LH Caudal Anterior Cingulate
203	SC RH Precuneus - LH Medial Orbitofrontal
204	SC RH Precuneus - LH Rostral Anterior Cingulate
205	SC RH Precuneus - LH Superior Frontal Gyrus
206	SC RH Precuneus - RH Caudal Anterior Cingulate
207	SC RH Precuneus - RH Medial Orbitofrontal
208	SC RH Precuneus - RH Rostral Anterior Cingulate
209	SC RH Precuneus - RH Frontal Pole
210	SC RH Rostral Middle Frontal Gyrus - RH Bank of the Superior Temporal Sulcus
211	SC RH Superior Parietal Lobule - LH Lateral Occipital Gyrus
212	SC RH Superior Parietal Lobule - RH Caudal Middle Frontal Gyrus
213	SC RH Superior Temporal Gyrus - RH Pars Triangularis
214	SC RH Superior Temporal Gyrus - RH Postcentral
215	SC LH Ventral Diencephalon - LH Superior Frontal Gyrus
216	SC LH Ventral Diencephalon - LH Temporal Pole
217	SC RH Supramarginal Gyrus - LH Rostral Anterior Cingulate
218	SC RH Temporal Pole - RH Caudate
219	SC RH Temporal Pole - RH Putamen
220	SC RH Insula - RH Postcentral
221	SC RH Thalamus - RH Middle Temporal Gyrus
222	Surface Area LH Medial Orbitofrontal
223	Surface Area LH Pericalcarine
224	Surface Area RH Medial Orbitofrontal
225	Surface Area RH Pars opercularis
226	Thickness LH Inferior Temporal Gyrus
227	Thickness LH Isthmus Cingulate
228	Thickness LH Postcentral
229	Thickness RH Middle Temporal Gyrus
230	Thickness RH Pars Triangularis
231	Thickness RH Postcentral

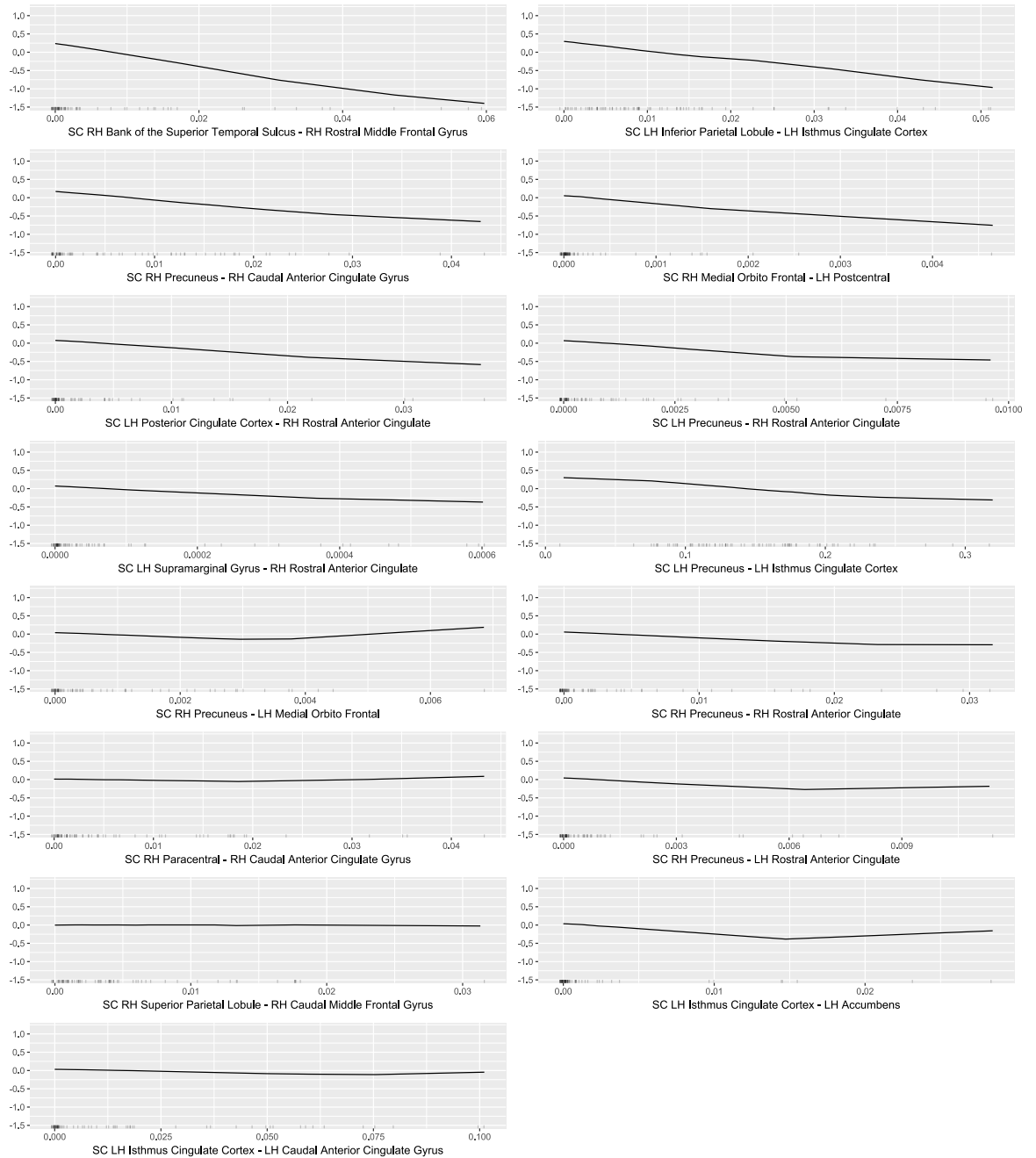
232	Thickness RH Rostral Middle Frontal Gyrus
233	Volume LH Lateral Orbitofrontal
234	Volume LH Temporal Pole
235	Volume RH Fusiform
236	Volume RH Parahippocampal
237	Volume RH Postcentral

Appendix I Feature effect on fatigue

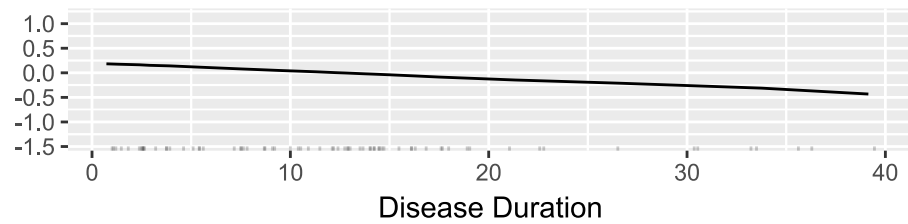
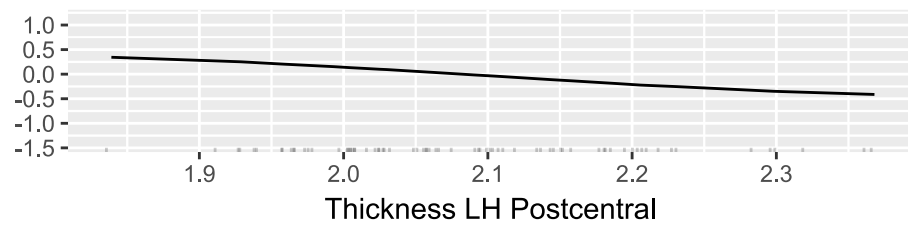
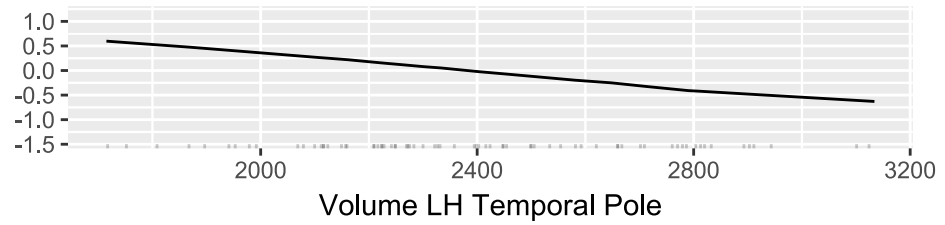
- Accumulated local effects (ALE) plots of the resting-state functional connectivity (rsFC) features in the best-performing multimodal model. The plots display the main effects of the features on the prediction of Chalder Fatigue after six months. The y-axes represent the size of the mean effect each feature has on the fatigue prediction relative to the overall model prediction.



- Accumulated local effects (ALE) plots of the structural connectivity (SC) features in the best-performing multimodal model. The plots display the main effects of the features on the prediction of Chalder Fatigue after six months. The y-axis represents the size of the mean effect each feature has on the fatigue prediction relative to the overall model prediction.



- Accumulated local effects (ALE) plots of the morphometric and clinical features in the best-performing multimodal model. The plots display the main effects of the features on the prediction of Chalder Fatigue after six months. The y-axes represent the size of the mean effect each feature has on the fatigue prediction relative to the overall model prediction.



List of references

- ABOSHIHA, A., GALLAGHER, R. & GARGAN, L. 2019. Chasing Value as AI Transforms Health Care. BCG.
- ALAA, A. M., BOLTON, T., DI ANGELANTONIO, E., RUDD, J. H. F. & VAN DER SCHAAR, M. 2019. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *Plos One*, 14.
- ALAMANOS, Y., VOULGARI, P. V. & DROSOS, A. A. 2006. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: A systematic review. *Seminars in Arthritis and Rheumatism*, 36, 182-188.
- ALLEN, E. A., DAMARAJU, E., PLIS, S. M., ERHARDT, E. B., EICHELE, T. & CALHOUN, V. D. 2014. Tracking Whole-Brain Connectivity Dynamics in the Resting State. *Cerebral Cortex*, 24, 663-676.
- ALLISON, P. D. 1990. Change scores as dependent variables in regression analysis. *Sociological methodology*, 93-114.
- ALMEIDA, C., CHOY, E. H. S., HEWLETT, S., KIRWAN, J. R., CRAMP, F., CHALDER, T., POLLOCK, J. & CHRISTENSEN, R. 2016. Biologic interventions for fatigue in rheumatoid arthritis. *Cochrane Database of Systematic Reviews*.
- ALMODOVAR, R., ZARCO, P., OTON, T. & CARMONA, L. 2018. Effect of weight loss on activity in psoriatic arthritis: A systematic review. *Reumatologia Clinica*, 14, 207-210.
- ALMUTAIRI, K., NOSSENT, J., PREEN, D., KEEN, H. & INDERJEETH, C. 2021. The global prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review. *Rheumatology International*, 41, 863-877.
- ALTMAN, D. G. & ROYSTON, P. 2006. Statistics notes - The cost of dichotomising continuous variables. *British Medical Journal*, 332, 1080-1080.
- ALTMANN, A., TOLOSI, L., SANDER, O. & LENGAUER, T. 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26, 1340-1347.
- AMBROSEN, K. S., FREDRIKSSON, F., ANHOJ, S., BAK, N., VAN DELLEN, E., DOMINICUS, L., LEMVIGH, C. K., SORENSEN, M. E., NIELSEN, M. O., BOJESEN, K. B., FAGERLUND, B., GLENTHOJ, B. Y., ORANJE, B., HANSEN, L. K. & EBDRUP, B. H. 2023. Clustering of antipsychotic-naive patients with schizophrenia based on functional connectivity from resting-state electroencephalography. *European Archives of Psychiatry and Clinical Neuroscience*.

- AMROLLAHI, F., SHASHIKUMAR, S. P., HOLDER, A. L. & NEMAT, S. 2022. Leveraging clinical data across healthcare institutions for continual learning of predictive risk models. *Scientific Reports*, 12.
- ANDERSON, J. S., FERGUSON, M. A., LOPEZ-LARSON, M. & YURGELUN-TODD, D. 2011. Reproducibility of Single-Subject Functional Connectivity Measurements. *American Journal of Neuroradiology*, 32, 548-555.
- ANGELES FERNANDEZ-GIL, M., PALACIOS-BOTE, R., LEO-BARAHONA, M. & MORA-ENCINAS, J. P. 2010. Anatomy of the Brainstem: A Gaze Into the Stem of Life. *Seminars in Ultrasound Ct and Mri*, 31, 196-219.
- ANTERAPER, S. A., NIETO-CASTANON, A. & WHITFIELD-GABRIELI, S. 2020. Functional MRI Methods. *Neuroimaging in Schizophrenia*. Cham: Springer International Publishing.
- ANTONENKO, D., HAYEK, D., NETZBAND, J., GRITNER, U. & FLOEEL, A. 2019. tDCS-induced episodic memory enhancement and its association with functional network coupling in older adults. *Scientific Reports*, 9.
- APLEY, D. W. & ZHU, J. 2020. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 82, 1059-1086.
- ASTER, H.-C., EVDOKIMOV, D., BRAUN, A., UCEYLER, N., KAMPF, T., PHAM, M., HOMOLA, G. A. & SOMMER, C. 2022. CNS imaging characteristics in fibromyalgia patients with and without peripheral nerve involvement. *Scientific Reports*, 12.
- AUDOIN, B., IBARROLA, D., DUONG, M. V. A., PELLETIER, J., CONFORT-GOUNY, S., MALIKOVA, I., ALI-CHÉRIF, A., COZZONE, P. J. & RANJEVA, J. P. 2005. Functional MRI study of PASAT in normal subjects. *Magnetic Resonance Materials in Physics Biology and Medicine*, 18, 96-102.
- AUSTIN, P. C. & BRUNNER, L. J. 2004. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Statistics in Medicine*, 23, 1159-1178.
- AZUR, M. J., STUART, E. A., FRANGAKIS, C. & LEAF, P. J. 2011. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20, 40-49.
- BACHMAIR, E.-M., MARTIN, K., AUCOTT, L., DHAUN, N., DURES, E., EMSLEY, R., GRAY, S. R., KIDD, E., KUMAR, V. & LOVELL, K. 2022. Remotely delivered cognitive behavioural and personalised exercise interventions for fatigue severity and impact in inflammatory rheumatic diseases (LIFT): a multicentre, randomised, controlled, open-label, parallel-group trial. *The Lancet Rheumatology*.
- BALACHANDAR, R., JOHN, J. P., SAINI, J., KUMAR, K. J., JOSHI, H., SADANAND, S., AIYAPPAN, S., SIVAKUMAR, P. T., LOGANATHAN, S., VARGHESE, M. &

- BHARATH, S. 2015. A study of structural and functional connectivity in early Alzheimer's disease using rest fMRI and diffusion tensor imaging. *International Journal of Geriatric Psychiatry*, 30, 497-504.
- BALL, T. M., SQUEGLIA, L. M., TAPERT, S. F. & PAULUS, M. P. 2020. Double Dipping in Machine Learning: Problems and Solutions. *Biological Psychiatry-Cognitive Neuroscience and Neuroimaging*, 5, 261-263.
- BARAN, T. M., ZHANG, Z., ANDERSON, A. J., MCDERMOTT, K. & LIN, F. 2020. Brain structural connectomes indicate shared neural circuitry involved in subjective experience of cognitive and physical fatigue in older adults. *Brain Imaging and Behavior*, 14, 2488-2499.
- BARRACLOUGH, M., MCKIE, S., PARKER, B., ELLIOTT, R. & BRUCE, I. N. 2021. The effects of disease activity, inflammation, depression and cognitive fatigue on resting state fMRI in systemic lupus erythematosus. *Rheumatology (Oxford, England)*.
- BARRACLOUGH, M., MCKIE, S., PARKER, B., JACKSON, A., PEMBERTON, P., ELLIOTT, R. & BRUCE, I. N. 2019. Altered cognitive function in systemic lupus erythematosus and associations with inflammation and functional and structural brain changes. *Annals of the Rheumatic Diseases*, 78, 934-940.
- BARTH, B., MAYER-CARIUS, K., STREHL, U., KELAVA, A., HAEUSSINGER, F. B., FALLGATTER, A. J. & EHLIS, A.-C. 2018. Identification of neurophysiological biotypes in attention deficit hyperactivity disorder. *Psychiatry and Clinical Neurosciences*, 72, 836-848.
- BASTA, M., SIMOS, N. J., ZIOGA, M., ZAGANAS, I., PANAGIOTAKIS, S., LIONIS, C. & VGONTZAS, A. N. 2023. Personalized screening and risk profiles for Mild Cognitive Impairment via a Machine Learning Framework: Implications for general practice. *International Journal of Medical Informatics*, 170.
- BASU, I., YOUSEFI, A., CROCKER, B., ZELMANN, R., PAULK, A. C., PELED, N., ELLARD, K. K., WEISHOLTZ, D. S., COSGROVE, G. R., DECKERSBACH, T., EDEN, U. T., ESKANDAR, E. N., DOUGHERTY, D. D., CASH, S. S. & WIDGE, A. S. 2021. Closed-loop enhancement and neural decoding of cognitive control in humans. *Nature Biomedical Engineering*.
- BASU, N., JONES, G. T., MACFARLANE, G. J. & DRUCE, K. L. 2017. Identification and Validation of Clinically Relevant Clusters of Severe Fatigue in Rheumatoid Arthritis. *Psychosomatic Medicine*, 79, 1051-1058.
- BASU, N., KAPLAN, C. M., ICHESCO, E., LARKIN, T., HARRIS, R. E., MURRAY, A., WAITER, G. & CLAUW, D. J. 2018. Neurobiologic Features of Fibromyalgia Are Also Present Among Rheumatoid Arthritis Patients. *Arthritis & Rheumatology*, 70, 1000-1007.
- BASU, N., KAPLAN, C. M., ICHESCO, E., LARKIN, T., SCHREPF, A., MURRAY, A. D., CLAUW, D. J., WAITER, G. D. & HARRIS, R. E. 2019. Functional and

structural magnetic resonance imaging correlates of fatigue in patients with rheumatoid arthritis. *Rheumatology (Oxford, England)*, 58, 1822-1830.

- BASU, N., MURRAY, A. D., JONES, G. T., REID, D. M., MACFARLANE, G. J. & WAITER, G. D. 2014. Neural correlates of fatigue in granulomatosis with polyangiitis: a functional magnetic resonance imaging study. *Rheumatology*, 53, 2080-2087.
- BEATTY, W. W., GORETTI, B., SIRACUSA, G., ZIPOLI, V., PORTACCIO, E. & AMATO, M. P. 2003. Changes in neuropsychological test performance over the workday in multiple sclerosis. *Clinical Neuropsychologist*, 17, 551-560.
- BECHMAN, K., YATES, M. & GALLOWAY, J. B. 2019. The new entries in the therapeutic armamentarium: The small molecule JAK inhibitors. *Pharmacological Research*, 147.
- BECKER, S. & SCHWEINHARDT, P. 2012. Dysfunctional Neurotransmitter Systems in Fibromyalgia, Their Role in Central Stress Circuitry and Pharmacological Actions on These Systems. *Pain Research and Treatment*, 2012.
- BECKMANN, C. F., DELUCA, M., DEVLIN, J. T. & SMITH, S. M. 2005. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 1001-1013.
- BEHRENS, T. E. J., BERG, H. J., JBABDI, S., RUSHWORTH, M. F. S. & WOOLRICH, M. W. 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34, 144-155.
- BEHZADI, Y., RESTOM, K., LIAU, J. & LIU, T. T. 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, 37, 90-101.
- BENEDEK, M., JAUK, E., BEATY, R. E., FINK, A., KOSCHUTNIG, K. & NEUBAUER, A. C. 2016. Brain mechanisms associated with internally directed attention and self-generated thought. *Scientific Reports*, 6.
- BENGTSSON, S. L., LAU, H. C. & PASSINGHAM, R. E. 2009. Motivation to do Well Enhances Responses to Errors and Self-Monitoring. *Cerebral Cortex*, 19, 797-804.
- BENJAMINI, Y. & HOCHBERG, Y. 1995. CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 57, 289-300.
- BERNER, C., ERLACHER, L., FENZL, K. H. & DORNER, T. E. 2018. A cross-sectional study on self-reported physical and mental health-related quality of life in rheumatoid arthritis and the role of illness perception. *Health and Quality of Life Outcomes*, 16.

- BETZEL, R. F., FUKUSHIMA, M., HE, Y., ZUO, X.-N. & SPORNS, O. 2016. Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks. *Neuroimage*, 127, 287-297.
- BEYNEL, L., POWERS, J. P. & APPELBAUM, L. G. 2020. Effects of repetitive transcranial magnetic stimulation on resting-state connectivity: A systematic review. *Neuroimage*, 211.
- BISECCO, A., DI NARDO, F., DOCIMO, R., CAIAZZO, G., D'AMBROSIO, A., BONAVITA, S., CAPUANO, R., SINISI, L., CIRILLO, M., ESPOSITO, F., TEDESCHI, G. & GALLO, A. 2018. Fatigue in multiple sclerosis: The contribution of resting-state functional connectivity reorganization. *Multiple Sclerosis Journal*, 24, 1696-1705.
- BISWAL, B., YETKIN, F. Z., HAUGHTON, V. M. & HYDE, J. S. 1995. FUNCTIONAL CONNECTIVITY IN THE MOTOR CORTEX OF RESTING HUMAN BRAIN USING ECHO-PLANAR MRI. *Magnetic Resonance in Medicine*, 34, 537-541.
- BOISSONEAULT, J., LETZEN, J., LAI, S., O'SHEA, A., CRAGGS, J., ROBINSON, M. E. & STAUD, R. 2016. Abnormal resting state functional connectivity in patients with chronic fatigue syndrome: an arterial spin-labeling fMRI study. *Magnetic Resonance Imaging*, 34, 603-608.
- BOKMA, W. A., ZHUTOVSKY, P., GILTAY, E. J., SCHOEVERS, R. A., PENNINX, B. W. J. H., VAN BALKOM, A. L. J. M., BATELAAN, N. M. & VAN WINGEN, G. A. 2022. Predicting the naturalistic course in anxiety disorders using clinical and biological markers: a machine learning approach. *Psychological Medicine*, 52, 57-67.
- BOLT, T., NOMI, J. S., RUBINOV, M. & UDDIN, L. Q. 2017. Correspondence Between Evoked and Intrinsic Functional Brain Network Configurations. *Human Brain Mapping*, 38, 1992-2007.
- BOMMERT, A., SUN, X., BISCHL, B., RAHNENFUEHRER, J. & LANG, M. 2020. Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis*, 143.
- BOOMERSHINE, C. S. 2015. Fibromyalgia: The Prototypical Central Sensitivity Syndrome. *Current Rheumatology Reviews*, 11, 131-145.
- BOS, D. J., ORANJE, B., ACHTERBERG, M., VLASKAMP, C., AMBROSINO, S., DE REUS, M. A., VAN DEN HEUVEL, M. P., ROMBOUTS, S. A. R. B. & DURSTON, S. 2017. Structural and functional connectivity in children and adolescents with and without attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, 58, 810-818.
- BOTVINIK-NEZER, R., HOLZMEISTER, F., CAMERER, C. F., DREBER, A., HUBER, J., JOHANNESSON, M., KIRCHLER, M., IWANIR, R., MUMFORD, J. A., ADCOCK, R. A., AVESANI, P., BACZKOWSKI, B. M., BAJRACHARYA, A., BAKST, L., BALL, S., BARILARI, M., BAULT, N., BEATON, D., BEITNER, J.,

- BENOIT, R. G., BERKERS, R. M. W. J., BHANJI, J. P., BISWAL, B. B., BOBADILLA-SUAREZ, S., BORTOLINI, T., BOTTENHORN, K. L., BOWRING, A., BRAEM, S., BROOKS, H. R., BRUDNER, E. G., CALDERON, C. B., CAMILLERI, J. A., CASTRELLON, J. J., CECCHETTI, L., CIESLIK, E. C., COLE, Z. J., COLLIGNON, O., COX, R. W., CUNNINGHAM, W. A., CZOSCHKE, S., DADI, K., DAVIS, C. P., LUCA, A. D., DELGADO, M. R., DEMETRIOU, L., DENNISON, J. B., DI, X., DICKIE, E. W., DOBRYAKOVA, E., DONNAT, C. L., DUKART, J., DUNCAN, N. W., DURNEZ, J., EED, A., EICKHOFF, S. B., ERHART, A., FONTANESI, L., FRICKE, G. M., FU, S., GALVAN, A., GAU, R., GENON, S., GLATARD, T., GLEREAN, E., GOEMAN, J. J., GOLOWIN, S. A. E., GONZALEZ-GARCIA, C., GORGOLEWSKI, K. J., GRADY, C. L., GREEN, M. A., GUASSI MOREIRA, J. F., GUEST, O., HAKIMI, S., HAMILTON, J. P., HANCOCK, R., HANDJARAS, G., HARRY, B. B., HAWCO, C., HERHOLZ, P., HERMAN, G., HEUNIS, S., HOFFSTAEDTER, F., HOGEVEEN, J., HOLMES, S., HU, C.-P., HUETTEL, S. A., HUGHES, M. E., IACOVELLA, V., IORDAN, A. D., ISAGER, P. M., ISIK, A. I., JAHN, A., JOHNSON, M. R., JOHNSTONE, T., JOSEPH, M. J. E., JULIANO, A. C., KABLE, J. W., KASSINOPOULOS, M., KOBAYASHI, C., KONG, X.-Z., et al. 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582, 84-+.
- BROWN, G. G., MATHALON, D. H., STERN, H., FORD, J., MUELLER, B., GREVE, D. N., MCCARTHY, G., VOYVODIC, J., GLOVER, G., DIAZ, M., YETTER, E., OZYURT, I. B., JORGENSEN, K. W., WIBLE, C. G., TURNER, J. A., THOMPSON, W. K., POTKIN, S. G. & FUNCTION BIOMED INFORMATICS RES, N. 2011. Multisite reliability of cognitive BOLD data. *Neuroimage*, 54, 2163-2175.
- BRUMMETT, C. M. & CLAUW, D. J. 2011. Fibromyalgia: a primer for the anesthesia community. *Current Opinion in Anesthesiology*, 24, 532-539.
- BUS, B. A. A., MOLENDIJK, M. L., PENNINX, B. W. J. H., BUITELAAR, J. K., PRICKAERTS, J., ELZINGA, B. M. & VOSHAAR, R. C. O. 2014. Low serum BDNF levels in depressed patients cannot be attributed to individual depressive symptoms or symptom cluster. *World Journal of Biological Psychiatry*, 15, 561-569.
- CAPONE, F., COLLORONE, S., CORTESE, R., DI LAZZARO, V. & MOCCIA, M. 2020. Fatigue in multiple sclerosis: The role of thalamus. *Multiple Sclerosis Journal*, 26, 6-16.
- CASH, F. H. R., WEIGAND, A., ZALESKY, A., SIDDIQI, H. S., DOWNAR, J., FITZGERALD, B. P. & FOX, D. M. 2021. Using Brain Imaging to Improve Spatial Targeting of Transcranial Magnetic Stimulation for Depression. *Biological Psychiatry*, 90, 689-700.
- CHAHAL, R., GOTLIB, I. H. & GUYER, A. E. 2020. Research Review: Brain network connectivity and the heterogeneity of depression in adolescence - a precision mental health perspective. *Journal of Child Psychology and Psychiatry*, 61, 1282-1298.
- CHAI, X. J., CASTANON, A. N., OENQUER, D. & WHITFIELD-GABRIELI, S. 2012. Anticorrelations in resting state networks without global signal regression. *Neuroimage*, 59, 1420-1428.

- CHALAH, M. A., RIACHI, N., AHDAB, R., CREANGE, A., LEFAUCHEUR, J.-P. & AYACHE, S. S. 2015. Fatigue in Multiple Sclerosis: Neural Correlates and the Role of Non-Invasive Brain Stimulation. *Frontiers in Cellular Neuroscience*, 9.
- CHANG, C. & GLOVER, G. H. 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage*, 50, 81-98.
- CHAO, G., SUN, S. & BI, J. 2021. A Survey on Multi-View Clustering. *IEEE transactions on artificial intelligence*, 2, 146-168.
- CHARRAD, M., GHAZZALI, N., BOITEAU, V. & NIKNAFS, A. 2014. Nbclust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61, 1-36.
- CHEN, C.-L., HWANG, T.-J., TUNG, Y.-H., YANG, L.-Y., HSU, Y.-C., LIU, C.-M., LIN, Y.-T., HSIEH, M.-H., LIU, C.-C., CHIEN, Y.-L., HWU, H.-G. & TSENG, W.-Y. I. 2022a. Detection of advanced brain aging in schizophrenia and its structural underpinning by using normative brain age metrics. *Neuroimage-Clinical*, 34.
- CHEN, J., PATIL, K. R., YEO, B. T. T. & EICKHOFF, S. B. 2023. Leveraging Machine Learning for Gaining Neurobiological and Nosological Insights in Psychiatric Research. *Biological Psychiatry*, 93, 18-28.
- CHEN, J., TAM, A., KEBETS, V., ORBAN, C., OOI, L. Q. R., ASPLUND, C. L., MAREK, S., DOSENBACH, N. U. F., EICKHOFF, S. B., BZDOK, D., HOLMES, A. J. & YEO, B. T. T. 2022b. Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nature Communications*, 13.
- CHEN, M. H., DELUCA, J., GENOVA, H. M., YAO, B. & WYLIE, G. R. 2020. Cognitive Fatigue Is Associated with Altered Functional Connectivity in Interoceptive and Reward Pathways in Multiple Sclerosis. *Diagnostics*, 10.
- CHEN, Z. S., KULKARNI, P. P., GALATZER-LEVY, I. R., BIGIO, B., NASCA, C. & ZHANG, Y. 2022c. Modern views of machine learning for precision psychiatry. *Patterns*, 3.
- CHONG, J. S. X., NG, G. J. P., LEE, S. C. & ZHOU, J. 2017a. Salience network connectivity in the insula is associated with individual differences in interoceptive accuracy. *Brain Structure & Function*, 222, 1635-1644.
- CHONG, T. T. J., APPS, M., GIEHL, K., SILLENCE, A., GRIMA, L. L. & HUSAIN, M. 2017b. Neurocomputational mechanisms underlying subjective valuation of effort costs. *Plos Biology*, 15.
- CHORUS, A. M. J., MIEDEMA, H. S., BOONEN, A. & VAN DER LINDEN, S. 2003. Quality of life and work in patients with rheumatoid arthritis and ankylosing spondylitis of working age. *Annals of the Rheumatic Diseases*, 62, 1178-1184.

- CHURPEK, M. M., YUEN, T. C., WINSLOW, C., MELTZER, D. O., KATTAN, M. W. & EDELSON, D. P. 2016. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Critical Care Medicine*, 44, 368-374.
- CIOLINO, J. D., PALAC, H. L., YANG, A., VACA, M. & BELLI, H. M. 2019. Ideal vs. real: a systematic review on handling covariates in randomized controlled trials. *Bmc Medical Research Methodology*, 19.
- CLERY-MELIN, M.-L., JOLLANT, F. & GORWOOD, P. 2019. Reward systems and cognitions in Major Depressive Disorder. *Cns Spectrums*, 24, 64-77.
- COHEN, J. R. & D'ESPOSITO, M. 2016. The Segregation and Integration of Distinct Brain Networks and Their Relationship to Cognition. *Journal of Neuroscience*, 36, 12083-12094.
- COLE, J. H., POUDEL, R. P. K., TSAGKRASOULIS, D., CAAN, M. W. A., STEVES, C., SPECTOR, T. D. & MONTANA, G. 2017. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage*, 163, 115-124.
- COLE, M. W., BASSETT, D. S., POWER, J. D., BRAVER, T. S. & PETERSEN, S. E. 2014. Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron*, 83, 238-251.
- COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE 2015. Guideline on adjustment for baseline covariates in clinical trials. *London: European Medicines Agency*.
- COOK, D. B., O'CONNOR, P. J., LANGE, G. & STEFFENER, J. 2007. Functional neuroimaging correlates of mental fatigue induced by cognition among chronic fatigue syndrome patients and controls. *Neuroimage*, 36, 108-122.
- COOKSEY, R., RAHMAN, M. A., KENNEDY, J., BROPHY, S. & CHOY, E. 2021. Biologic use in psoriatic arthritis and ankylosing spondylitis patients: a descriptive epidemiological study using linked, routine data in Wales, UK. *Rheumatology advances in practice*, 5, rkab042-rkab042.
- CRAIG, A. D. 2002. How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3, 655-666.
- CRAMP, F., HEWLETT, S., ALMEIDA, C., KIRWAN, J. R., CHOY, E. H. S., CHALDER, T., POLLOCK, J. & CHRISTENSEN, R. 2013. Non-pharmacological interventions for fatigue in rheumatoid arthritis. *Cochrane Database of Systematic Reviews*.
- CRITCHLEY, H. D., WIENS, S., ROTSHTEIN, P., ÖHMAN, A. & DOLAN, R. J. 2004. Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7, 189-195.

- DALE, A. M., FISCHL, B. & SERENO, M. I. 1999. Cortical surface-based analysis - I. Segmentation and surface reconstruction. *Neuroimage*, 9, 179-194.
- DAMOISEAUX, J. S., ROMBOUTS, S. A. R. B., BARKHOF, F., SCHELTENS, P., STAM, C. J., SMITH, S. M. & BECKMANN, C. F. 2006. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 13848-13853.
- DANTZER, R., O'CONNOR, J. C., FREUND, G. G., JOHNSON, R. W. & KELLEY, K. W. 2008. From inflammation to sickness and depression: when the immune system subjugates the brain. *Nature Reviews Neuroscience*, 9, 46-57.
- DAVID, A., PELOSI, A., MCDONALD, E., STEPHENS, D., LEDGER, D., RATHBONE, R. & MANN, A. 1990. TIRED, WEAK, OR IN NEED OF REST - FATIGUE AMONG GENERAL-PRACTICE ATTENDERS. *Bmj-British Medical Journal*, 301, 1199-1202.
- DAVIES, K., DURES, E. & NG, W.-F. 2021. Fatigue in inflammatory rheumatic diseases: current knowledge and areas for future research. *Nature Reviews Rheumatology*.
- DEMISSIE, S., LAVALLEY, M. P., HORTON, N. J., GLYNN, R. J. & CUPPLES, L. A. 2003. Bias due to missing exposure data using complete-case analysis in the proportional hazards regression model. *Statistics in Medicine*, 22, 545-557.
- DESIKAN, R. S., SEGONNE, F., FISCHL, B., QUINN, B. T., DICKERSON, B. C., BLACKER, D., BUCKNER, R. L., DALE, A. M., MAGUIRE, R. P., HYMAN, B. T., ALBERT, M. S. & KILLIANY, R. J. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31, 968-980.
- DHAMALA, E., RONG OOI, L. Q., CHEN, J., RICARD, J. A., BERKELEY, E., CHOPRA, S., QU, Y., ZHANG, X.-H., LAWHEAD, C., YEO, B. T. T. & HOLMES, A. J. 2023a. Brain-Based Predictions of Psychiatric Illness-Linked Behaviors Across the Sexes. *Biological psychiatry*, 94, 479-491.
- DHAMALA, E., YEO, B. T. T. & HOLMES, A. J. 2023b. One Size Does Not Fit All: Methodological Considerations for Brain-Based Predictive Modeling in Psychiatry. *Biological Psychiatry*, 93, 717-728.
- DINGA, R., MARQUAND, A. F., VELTMAN, D. J., BEEKMAN, A. T. F., SCHOEVERS, R. A., VAN HEMERT, A. M., PENNINX, B. W. J. H. & SCHMAAL, L. 2018. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Translational Psychiatry*, 8.
- DINGA, R., SCHMAAL, L., PENNINX, B. W. J. H., VAN TOL, M. J., VELTMAN, D. J., VAN VELZEN, L., MENNES, M., VAN DER WEE, N. J. A. & MARQUAND, A. F. 2019. Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017). *NeuroImage: Clinical*, 22 (no pagination).

- DINNO, A. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *Stata Journal*, 15, 292-300.
- DO ESPIRITO SANTO, R. C., POMPERMAYER, M. G., BINI, R. R., OLSZEWSKI, V., TEIXEIRA, E. G., CHAKR, R., XAVIER, R. M. & BRENOL, C. V. 2018. Neuromuscular fatigue is weakly associated with perception of fatigue and function in patients with rheumatoid arthritis. *Rheumatology International*, 38, 415-423.
- DOUMEN, M., PAZMINO, S., BERTRAND, D., DE COCK, D., JOLY, J., WESTHOVENS, R. & VERSCHUEREN, P. 2022. Longitudinal trajectories of fatigue in early RA: the role of inflammation, perceived disease impact and early treatment response. *Annals of the Rheumatic Diseases*, 81, 1385-1391.
- DRUCE, K. L., AIKMAN, L., DILLEEN, M., BURDEN, A., SZCZYPA, P. & BASU, N. 2018. Fatigue independently predicts different work disability dimensions in etanercept-treated rheumatoid arthritis and ankylosing spondylitis patients. *Arthritis Research & Therapy*, 20.
- DRUCE, K. L., BHATTACHARYA, Y., JONES, G. T., MACFARLANE, G. J. & BASU, N. 2016. Most patients who reach disease remission following anti-TNF therapy continue to report fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Rheumatology*, 55, 1786-1790.
- DRUCE, K. L., JONES, G. T., MACFARLANE, G. J. & BASU, N. 2015. Patients receiving anti-TNF therapies experience clinically important improvements in RA-related fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Rheumatology*, 54, 964-971.
- DRUCE, K. L. & MCBETH, J. 2019. Central sensitization predicts greater fatigue independently of musculoskeletal pain. *Rheumatology*, 58, 1923-1927.
- DRYSDALE, A. T., GROSENICK, L., DOWNAR, J., DUNLOP, K., MANSOURI, F., MENG, Y., FETCHO, R. N., ZEBLEY, B., OATHES, D. J., ETKIN, A., SCHATZBERG, A. F., SUDHEIMER, K., KELLER, J., MAYBERG, H. S., GUNNING, F. M., ALEXOPOULOS, G. S., FOX, M. D., PASCUAL-LEONE, A., VOSS, H. U., CASEY, B. J., DUBIN, M. J. & LISTON, C. 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, 23, 28-38.
- DUNAS, T., WAHLIN, A., NYBERG, L. & BORAXBEBK, C.-J. 2021. Multimodal Image Analysis of Apparent Brain Age Identifies Physical Fitness as Predictor of Brain Maintenance. *Cerebral Cortex*, 31, 3393-3407.
- DUNNE, J., FLORES, M., GAWANDE, R. & SCHUMAN-OLIVIER, Z. 2021. Losing trust in body sensations: Interoceptive awareness and depression symptom severity among primary care patients. *Journal of Affective Disorders*, 282, 1210-1219.
- DURES, E., KITCHEN, K., ALMEIDA, C., AMBLER, N., CLISS, A., HAMMOND, A., KNOPS, B., MORRIS, M., SWINKELS, A. & HEWLETT, S. 2012. "They didn't tell us, they made us work it out ourselves": Patient perspectives of a cognitive-

behavioral program for rheumatoid arthritis fatigue. *Arthritis Care & Research*, 64, 494-501.

DZIURA, J. D., POST, L. A., ZHAO, Q., FU, Z. & PEDUZZI, P. 2013. Strategies for Dealing with Missing Data in Clinical Trials: From Design to Analysis. *Yale Journal of Biology and Medicine*, 86, 343-358.

EGGART, M. & VALDES-STAUBER, J. 2021. Can changes in multidimensional self-reported interoception be considered as outcome predictors in severely depressed patients? A moderation and mediation analysis. *Journal of Psychosomatic Research*, 141.

EIJLERS, A. J. C., WINK, A. M., MEIJER, K. A., DOUW, L., GEURTS, J. J. G. & SCHOONHEIM, M. M. 2019. Reduced Network Dynamics on Functional MRI Signals Cognitive Impairment in Multiple Sclerosis. *Radiology*, 292, 449-457.

ERIKSSON, L., JAWORSKA, J., WORTH, A. P., CRONIN, M. T. D., MCDOWELL, R. M. & GRAMATICA, P. 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives*, 111, 1361-1375.

ERNST, J., NORTHOFF, G., BOEKER, H., SEIFRITZ, E. & GRIMM, S. 2013. Interoceptive awareness enhances neural activity during empathy. *Human Brain Mapping*, 34, 1615-1624.

ESTEVA, A., KUPREL, B., NOVOA, R. A., KO, J., SWETTER, S. M., BLAU, H. M. & THRUN, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115-+.

FELDTHUSEN, C., BJORK, M., FORSBLAD-D'ELIA, H., MANNERKORPI, K. & UNIV GOTHENBURG CTR PERSON-CTR, C. 2013. Perception, consequences, communication, and strategies for handling fatigue in persons with rheumatoid arthritis of working age-a focus group study. *Clinical Rheumatology*, 32, 557-566.

FELGER, J. C., LI, L., MARVAR, P. J., WOOLWINE, B. J., HARRISON, D. G., RAISON, C. L. & MILLER, A. H. 2013. Tyrosine metabolism during interferon-alpha administration: Association with fatigue and CSF dopamine concentrations. *Brain Behavior and Immunity*, 31, 153-160.

FENG, R., GRANA, D. & BALLING, N. 2021. Imputation of missing well log data by random forest and its uncertainty analysis. *Computers & Geosciences*, 152.

FISCHL, B. 2012. FreeSurfer. *Neuroimage*, 62, 774-781.

FISCHL, B., SALAT, D. H., BUSA, E., ALBERT, M., DIETERICH, M., HASELGROVE, C., VAN DER KOUWE, A., KILLIANY, R., KENNEDY, D., KLAVENESS, S., MONTILLO, A., MAKRIS, N., ROSEN, B. & DALE, A. M. 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33, 341-355.

- FISHER, A., RUDIN, C. & DOMINICI, F. 2019. All Models are Wrong, but *<i>Many</i>* are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20.
- FITZGERALD, K. C., MORRIS, B., SOROOSH, A., BALSHI, A., MAHER, D., KAPLIN, A. & NOURBAKHS, B. 2021. Pilot randomized active-placebo-controlled trial of low-dose ketamine for the treatment of multiple sclerosis-related fatigue. *Multiple Sclerosis Journal*, 27, 942-953.
- FLODIN, P., MARTINSEN, S., LOFGREN, M., BILEVICIUTE-LJUNGAR, I., KOSEK, E. & FRANSSON, P. 2014. Fibromyalgia Is Associated with Decreased Connectivity Between Pain- and Sensorimotor Brain Areas. *Brain Connectivity*, 4, 587-594.
- FOROUZANNEZHAD, P., ABBASPOUR, A., FANG, C., CABRERIZO, M., LOEWENSTEIN, D., DUARA, R. & ADJOUADI, M. 2019. A survey on applications and analysis methods of functional magnetic resonance imaging for Alzheimer's disease. *Journal of Neuroscience Methods*, 317, 121-140.
- FOX, M. D., LIU, H. & PASCUAL-LEONE, A. 2013. Identification of reproducible individualized targets for treatment of depression with TMS based on intrinsic connectivity. *Neuroimage*, 66, 151-160.
- FOX, M. D. & RAICHLER, M. E. 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8, 700-711.
- FOX, M. D., SNYDER, A. Z., VINCENT, J. L., CORBETTA, M., VAN ESSEN, D. C. & RAICHLER, M. E. 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 9673-9678.
- FRIEDMAN, L., STERN, H., BROWN, G. G., MATHALON, D. H., TURNER, J., GLOVER, G. H., GOLLUB, R. L., LAURIELLO, J., LIM, K. O., CANNON, T., GREVE, D. N., BOCKHOLT, H. J., BELGER, A., MUELLER, B., DOTY, M. J., HE, J., WELLS, W., SMYTH, P., PIEPER, S., KIM, S., KUBICKI, M., VANGEL, M. & POTKIN, S. G. 2008. Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, 29, 958-972.
- FRISTON, K. J. 1994. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2, 56-78.
- FRISTON, K. J., WILLIAMS, S., HOWARD, R., FRACKOWIAK, R. S. J. & TURNER, R. 1996. Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, 35, 346-355.
- GAEDE, G., TIEDE, M., LORENZ, I., BRANDT, A. U., PFUELLER, C., DOERR, J., BELLMANN-STROBL, J., PIPER, S. K., ROTH, Y., ZANGEN, A., SCHIPPLING, S. & PAUL, F. 2018. Safety and preliminary efficacy of deep transcranial magnetic

stimulation in MS-related fatigue. *Neurology-Neuroimmunology & Neuroinflammation*, 5.

- GALLOWAY, J., CAPRON, J.-P., DE LEONARDIS, F., FAKHOURI, W., ROSE, A., KOURIS, I. & BURKE, T. 2020. The impact of disease severity and duration on cost, early retirement and ability to work in rheumatoid arthritis in Europe: an economic modelling study. *Rheumatology Advances in Practice*, 4.
- GAO, C. X., DWYER, D., ZHU, Y., SMITH, C. L., DU, L., FILIA, K. M., BAYER, J., MENSSINK, J. M., WANG, T., BERGMEIR, C., WOOD, S. & COTTON, S. M. 2023. An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research*, 327.
- GARALI, I., ADANYEGUH, I. M., ICHOU, F., PERLBARG, V., SEYER, A., COLSCH, B., MOSZER, I., GUILLEMOT, V., DURR, A., MOCHEL, F. & TENENHAUS, A. 2018. A strategy for multimodal data integration: application to biomarkers identification in spinocerebellar ataxia. *Briefings in Bioinformatics*, 19, 1356-1369.
- GARFINKEL, S. N., SETH, A. K., BARRETT, A. B., SUZUKI, K. & CRITCHLEY, H. D. 2015. Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, 104, 65-74.
- GELLER, W. N., LIU, K. & WARREN, S. L. 2021. Specificity of anhedonic alterations in resting-state network connectivity and structure: A transdiagnostic approach. *Psychiatry Research-Neuroimaging*, 317.
- GENOVA, H. M., RAJAGOPALAN, V., DELUCA, J., DAS, A., BINDER, A., ARJUNAN, A., CHIARAVALLOTI, N. & WYLIE, G. 2013. Examination of Cognitive Fatigue in Multiple Sclerosis using Functional Magnetic Resonance Imaging and Diffusion Tensor Imaging. *Plos One*, 8.
- GENOVESE, C. R., LAZAR, N. A. & NICHOLS, T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15, 870-878.
- GERGELYFI, M., SANZ-ARIGITA, E. J., SOLOPCHUK, O., DRICOT, L., JACOB, B. & ZENON, A. 2021. Mental fatigue correlates with depression of task-related network and augmented DMN activity but spares the reward circuit. *Neuroimage*, 243.
- GHAFOURI-FARD, S., TAHERI, M., OMRANI, M. D., DAAEE, A., MOHAMMAD-RAHIMI, H. & KAZAZI, H. 2019. Application of Single-Nucleotide Polymorphisms in the Diagnosis of Autism Spectrum Disorders: A Preliminary Study with Artificial Neural Networks. *Journal of Molecular Neuroscience*, 68, 515-521.
- GIANNI, C., BELVISI, D., CONTE, A., TOMMASIN, S., CORTESE, A., PETSAS, N., BAIONE, V., TARTAGLIA, M., MILLEFIORINI, E., BERARDELLI, A. & PANTANO, P. 2021. Altered sensorimotor integration in multiple sclerosis: A combined neurophysiological and functional MRI study. *Clinical Neurophysiology*, 132, 2191-2198.

- GIRKA, F., CAMENEN, E., PELTIER, C., GLOAGUEN, A., GUILLEMOT, V., LE BRUSQUET, L. & TENENHAUS, A. 2023. Multiblock data analysis with the RGCCA package. *Journal of Statistical Software*, 1-36.
- GLINTBORG, B., OSTERGAARD, M., DREYER, L., KROGH, N. S., TARP, U., HANSEN, M. S., RIFBJERG-MADSEN, S., LORENZEN, T. & HETLAND, M. L. 2011. Treatment Response, Drug Survival, and Predictors Thereof in 764 Patients With Psoriatic Arthritis Treated With Anti-Tumor Necrosis Factor alpha Therapy Results From the Nationwide Danish DANBIO Registry. *Arthritis and Rheumatism*, 63, 382-390.
- GLYMOUR, M. M., WEUVE, J., BERKMAN, L. F., KAWACHI, I. & ROBINS, J. M. 2005. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *American Journal of Epidemiology*, 162, 267-278.
- GOERTZ, Y. M. J., BRAAMSE, A. M. J., SPRUIT, M. A., JANSSEN, D. J. A., EBADI, Z., VAN HERCK, M., BURTIN, C., PETERS, J. B., SPRANGERS, M. A. G., LAMERS, F., TWISK, J. W. R., THONG, M. S. Y., VERCOULEN, J. H., GEERLINGS, S. E., VAES, A. W., BEIJERS, R. J. H. C. G., VAN BEERS, M., SCHOLS, A. M. W. J., ROSMALEN, J. G. M. & KNOOP, H. 2021. Fatigue in patients with chronic disease: results from the population-based Lifelines Cohort Study. *Scientific Reports*, 11.
- GONI, M., BASU, N., MURRAY, A. D. & WAITER, G. D. 2018. Neural Indicators of Fatigue in Chronic Diseases: A Systematic Review of MRI Studies. *Diagnostics*, 8.
- GONI, M., BASU, N., MURRAY, A. D. & WAITER, G. D. 2022. Brain predictors of fatigue in rheumatoid arthritis: A machine learning study. *PloS one*, 17, e0269952-e0269952.
- GONZALEZ CAMPO, C., SALAMONE, P. C., RODRIGUEZ-ARRIAGADA, N., RICHTER, F., HERRERA, E., BRUNO, D., PAGANI CASSARA, F., SINAY, V., GARCIA, A. M., IBANEZ, A. & SEDENO, L. 2020. Fatigue in multiple sclerosis is associated with multimodal interoceptive abnormalities. *Multiple Sclerosis Journal*, 26, 1845-1853.
- GONZALEZ-CASTILLO, J. & BANDETTINI, P. A. 2018. Task-based dynamic functional connectivity: Recent findings and open questions. *Neuroimage*, 180, 526-533.
- GOSS-SAMPSON, M. 2019. Statistical analysis in JASP: A guide for students. JASP.
- GOSSEC, L., BARALIAKOS, X., KERSCHBAUMER, A., DE WIT, M., MCINNES, I., DOUGADOS, M., PRIMDAHL, J., MCGONAGLE, D. G., ALETAHA, D., BALANESCU, A., BALINT, P. V., BERTHEUSSEN, H., BOEHNCKE, W.-H., BURMESTER, G. R., CANETE, J. D., DAMJANOV, N. S., KRAGSTRUP, T. W., KVIEN, T. K., LANDEWE, R. B. M., LORIES, R. J. U., MARZO-ORTEGA, H., PODDUBNYI, D., RODRIGUES MANICA, S. A., SCHETT, G., VEALE, D. J., VAN DEN BOSCH, F. E., VAN DER HEIJDE, D. & SMOLEN, J. S. 2020. EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2019 update. *Annals of the Rheumatic Diseases*, 79, 700-712.

- GOWIN, J. L., ERNST, M., BALL, T., MAY, A. C., SLOAN, M. E., TAPERT, S. F. & PAULUS, M. P. 2019. Using neuroimaging to predict relapse in stimulant dependence: A comparison of linear and machine learning models. *Neuroimage-Clinical*, 21.
- GRAMACY, R. B. 2020. *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*, CRC press.
- GRATTON, C., LAUMANN, T. O., NIELSEN, A. N., GREENE, D. J., GORDON, E. M., GILMORE, A. W., NELSON, S. M., COALSON, R. S., SNYDER, A. Z., SCHLAGGAR, B. L., DOSENBACH, N. U. F. & PETERSEN, S. E. 2018. Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, 98, 439-+.
- GREB, J. E., GOLDMINZ, A. M., ELDER, J. T., LEBWOHL, M. G., GLADMAN, D., WU, J. J., MEHTA, N. N., FINLAY, A. Y. & GOTTLIEB, A. B. 2016. Psoriasis. *Nature Reviews Disease Primers*, 2.
- GREENE, A. S., GAO, S., SCHEINOST, D. & CONSTABLE, R. T. 2018. Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, 9.
- GREENER, J. G., KANDATHIL, S. M., MOFFAT, L. & JONES, D. T. 2022. A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, 23, 40-55.
- GROSENICK, L., SHI, T. C., GUNNING, F. M., DUBIN, M. J., DOWNAR, J. & LISTON, C. 2019. Functional and Optogenetic Approaches to Discovering Stable Subtype-Specific Circuit Mechanisms in Depression. *Biological Psychiatry-Cognitive Neuroscience and Neuroimaging*, 4, 554-566.
- GU, Z., GU, L., EILS, R., SCHLESNER, M. & BRORS, B. 2014. circize implements and enhances circular visualization in R. *Bioinformatics*, 30, 2811-2812.
- GUAN, Y., ZHANG, H., QUANG, D., WANG, Z., PARKER, S. C. J., PAPPAS, D. A., KREMER, J. M. & ZHU, F. 2019. Machine Learning to Predict Anti-Tumor Necrosis Factor Drug Responses of Rheumatoid Arthritis Patients by Integrating Clinical and Genetic Markers. *Arthritis & Rheumatology*, 71, 1987-1996.
- GUDU, T., ETCHETO, A., DE WIT, M., HEIBERG, T., MACCARONE, M., BALANESCU, A., BALINT, P. V., NIEDERMAYER, D. S., CANETE, J. D., HELLIWELL, P., KALYONCU, U., KILTZ, U., OTSAL, K., VEALE, D. J., DE VLAM, K., SCRIVO, R., STAMM, T., KVIEN, T. K. & GOSSEC, L. 2016. Fatigue in psoriatic arthritis - a cross-sectional study of 246 patients from 13 countries. *Joint Bone Spine*, 83, 439-443.
- GUISAN, A., ZIMMERMANN, N. E. & THUILLER, W. 2017. Assessing Model Performance: Which Data to Use? *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge: Cambridge University Press.
- HALLETT, M. 2007. Transcranial magnetic stimulation: A primer. *Neuron*, 55, 187-199.

- HALLQUIST, M. N., HWANG, K. & LUNA, B. 2013. The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *Neuroimage*, 82, 208-225.
- HANKEN, K., ELING, P. & HILDEBRANDT, H. 2014. The representation of inflammatory signals in the brain - a model for subjective fatigue in multiple sclerosis. *Frontiers in Neurology*, 5.
- HANKEN, K., ELING, P., KLEIN, J., KLAENE, E. & HILDEBRANDT, H. 2016. Different cortical underpinnings for fatigue and depression in MS? *Multiple Sclerosis and Related Disorders*, 6, 81-86.
- HARREWIJN, A., ABEND, R., LINKE, J., BROTMAN, M. A., FOX, N. A., LEIBENLUFT, E., WINKLER, A. M. & PINE, D. S. 2020. Combining fMRI during resting state and an attention bias task in children. *Neuroimage*, 205.
- HARRISON, N. A., BRYDON, L., WALKER, C., GRAY, M. A., STEPTOE, A., DOLAN, R. J. & CRITCHLEY, H. D. 2009. Neural Origins of Human Sickness in Interoceptive Responses to Inflammation. *Biological Psychiatry*, 66, 415-422.
- HARTE, S. E., HARRIS, R. E. & CLAUW, D. J. 2018. The neurobiology of central sensitization. *Journal of Applied Biobehavioral Research*, 23.
- HARUKI, Y. & OGAWA, K. 2021. Role of anatomical insular subdivisions in interoception: Interoceptive attention and accuracy have dissociable substrates. *European Journal of Neuroscience*, 53, 2669-2680.
- HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. H. 2009. *The elements of statistical learning: data mining, inference, and prediction*, Springer.
- HAWCO, C., VIVIANO, J. D., CHAVEZ, S., DICKIE, E. W., CALARCO, N., KOCHUNOV, P., ARGYELAN, M., TURNER, J. A., MALHOTRA, A. K., BUCHANAN, R. W., VOINESKOS, A. N. & GRP, S. 2018. A longitudinal human phantom reliability study of multi-center T1-weighted, DTI, and resting state fMRI data. *Psychiatry Research-Neuroimaging*, 282, 134-142.
- HEARNE, L. J., COCCHI, L., ZALESKY, A. & MATTINGLEY, J. B. 2017. Reconfiguration of Brain Network Architectures between Resting-State and Complexity-Dependent Cognitive Reasoning. *Journal of Neuroscience*, 37, 8399-8411.
- HELWEGEN, K., LIBEDINSKY, I. & VAN DEN HEUVEL, M. P. 2023. Statistical power in network neuroscience. *Trends in Cognitive Sciences*, 27, 282-301.
- HENNIG, C. 2019. *Cluster Validation Statistics* [Online]. Available: <https://www.rdocumentation.org/packages/fpc/versions/2.2-3/topics/cluster.stats> [Accessed April 2, 2019].

- HERNANDEZ, M., GUERRERO, G. D., CECILIA, J. M., GARCIA, J. M., INUGGI, A., JBABDI, S., BEHRENS, T. E. J. & SOTIROPOULOS, S. N. 2013. Accelerating Fibre Orientation Estimation from Diffusion Weighted Magnetic Resonance Imaging Using GPUs. *Plos One*, 8.
- HERNESNIEMI, J. A., MAHDIANI, S., TYNKKYNNEN, J. A., LYYTIKAINEN, L.-P., MISHRA, P. P., LEHTIMAKI, T., ESKOLA, M., NIKUS, K., ANTILA, K. & OKSALA, N. 2019. Extensive phenotype data and machine learning in prediction of mortality in acute coronary syndrome - the MADDEC study. *Annals of Medicine*, 51, 156-163.
- HEWLETT, S., AMBLER, N., ALMEIDA, C., BLAIR, P. S., CHOY, E., DURES, E., HAMMOND, A., HOLLINGWORTH, W., KIRWAN, J., PLUMMER, Z., ROOKE, C., THORN, J., TOMKINSON, K., POLLOCK, J. & TEAM, R. S. 2015. Protocol for a randomised controlled trial for Reducing Arthritis Fatigue by clinical Teams (RAFT) using cognitive-behavioural approaches. *Bmj Open*, 5.
- HEWLETT, S., COCKSHOTT, Z., BYRON, M., KITCHEN, K., TIPLER, S., POPE, D. & HEHIR, M. 2005. Patients' perceptions of fatigue in rheumatoid arthritis: Overwhelming, uncontrollable, ignored. *Arthritis & Rheumatism-Arthritis Care & Research*, 53, 697-702.
- HODSON, T. O. 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15, 5481-5487.
- HOGESTOL, E. A., NYGAARD, G. O., ALNAES, D., BEYER, M. K., WESTLYE, L. T. & HARBO, H. F. 2019. Symptoms of fatigue and depression is reflected in altered default mode network connectivity in multiple sclerosis. *Plos One*, 14.
- HOLDREN, M., SCHIEIR, O., BARTLETT, S. J., BESSETTE, L., BOIRE, G., HAZLEWOOD, G., HITCHON, C., KEYSTONE, E., TIN, D., THORNE, C., BYKERK, V. & POPE, J. 2019. ACHIEVING A LOW DISEASE STATE WITHIN FIRST 3 MONTHS IN EARLY RHEUMATOID ARTHRITIS RESULTS IN LOWER FATIGUE OVER 5 YEARS. *Annals of the Rheumatic Diseases*, 78, 240-240.
- HOLTEN, K., SUNDLISATER, N. P., LILLEGRAVEN, S., SEXTON, J., NORDBERG, L. B., MOHOLT, E., HAMMER, H. B., UHLIG, T., KVIEN, T. K., HAAVARDSHOLM, E. A. & AGA, A.-B. 2022. Fatigue in patients with early rheumatoid arthritis undergoing treat-to-target therapy: predictors and response to treatment. *Annals of the Rheumatic Diseases*, 81, 344-350.
- HONG, S.-J., VALK, S. L., DI MARTINO, A., MILHAM, M. P. & BERNHARDT, B. C. 2018. Multidimensional Neuroanatomical Subtyping of Autism Spectrum Disorder. *Cerebral Cortex*, 28, 3578-3588.
- HUBERT, L. & ARABIE, P. 1985. COMPARING PARTITIONS. *Journal of Classification*, 2, 193-218.

- HUETTEL, S. A., SONG, A. W. & MCCARTHY, G. 2014. Functional Magnetic Resonance Imaging, Third Edition. *Functional Magnetic Resonance Imaging, Third Edition*, 1-573.
- HUSTED, J. A., TOM, B. D., SCHENTAG, C. T., FAREWELL, V. T. & GLADMAN, D. D. 2009. Occurrence and correlates of fatigue in psoriatic arthritis. *Annals of the Rheumatic Diseases*, 68, 1553-1558.
- HUTCHISON, R. M., WOMELSDORF, T., ALLEN, E. A., BANDETTINI, P. A., CALHOUN, V. D., CORBETTA, M., DELLA PENNA, S., DUYN, J. H., GLOVER, G. H., GONZALEZ-CASTILLO, J., HANDWERKER, D. A., KEILHOLZ, S., KIVINIEMI, V., LEOPOLD, D. A., DE PASQUALE, F., SPORNS, O., WALTER, M. & CHANG, C. 2013. Dynamic functional connectivity: Promise, issues, and interpretations. *Neuroimage*, 80, 360-378.
- HVITFELDT, E., PEDERSEN, T. L. & BENESTY, M. 2022. Package 'lime'.
- ING, A., SAEMANN, P. G., CHU, C., TAY, N., BIONDO, F., ROBERTL, G., JIA, T., WOLFERS, T., DESRIVIERES, S., BANASCHEWSKI, T., BOKDE, A. L. W., BROMBERG, U., BUECHEL, C., CONROD, P., FADAI, T., FLOR, H., FROUIN, V., GARAVAN, H., SPECHLER, P. A., GOWLANDL, P., GRIMMER, Y., HEINZ, A., ITTERMANN, B., KAPPEL, V., MARTINOT, J.-L., MEYER-LINDENBERG, A., MILLENETS, S., NESS, F., VAN NOORT, B., ORFANOS, D. P., MARTINOT, M.-L. P., PENTTILAE, J., POUSTKA, L., QUINLAN, E. B., SMOLKA, M. N., STRINGARIS, A., STRUVE, M., VEER, I. M., WALTER, H., WHELAN, R., ANDREASSEN, O. A., AGARTZ, I., LEMAITRE, H., BARKER, E. D., ASHBURNER, J., BINDER, E., BUITELAAR, J., MARQUAND, A., ROBBINS, T. W., SCHUMANN, G. & CONSORTIUM, I. 2019. Identification of neurobehavioural symptom groups based on shared brain mechanisms. *Nature Human Behaviour*, 3, 1306-1318.
- JAEGER, B. C., TIERNEY, N. J. & SIMON, N. R. 2020. When to Impute? Imputation before and during cross-validation. *arXiv preprint arXiv:2010.00718*.
- JAIME-LARA, R. B., KOONS, B. C., MATURA, L. A., HODGSON, N. A. & RIEGEL, B. 2020. A Qualitative Metasynthesis of the Experience of Fatigue Across Five Chronic Conditions. *Journal of Pain and Symptom Management*, 59, 1320-1343.
- JAKOBSEN, J. C., GLUUD, C., WETTERSLEV, J. & WINKEL, P. 2017. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *Bmc Medical Research Methodology*, 17.
- JAMEEN, A. R. M., RIBBONS, K., LECHNER-SCOTT, J. & RAMADAN, S. 2019. Evaluation of MS related central fatigue using MR neuroimaging methods: Scoping review. *Journal of the Neurological Sciences*, 400, 52-71.
- JAWABRI, K. H. & SHARMA, S. 2019. Physiology, Cerebral Cortex Functions. *StatPearls [Internet]*. StatPearls Publishing.

- JENNEKENSSCHINKEL, A., SANDERS, E., LANSER, J. B. K. & VANDERVELDE, E. A. 1988. REACTION-TIME IN AMBULANT MULTIPLE-SCLEROSIS PATIENTS .1. INFLUENCE OF PROLONGED COGNITIVE EFFORT. *Journal of the Neurological Sciences*, 85, 173-186.
- JENSEN, K. B., SRINIVASAN, P., SPAETH, R., TAN, Y., KOSEK, E., PETZKE, F., CARVILLE, S., FRANSSON, P., MARCUS, H., WILLIAMS, S. C. R., CHOY, E., VITTON, O., GRACELY, R., INGVAR, M. & KONG, J. 2013. Overlapping Structural and Functional Brain Changes in Patients With Long-Term Exposure to Fibromyalgia Pain. *Arthritis and Rheumatism*, 65, 3293-3303.
- JIN, T., NGUYEN, N. D., TALOS, F. & WANG, D. 2021. ECMarker: interpretable machine learning model identifies gene expression biomarkers predicting clinical outcomes and reveals molecular mechanisms of human disease in early stages. *Bioinformatics*, 37, 1115-1124.
- JOEL, L. O., DOORSAMY, W. & PAUL, B. S. 2022. A review of missing data handling techniques for machine learning. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 5, 971-1005.
- JOHNSON, S. K., LANGE, G., DELUCA, J., KORN, L. R. & NATELSON, B. 1997. The effects of fatigue on neuropsychological performance in patients with chronic fatigue syndrome, multiple sclerosis, and depression. *Applied neuropsychology*, 4, 145-53.
- JOLLANS, L., BOYLE, R., ARTIGES, E., BANASCHEWSKI, T., DESRIVIERES, S., GRIGIS, A., MARTINOT, J.-L., PAUS, T., SMOLKA, M. N., WALTER, H., SCHUMANN, G., GARAVAN, H. & WHELAN, R. 2019. Quantifying performance of machine learning methods for neuroimaging data. *Neuroimage*, 199, 351-365.
- JUAREZ-OROZCO, L. E., MARTINEZ-MANZANERA, O., NESTEROV, S. V., KAJANDER, S. & KNUUTI, J. 2018. The machine learning horizon in cardiac hybrid imaging. *European Journal of Hybrid Imaging*, 2.
- KANG, S.-G. & CHO, S.-E. 2020. Neuroimaging Biomarkers for Predicting Treatment Response and Recurrence of Major Depressive Disorder. *International Journal of Molecular Sciences*, 21.
- KAPLAN, C., MINC, A., BASU, N. & SCHREPF, A. 2019. Inflammation and the Central Nervous System in Inflammatory Rheumatic Disease. *Current Rheumatology Reports*, 21.
- KARSHIKOFF, B., SUNDELIN, T. & LASSELIN, J. 2017. Role of inflammation in Human Fatigue: Relevance of Multidimensional Assessments and Potential Neuronal Mechanisms. *Frontiers in Immunology*, 8.
- KEROLA, A. M., KAZEMI, A., ROLLEFSTAD, S., LILLEGRAVEN, S., SEXTON, J., WIBETOE, G., HAAVARDSHOLM, E. A., KVIEN, T. K. & SEMB, A. G. 2022. All-cause and cause-specific mortality in rheumatoid arthritis, psoriatic arthritis and axial spondyloarthritis: a nationwide registry study. *Rheumatology*, 61, 4656-4666.

- KIM, S. B. 2018. Explaining Lord's Paradox in Introductory Statistical Theory Courses. *International Journal of Statistics and Probability*, 7.
- KJELDGAARD, L. 2018. modelgrid: A framework for creating, managing and training multiple caret models (2018). *R package version*, 1.
- KLISTORNER, A., WANG, C., YIANNIKAS, C., GRAHAM, S. L., PARRATT, J. & BARNETT, M. H. 2016. Progressive Injury in Chronic Multiple Sclerosis Lesions Is Gender-Specific: A DTI Study. *Plos One*, 11.
- KOOB, G. F. & VOLKOW, N. D. 2016. Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry*, 3, 760-773.
- KOPPE, G., MEYER-LINDENBERG, A. & DURSTEWITZ, D. 2021. Deep learning for small and big data in psychiatry. *Neuropsychopharmacology*, 46, 176-190.
- KORTE, S. M. & STRAUB, R. H. 2019. Fatigue in inflammatory rheumatic disorders: pathophysiological mechanisms. *Rheumatology*, 58, 35-50.
- KORZHOVA, J., BAKULIN, I., SINITSYN, D., POYDASHEVA, A., SUPONEVA, N., ZAKHAROVA, M. & PIRADOV, M. 2019. High-frequency repetitive transcranial magnetic stimulation and intermittent theta-burst stimulation for spasticity management in secondary progressive multiple sclerosis. *European Journal of Neurology*, 26, 680-+.
- KOUBIYR, I., BESSON, P., DELOIRE, M., CHARRE-MORIN, J., SAUBUSSE, A., TOURDIAS, T., BROCHET, B. & RUET, A. 2019. Dynamic modular-level alterations of structural-functional coupling in clinically isolated syndrome. *Brain*, 142, 3428-3439.
- KRATZ, A. L., SCHILLING, S., GOESLING, J. & WILLIAMS, D. A. 2016. The PROMIS Fatigue(FM) Profile: a self-report measure of fatigue for use in fibromyalgia. *Quality of Life Research*, 25, 1803-1813.
- KRAYNAK, T. E., MARSLAND, A. L., WAGER, T. D. & GIANAROS, P. J. 2018. Functional neuroanatomy of peripheral inflammatory physiology: A meta-analysis of human neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 94, 76-92.
- KRIENEN, F. M., YEO, B. T. T. & BUCKNER, R. L. 2014. Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 369.
- KUHN, M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28, 1-26.
- KUHN, M. & JOHNSON, K. 2013. *Applied predictive modeling*, Springer.

- KUPPUSWAMY, A. 2022. The Neurobiology of Pathological Fatigue: New Models, New Questions. *Neuroscientist*, 28, 238-253.
- LABITIGAN, M., BAHCE-ALTUNTAS, A., KREMER, J. M., REED, G., GREENBERG, J. D., JORDAN, N., PUTTERMAN, C. & BRODER, A. 2014. Higher Rates and Clustering of Abnormal Lipids, Obesity, and Diabetes Mellitus in Psoriatic Arthritis Compared With Rheumatoid Arthritis. *Arthritis Care & Research*, 66, 600-607.
- LAKE, E. M. R., FINN, E. S., NOBLE, S. M., VANDERWAL, T., SHEN, X., ROSENBERG, M. D., SPANN, M. N., CHUN, M. M., SCHEINOST, D. & CONSTABLE, R. T. 2019. The Functional Brain Organization of an Individual Allows Prediction of Measures of Social Abilities Transdiagnostically in Autism and Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, 86, 315-326.
- LASSELIN, J. 2021. Back to the future of psychoneuroimmunology: Studying inflammation-induced sickness behavior. *Brain, Behavior, & Immunity - Health*, 18.
- LE, T. T., URBANOWICZ, R. J., MOORE, J. H. & MCKINNEY, B. A. 2019. STatistical Inference Relief (STIR) feature selection. *Bioinformatics*, 35, 1358-1365.
- LEACH, L. F. & HENSON, R. K. 2014. Bias and precision of the squared canonical correlation coefficient under nonnormal data condition. *Journal of Modern Applied Statistical Methods*, 13, 8.
- LEE, J., CHI, S. & LEE, M.-S. 2022. Personalized Diagnosis and Treatment for Neuroimaging in Depressive Disorders. *Journal of Personalized Medicine*, 12.
- LEE, Y. C., FRITS, M. L., IANNACCONE, C. K., WEINBLATT, M. E., SHADICK, N. A., WILLIAMS, D. A. & CUI, J. 2014. Subgrouping of Patients With Rheumatoid Arthritis Based on Pain, Fatigue, Inflammation, and Psychosocial Factors. *Arthritis & Rheumatology*, 66, 2006-2014.
- LEONARDI, N. & VAN DE VILLE, D. 2015. On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage*, 104, 430-436.
- LEVY, A., TAIB, S., ARBUS, C., PERAN, P., SAUVAGET, A., SCHMITT, L. & YRONDI, A. 2019. Neuroimaging Biomarkers at Baseline Predict Electroconvulsive Therapy Overall Clinical Response in Depression A Systematic Review. *Journal of Ect*, 35, 77-83.
- LI, B.-J., FRISTON, K., MODY, M., WANG, H.-N., LU, H.-B. & HU, D.-W. 2018. A brain network model for depression: From symptom understanding to disease intervention. *Cns Neuroscience & Therapeutics*, 24, 1004-1019.
- LI, Y., ZHANG, Q. & YOON, S. W. 2021. Gaussian process regression-based learning rate optimization in convolutional neural networks for medical images classification. *Expert Systems with Applications*, 184.

- LIBEDINSKY, I., HELWEGEN, K., DANNLOWSKI, U., FORNITO, A., REPPLE, J., ZALESKY, A., BREAKSPEAR, M., VAN DEN HEUVEL, M., ALZHEIMER'S DISEASE NEUROIMAGING, I. & ALZHEIMER'S DISEASE REPOSITORY WITHOUT BORDERS, I. 2022. Reproducibility of neuroimaging studies of brain disorders with hundreds-not thousands-of participants. *bioRxiv*.
- LINNHOF, S., FIENE, M., HEINZE, H.-J. & ZAEHLE, T. 2019. Cognitive Fatigue in Multiple Sclerosis: An Objective Approach to Diagnosis and Treatment by Transcranial Electrical Stimulation. *Brain Sciences*, 9.
- LIU, S., CAI, W., LIU, S., ZHANG, F., FULHAM, M., FENG, D., PUJOL, S. & KIKINIS, R. 2015. Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders. *Brain informatics*, 2, 167-180.
- LIU, Y., ARYEE, M. J., PADYUKOV, L., FALLIN, M. D., HESSELBERG, E., RUNARSSON, A., REINIUS, L., ACEVEDO, N., TAUB, M., RONNINGER, M., SHCHETYNSKY, K., SCHEYNIUS, A., KERE, J., ALFREDSSON, L., KLARESKOG, L., EKSTROM, T. J. & FEINBERG, A. P. 2013. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology*, 31, 142-147.
- LIU, Y., HAYES, D. N., NOBEL, A. & MARRON, J. S. 2008. Statistical Significance of Clustering for High-Dimension, Low-Sample Size Data. *Journal of the American Statistical Association*, 103, 1281-1293.
- LLUFRIU, S., MARTINEZ-HERAS, E., SOLANA, E., SOLA-VALLS, N., SEPULVEDA, M., BLANCO, Y., MARTINEZ-LAPISCINA, E. H., ANDORRA, M., VILLOSLADA, P., PRATS-GALINO, A. & SAIZ, A. 2017. Structural networks involved in attention and executive functions in multiple sclerosis. *Neuroimage-Clinical*, 13, 288-296.
- LO VERCIO, L., AMADOR, K., BANNISTER, J. J., CRITES, S., GUTIERREZ, A., MACDONALD, M. E., MOORE, J., MOUCHES, P., RAJASHEKAR, D., SCHIMERT, S., SUBBANNA, N., TULADHAR, A., WANG, N., WILMS, M., WINDER, A. & FORKERT, N. D. 2020. Supervised machine learning tools: a tutorial for clinicians. *Journal of Neural Engineering*, 17.
- LOPEZ-SOLA, M., PUJOL, J., WAGER, T. D., GARCIA-FONTANALS, A., BLANCO-HINOJO, L., GARCIA-BLANCO, S., POCA-DIAS, V., HARRISON, B. J., CONTRERAS-RODRIGUEZ, O., MONFORT, J., GARCIA-FRUCTUOSO, F. & DEUS, J. 2014. Altered Functional Magnetic Resonance Imaging Responses to Nonpainful Sensory Stimulation in Fibromyalgia Patients. *Arthritis & Rheumatology*, 66, 3200-3209.
- LOW, A. Y. T., GOLDSTEIN, N., GAUNT, J. R., HUANG, K.-P., ZAINOLABIDIN, N., YIP, A. K. K., CARTY, J. R. E., CHOI, J. Y., MILLER, A. M., HO, H. S. T., LENHERR, C., BALTAR, N., AZIM, E., SESSIONS, O. M., CH'NG, T. H., BRUCE, A. S., MARTIN, L. E., HALKO, M. A., BRADY, R. O., JR., HOLSEN, L. M., ALHADEFF, A. L., CHEN, A. I. & BETLEY, J. N. 2021. Reverse-translational identification of a cerebellar satiation network. *Nature*, 600, 269-+.

- MACCALLUM, R. C., ZHANG, S. B., PREACHER, K. J. & RUCKER, D. D. 2002. On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19-40.
- MACEDO, A. M., OAKLEY, S. P., PANAYI, G. S. & KIRKHAM, B. W. 2009. Functional and Work Outcomes Improve in Patients With Rheumatoid Arthritis Who Receive Targeted, Comprehensive Occupational Therapy. *Arthritis & Rheumatism-Arthritis Care & Research*, 61, 1522-1530.
- MACGREGOR, A. J., SNIEDER, H., RIGBY, A. S., KOSKENVUO, M., KAPRIO, J., AHO, K. & SILMAN, A. J. 2000. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis and Rheumatism*, 43, 30-37.
- MACIUKIEWICZ, M., MARSHE, V. S., HAUSCHILD, A.-C., FOSTER, J. A., ROTZINGER, S., KENNEDY, J. L., KENNEDY, S. H., MUELLER, D. J. & GERACI, J. 2018. GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. *Journal of Psychiatric Research*, 99, 62-68.
- MADRID-GARCIA, A., MERINO-BARBANCHO, B., RODRIGUEZ-GONZALEZ, A., FERNANDEZ-GUTIERREZ, B., RODRIGUEZ-RODRIGUEZ, L. & MENASALVAS-RUIZ, E. 2023. Understanding the role and adoption of artificial intelligence techniques in rheumatology research: An in-depth review of the literature. *Seminars in Arthritis and Rheumatism*, 61.
- MAGALHAES, R., MARQUES, P., SOARES, J., ALVES, V. & SOUSA, N. 2015. The Impact of Normalization and Segmentation on Resting-State Brain Networks. *Brain Connectivity*, 5, 166-176.
- MALLEY, J. D., MALLEY, K. G. & PAJEVIC, S. 2011. *Statistical learning for biomedical data*, Cambridge University Press.
- MANJALY, Z.-M., HARRISON, N. A., CRITCHLEY, H. D., DO, C. T., STEFANICS, G., WENDEROTH, N., LUTTEROTTI, A., MUELLER, A. & STEPHAN, K. E. 2019. Pathophysiological and cognitive mechanisms of fatigue in multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 90, 642-651.
- MARDIA, K. V., KENT, J. T. & BIBBY, J. M. 1980. *Multivariate Analysis (Probability and Mathematical Statistics)* Academic Press. London.
- MAREK, S., TERVO-CLEMMENS, B., CALABRO, F. J., MONTEZ, D. F., KAY, B. P., HATOUM, A. S., DONOHUE, M. R., FORAN, W., MILLER, R. L., HENDRICKSON, T. J., MALONE, S. M., KANDALA, S., FECZKO, E., MIRANDA-DOMINGUEZ, O., GRAHAM, A. M., EARL, E. A., PERRONE, A. J., CORDOVA, M., DOYLE, O., MOORE, L. A., CONAN, G. M., URIARTE, J., SNIDER, K., LYNCH, B. J., WILGENBUSCH, J. C., PENGO, T., TAM, A., CHEN, J., NEWBOLD, D. J., ZHENG, A., SEIDER, N. A., VAN, A. N., METOKI, A., CHAUVIN, R. J., LAUMANN, T. O., GREENE, D. J., PETERSEN, S. E., GARAVAN, H., THOMPSON, W. K., NICHOLS, T. E., YEO, B. T. T., BARCH,

- D. M., LUNA, B., FAIR, D. A. & DOSENBACH, N. U. F. 2022. Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603, 654-+.
- MARQUAND, A., HOWARD, M., BRAMMER, M., CHU, C., COEN, S. & MOURAO-MIRANDA, J. 2010. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *Neuroimage*, 49, 2178-2189.
- MARRELLI, K., CHENG, A. J., BROPHY, J. D. & POWER, G. A. 2018. Perceived Versus Performance Fatigability in Patients With Rheumatoid Arthritis. *Frontiers in Physiology*, 9.
- MARTIN, J. H. 2003. *Neuroanatomy: text and atlas*, London;New York, N.Y.;, McGraw-Hill.
- MARTIN, K. R., BACHMAIR, E.-M., AUCOTT, L., DURES, E., EMSLEY, R., GRAY, S. R., HEWLETT, S., KUMAR, V., LOVELL, K., MACFARLANE, G. J., MACLENNAN, G., MCNAMEE, P., NORRIE, J., PAUL, L., RALSTON, S., SIEBERT, S., WEARDEN, A., WHITE, P. D. & BASU, N. 2019. Protocol for a multicentre randomised controlled parallel-group trial to compare the effectiveness of remotely delivered cognitive-behavioural and graded exercise interventions with usual care alone to lessen the impact of fatigue in inflammatory rheumatic diseases (LIFT). *Bmj Open*, 9.
- MATCHAM, F., ALI, S., HOTOPIF, M. & CHALDER, T. 2015. Psychological correlates of fatigue in rheumatoid arthritis: A systematic review. *Clinical Psychology Review*, 39, 16-29.
- MATCHAM, F., DAVIES, R., HOTOPIF, M., HYRICH, K. L., NORTON, S., STEER, S. & GALLOWAY, J. 2018. The relationship between depression and biologic treatment response in rheumatoid arthritis: An analysis of the British Society for Rheumatology Biologics Register. *Rheumatology*, 57, 835-843.
- MATSUMOTO, K., TAKATA, K., YAMADA, D., USUDA, H., WADA, K., TADA, M., MISHIMA, Y., ISHIHARA, S., HORIE, S., SAITOH, A. & KATO, S. 2021. Juvenile social defeat stress exposure favors in later onset of irritable bowel syndrome-like symptoms in male mice. *Scientific Reports*, 11.
- MATURA, L. A., MALONE, S., JAIME-LARA, R. & RIEGEL, B. 2018. A Systematic Review of Biological Mechanisms of Fatigue in Chronic Illness. *Biological Research for Nursing*, 20, 410-421.
- MAZZIOTTA, J. C., TOGA, A. W., EVANS, A., FOX, P. & LANCASTER, J. 1995. A PROBABILISTIC ATLAS OF THE HUMAN BRAIN - THEORY AND RATIONALE FOR ITS DEVELOPMENT. *Neuroimage*, 2, 89-101.
- MC ARDLE, A., FLATLEY, B., PENNINGTON, S. R. & FITZGERALD, O. 2015. Early biomarkers of joint damage in rheumatoid and psoriatic arthritis. *Arthritis Research & Therapy*, 17.

- MCGRATH, C. L., KELLEY, M. E., HOLTZHEIMER, P. E., III, DUNLOP, B. W., CRAIGHEAD, W. E., FRANCO, A. R., CRADDOCK, C. & MAYBERG, H. S. 2013. Toward a Neuroimaging Treatment Selection Biomarker for Major Depressive Disorder. *Jama Psychiatry*, 70, 821-829.
- MEEUS, M. & NIJS, J. 2007. Central sensitization: a biopsychosocial explanation for chronic widespread pain in patients with fibromyalgia and chronic fatigue syndrome. *Clinical Rheumatology*, 26, 465-473.
- MEHLING, W. E., ACREE, M., STEWART, A., SILAS, J. & JONES, A. 2018. The Multidimensional Assessment of Interoceptive Awareness, Version 2 (MAIA-2). *Plos One*, 13.
- MEINSHAUSEN, N. & BUEHLMANN, P. 2010. Stability selection. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 72, 417-473.
- MENNES, M., KELLY, C., ZUO, X.-N., DI MARTINO, A., BISWAL, B. B., CASTELLANOS, F. X. & MILHAM, M. P. 2010. Inter-individual differences in resting-state functional connectivity predict task-induced BOLD activity. *Neuroimage*, 50, 1690-1701.
- MENON, V. & UDDIN, L. Q. 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Structure & Function*, 214, 655-667.
- MEROLA, J. F., ESPINOZA, L. R. & FLEISCHMANN, R. 2018. Distinguishing rheumatoid arthritis from psoriatic arthritis. *Rmd Open*, 4.
- MERTENS, B. J. A., BANZATO, E. & DE WREEDE, L. C. 2020. Construction and assessment of prediction rules for binary outcome in the presence of missing predictor data using multiple imputation and cross-validation: Methodological approach and data-based evaluation. *Biometrical Journal*, 62, 724-741.
- MICHELSEN, B., KRISTIANSUND, E. K., SEXTON, J., HAMMER, H. B., FAGERLI, K. M., LIE, E., WIEROD, A., KALSTAD, S., RODEVAND, E., KROLL, F., HAUGEBERG, G. & KVIEN, T. K. 2017. Do depression and anxiety reduce the likelihood of remission in rheumatoid arthritis and psoriatic arthritis? Data from the prospective multicentre NORDMARD study. *Annals of the Rheumatic Diseases*, 76, 1906-1910.
- MIHALIK, A., ADAMS, R. A. & HUYS, Q. 2020. Canonical Correlation Analysis for Identifying Biotypes of Depression. *Biological psychiatry. Cognitive neuroscience and neuroimaging*.
- MILANESCHI, Y., HOOGENDIJK, W., LIPS, P., HEIJBOER, A. C., SCHOEVEERS, R., VAN HEMERT, A. M., BEEKMAN, A. T. F., SMIT, J. H. & PENNINX, B. W. J. H. 2014. The association between low vitamin D and depressive disorders. *Molecular Psychiatry*, 19, 444-451.

- MOLENT, C., OLIVO, D., WOLF, R. C., BALESTRIERI, M. & SAMBATARO, F. 2019. Functional neuroimaging in treatment resistant schizophrenia: A systematic review. *Neuroscience and Biobehavioral Reviews*, 104, 178-190.
- MOLNAR, C. 2020. *Interpretable machine learning*, Lulu. com.
- MOLNAR, C., CASALICCHIO, G. & BISCHL, B. 2018. iml: An R package for interpretable machine learning. *Journal of Open Source Software*, 3, 786.
- MORFINI, F., WHITFIELD-GABRIELI, S. & NIETO-CASTANON, A. 2023. Functional connectivity MRI quality control procedures in CONN. *Frontiers in Neuroscience*, 17.
- MORRIS, G., BERK, M., WALDER, K. & MAES, M. 2015. Central pathways causing fatigue in neuro-inflammatory and autoimmune illnesses. *Bmc Medicine*, 13.
- MULLER, T. & APPS, M. A. J. 2019. Motivational fatigue: A neurocognitive framework for the impact of effortful exertion on subsequent motivation. *Neuropsychologia*, 123, 141-151.
- NAMKUNG, H., KIM, S.-H. & SAWA, A. 2017. The Insula: An Underestimated Brain Area in Clinical Neuroscience, Psychiatry, and Neurology. *Trends in Neurosciences*, 40, 200-207.
- NGUYEN, K. P., FATT, C. C., TREACHER, A., MELLEMA, C., COOPER, C., JHA, M. K., KURIAN, B., FAVA, M., MCGRATH, P. J., WEISSMAN, M., PHILLIPS, M. L., TRIVEDI, M. H. & MONTILLO, A. A. 2022. Patterns of Pretreatment Reward Task Brain Activation Predict Individual Antidepressant Response: Key Results From the EMBARC Randomized Clinical Trial. *Biological Psychiatry*, 91, 550-560.
- NHU, N. T., CHEN, D. Y.-T. & KANG, J.-H. 2023. Functional Connectivity and Structural Signatures of the Visual Cortical System in Fibromyalgia: A Magnetic Resonance Imaging Study. *The Journal of rheumatology*, 50, 1063-1070.
- NICKLIN, J., CRAMP, F., KIRWAN, J., GREENWOOD, R., URBAN, M. & HEWLETT, S. 2010. Measuring Fatigue in Rheumatoid Arthritis: A Cross-Sectional Study to Evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire, Visual Analog Scales, and Numerical Rating Scales. *Arthritis Care & Research*, 62, 1559-1568.
- NIELSEN, A. N., BARCH, D. M., PETERSEN, S. E., SCHLAGGAR, B. L. & GREENE, D. J. 2020. Machine Learning With Neuroimaging: Evaluating Its Applications in Psychiatry. *Biological Psychiatry-Cognitive Neuroscience and Neuroimaging*, 5, 791-798.
- NIETO-CASTANON, A. 2020. *Handbook of functional connectivity magnetic resonance imaging methods in CONN*, Hilbert Press.

- NIETO-CASTANON, A. 2022. Brain-wide connectome inferences using functional connectivity MultiVariate Pattern Analyses (fc-MVPA). *Plos Computational Biology*, 18.
- NIETO-CASTANON, A. & WHITFIELD-GABRIELI, S. 2021. CONN functional connectivity toolbox (RRID: SCR_009550), Version 21. Hilbert Press.
- NIKOLAUS, S., BODE, C., TAAL, E. & VAN DE LAAR, M. A. F. J. 2013. Fatigue and factors related to fatigue in rheumatoid arthritis: a systematic review. *Arthritis care & research*, 65, 1128-1146.
- NIU, X., ZHANG, F., KOUNIOS, J. & LIANG, H. 2020. Improved prediction of brain age using multimodal neuroimaging data. *Human Brain Mapping*, 41, 1626-1643.
- NUDEL, J., BISHARA, A. M., DE GEUS, S. W. L., PATIL, P., SRINIVASAN, J., HESS, D. T. & WOODSON, J. 2021. Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSAQIP database. *Surgical Endoscopy and Other Interventional Techniques*, 35, 182-191.
- NUGRAHA, B., KARST, M., ENGELI, S. & GUTENBRUNNER, C. 2012. Brain-derived neurotrophic factor and exercise in fibromyalgia syndrome patients: a mini review. *Rheumatology International*, 32, 2593-2599.
- OGAWA, S., LEE, T. M., KAY, A. R. & TANK, D. W. 1990. BRAIN MAGNETIC-RESONANCE-IMAGING WITH CONTRAST DEPENDENT ON BLOOD OXYGENATION. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 9868-9872.
- OLIVER, L. D., HAWCO, C., VIVIANO, J. D. & VOINESKOS, A. N. 2022. From the Group to the Individual in Schizophrenia Spectrum Disorders: Biomarkers of Social Cognitive Impairments and Therapeutic Translation. *Biological Psychiatry*, 91, 699-708.
- ORBAI, A.-M., DE WIT, M., MEASE, P., SHEA, J. A., GOSSEC, L., LEUNG, Y. Y., TILLET, W., ELMAMOUN, M., DUFFIN, K. C., CAMPBELL, W., CHRISTENSEN, R., COATES, L., DURES, E., EDER, L., FITZGERALD, O., GLADMAN, D., GOEL, N., GRIEB, S. D., HEWLETT, S., HOEJGAARD, P., KALYONCU, U., LINDSAY, C., MCHUGH, N., SHEA, B., STEINKOENIG, I., STRAND, V. & OGDIE, A. 2017. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. *Annals of the Rheumatic Diseases*, 76.
- OVERMAN, C. L., KOOL, M. B., DA SILVA, J. A. P. & GEENEN, R. 2016. The prevalence of severe fatigue in rheumatic diseases: an international study. *Clinical Rheumatology*, 35, 409-415.
- PAIK, E. S., LEE, J.-W., PARK, J.-Y., KIM, J.-H., KIM, M., KIM, T.-J., CHOI, C. H., KIM, B.-G., BAE, D.-S. & SEO, S. W. 2019. Prediction of survival outcomes in patients

with epithelial ovarian cancer using machine learning methods. *Journal of Gynecologic Oncology*, 30.

- PAOLINI, M., HARRINGTON, Y., COLOMBO, F., BETTONAGLI, V., POLETTI, S., CARMINATI, M., COLOMBO, C., BENEDETTI, F. & ZANARDI, R. 2023. Hippocampal and parahippocampal volume and function predict antidepressant response in patients with major depression: A multimodal neuroimaging study. *Journal of Psychopharmacology*, 37, 1070-1081.
- PARKES, L., TIEGO, J., AQUINO, K., BRAGANZA, L., CHAMBERLAIN, S. R., FONTENELLE, L. F., HARRISON, B. J., LORENZETTI, V., PATON, B., RAZI, A., FORNITO, A. & YUCCEL, M. 2019. Transdiagnostic variations in impulsivity and compulsivity in obsessive-compulsive disorder and gambling disorder correlate with effective connectivity in cortical-striatal-thalamic-cortical circuits. *Neuroimage*, 202.
- PAT, N., WANG, Y., BARTONICEK, A., CANDIA, J. & STRINGARIS, A. 2023. Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cerebral cortex (New York, N.Y. : 1991)*, 33, 2682-2703.
- PENG, L., LUO, Z., ZENG, L.-L., HOU, C., SHEN, H., ZHOU, Z. & HU, D. 2022. Parcellating the human brain using resting-state dynamic functional connectivity. *Cerebral Cortex*.
- PEPYS, M. B. 1981. C-REACTIVE PROTEIN 50 YEARS ON. *Lancet*, 1, 653-657.
- PEREIRA, F., MITCHELL, T. & BOTVINICK, M. 2009. Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, 45, S199-S209.
- PERLMAN, K., BENRIMOH, D., ISRAEL, S., ROLLINS, C., BROWN, E., TUNTENG, J.-F., YOU, R., YOU, E., TANGUAY-SELA, M., SNOOK, E., MIRESCO, M. & BERLIM, M. T. 2019. A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *Journal of Affective Disorders*, 243, 503-515.
- PITTERI, M., GALAZZO, I. B., BRUSINI, L., CRUCIANI, F., DAPOR, C., MARASTONI, D., MENEGAZ, G. & CALABRESE, M. 2021. Microstructural MRI Correlates of Cognitive Impairment in Multiple Sclerosis: The Role of Deep Gray Matter. *Diagnostics*, 11.
- PITZALIS, C., CHOY, E. H. S. & BUCH, M. H. 2020. Transforming clinical trials in rheumatology: towards patient-centric precision medicine. *Nature Reviews Rheumatology*, 16, 590-599.
- PLANCHUELO-GOMEZ, A., LUBEIRO, A., NUNEZ-NOVO, P., GOMEZ-PILAR, J., DE LUIS-GARCIA, R., DEL VALLE, P., MARTIN-SANTIAGO, O., PEREZ-ESCUADERO, A. & MOLINA, V. 2020. Identificación of MRI-based psychosis subtypes: Replication and refinement. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 100.

- POLDRACK, R. A. 2000. Imaging brain plasticity: Conceptual and methodological issues - A theoretical review. *Neuroimage*, 12, 1-13.
- POLLARD, L. C., CHOY, E. H., GONZALEZ, J., KHOSHABA, B. & SCOTT, D. L. 2006. Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology*, 45, 885-889.
- POOLEY, R. A. 2005. AAPM/RSNA physics tutorial for residents - Fundamental physics of MR imaging. *Radiographics*, 25, 1087-1099.
- POPE, J. E. 2020. Management of Fatigue in Rheumatoid Arthritis. *Rmd Open*, 6.
- PORTUGAL, L. C. L., SCHROUFF, J., STIFFLER, R., BERTOCCI, M., BEBKO, G., CHASE, H., LOCKOVITCH, J., ASLAM, H., GRAUR, S., GREENBERG, T., PEREIRA, M., OLIVEIRA, L., PHILLIPS, M. & MOURAO-MIRANDA, J. 2019. Predicting anxiety from wholebrain activity patterns to emotional faces in young adults: a machine learning approach. *Neuroimage-Clinical*, 23.
- POWER, J. D., MITRA, A., LAUMANN, T. O., SNYDER, A. Z., SCHLAGGAR, B. L. & PETERSEN, S. E. 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, 84, 320-341.
- PRETI, M. G., BOLTON, T. A. W. & VAN DE VILLE, D. 2017. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160, 41-54.
- PRIMDAHL, J., HEGELUND, A., LORENZEN, A. G., LOEPPEHIN, K., DURES, E. & ESBENSEN, B. A. 2019. The Experience of people with rheumatoid arthritis living with fatigue: a qualitative metasynthesis. *Bmj Open*, 9.
- PUDJIHARTONO, N., FADASON, T., KEMPA-LIEHR, A. W. & O'SULLIVAN, J. M. 2022. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in bioinformatics*, 2, 927312-927312.
- PUJOL, J., MACIA, D., GARCIA-FONTANALS, A., BLANCO-HINOJO, L., LOPEZ-SOLA, M., GARCIA-BLANCO, S., POCA-DIAS, V., HARRISON, B. J., CONTRERAS-RODRIGUEZ, O., MONFORT, J., GARCIA-FRUCTUOSO, F. & DEUS, J. 2014. The contribution of sensory system functional connectivity reduction to clinical pain in fibromyalgia. *Pain*, 155, 1492-1503.
- PUNZI, L., PODSWIADEK, M., OLIVIERO, F., LONIGRO, A., MODESTI, V., RAMONDA, R. & TODESCO, S. 2007. Laboratory findings in psoriatic arthritis. *Reumatismo*, 59 Suppl 1, 52-5.
- QAZI, J. 2021. *Regularization: Ridge, LASSO, Elastic-Net — How do they work* [Online]. Available: <https://python.plainenglish.io/a26-regularization-ridge-lasso-elastic-net-how-do-they-work-72f57a7eb1be> [Accessed 05/09/2023].
- QUITKIN, F. M., STEWART, J. W., MCGRATH, P. J., TRICAMO, E., RABKIN, J. G., OCEPEKWELIKSON, K., NUNES, E., HARRISON, W. & KLEIN, D. F. 1993.

COLUMBIA ATYPICAL DEPRESSION - A SUBGROUP OF DEPRESSIVES WITH BETTER RESPONSE TO MAOI THAN TO TRICYCLIC ANTIDEPRESSANTS OR PLACEBO. *British Journal of Psychiatry*, 163, 30-34.

- RAJESH, M., ANISHKA, S., VIKSIT, P. S., AROHI, S. & REHANA, S. 2023. Improving Short-range Reservoir Inflow Forecasts with Machine Learning Model Combination. *Water Resources Management*, 37, 75-90.
- REVELL, A. Y., SILVA, A. B., ARNOLD, T. C., STEIN, J. M., DAS, S. R., SHINOHARA, R. T., BASSETT, D. S., LITT, B. & DAVIS, K. A. 2022. A framework For brain atlases: Lessons from seizure dynamics. *Neuroimage*, 254.
- REYGAERTS, T., MITROVIC, S., FAUTREL, B. & GOSSEC, L. 2018. Effect of biologics on fatigue in psoriatic arthritis: A systematic literature review with meta-analysis. *Joint Bone Spine*, 85, 405-410.
- RIBEIRO, M. T., SINGH, S., GUESTRIN, C. & ASSOC COMP, M. 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 1135-1144.
- RIEMSMA, R. P., RASKER, J. J., TAAL, E., GRIEP, E. N., WOUTERS, J. & WIEGMAN, O. 1998. Fatigue in rheumatoid arthritis: The role of self-efficacy and problematic social support. *British Journal of Rheumatology*, 37, 1042-1046.
- ROBNIK-SIKONJA, M., SAVICKY, P. & ROBNIK-SIKONJA, M. M. 2013. Package 'CORElearn'.
- RODRIGUEZ-PEREZ, R. & BAJORATH, J. 2022. Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. *Journal of Computer-Aided Molecular Design*, 36, 355-362.
- ROSEN, J., LANDRISCINA, A. & FRIEDMAN, A. J. 2016. Psoriasis-Associated Fatigue: Pathogenesis, Metrics, and Treatment. *Cutis*, 97, 125-132.
- ROSENBERG, M. D., FINN, E. S., SCHEINOST, D., PAPADEMETRIS, X., SHEN, X., CONSTABLE, R. T. & CHUN, M. M. 2016. A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, 19, 165-+.
- ROUAULT, M., PEREIRA, I., GALIOULLINE, H., FLEMING, S. M., STEPHAN, K. E. & MANJALY, Z.-M. 2023. Interoceptive and metacognitive facets of fatigue in multiple sclerosis. *European Journal of Neuroscience*, 58, 2603-2622.
- RUAN, Q., D'ONOFRIO, G., SANCARLO, D., BAO, Z., GRECO, A. & YU, Z. 2016. Potential neuroimaging biomarkers of pathologic brain changes in Mild Cognitive Impairment and Alzheimer's disease: a systematic review. *Bmc Geriatrics*, 16.
- RUDEBECK, P. H., RICH, E. L. & MAYBERG, H. S. 2019. From bed to bench side: Reverse translation to optimize neuromodulation for mood disorders. *Proceedings of*

the National Academy of Sciences of the United States of America, 116, 26288-26296.

- RUSSELL, A., HEPGUL, N., NIKKHESLAT, N., BORSINI, A., ZAJKOWSKA, Z., MOLL, N., FORTON, D., AGARWAL, K., CHALDER, T., MONDELLI, V., HOTOPF, M., CLEARE, A., MURPHY, G., FOSTER, G., WONG, T., SCHUETZE, G. A., SCHWARZ, M. J., HARRISON, N., ZUNSZAIN, P. A. & PARIANTE, C. M. 2019. Persistent fatigue induced by interferon-alpha: a novel, inflammation-based, proxy model of chronic fatigue syndrome. *Psychoneuroendocrinology*, 100, 276-285.
- SADAGHIANI, S. & WIRSICH, J. 2020. Intrinsic connectome organization across temporal scales: New insights from cross-modal approaches. *Network Neuroscience*, 4, 1-29.
- SALAFFI, F., CAROTTI, M., GASPARINI, S., INTORCIA, M. & GRASSI, W. 2009. The health-related quality of life in rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis: a comparison with a selected sample of healthy people. *Health and Quality of Life Outcomes*, 7.
- SALIGAN, L. N., LUCKENBAUGH, D. A., SLONENA, E. E., MACHADO-VIEIRA, R. & ZARATE, C. A., JR. 2016. An assessment of the anti-fatigue effects of ketamine from a double-blind, placebo-controlled, crossover study in bipolar disorder. *Journal of Affective Disorders*, 194, 115-119.
- SANCHEZ-PINTO, L. N., LUO, Y. & CHURPEK, M. M. 2018. Big Data and Data Science in Critical Care. *Chest*, 154, 1239-1248.
- SCALA, J. J., GANZ, A. B. & SNYDER, M. P. 2023. Precision Medicine Approaches to Mental Health Care. *Physiology*, 38, 82-98.
- SCHAFFER, J. & STRIMMER, K. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4.
- SCHEINOST, D., NOBLE, S., HORIEN, C., GREENE, A. S., LAKE, E. M. R., SALEHI, M., GAO, S., SHEN, X., O'CONNOR, D., BARRON, D. S., YIP, S. W., ROSENBERG, M. D. & CONSTABLE, R. T. 2019. Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage*, 193, 35-45.
- SCHEINOST, D., SINHA, R., CROSS, S. N., KWON, S. H., SZE, G., CONSTABLE, R. T. & MENT, L. R. 2017. Does prenatal stress alter the developing connectome? *Pediatric Research*, 81, 214-226.
- SCHETT, G., LORIES, R. J., D'AGOSTINO, M.-A., ELEWAUT, D., KIRKHAM, B., SORIANO, E. R. & MCGONAGLE, D. 2017. Enthesitis: from pathophysiology to treatment. *Nature Reviews Rheumatology*, 13, 731-741.
- SCHETT, G., RAHMAN, P., RITCHLIN, C., MCINNES, I. B., ELEWAUT, D. & SCHER, J. U. 2022. Psoriatic arthritis from a mechanistic perspective. *Nature Reviews Rheumatology*, 18, 311-325.

- SCHMETTOW, M. 2021. Generalized linear models. *New Statistics for Design Researchers: A Bayesian Workflow in Tidy R*, 323-399.
- SCHNACK, H. G. & KAHN, R. S. 2016. Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Frontiers in Psychiatry*, 7.
- SCHREIBER, H., LANG, M., KILTZ, K. & LANG, C. 2015. Is personality profile a relevant determinant of fatigue in multiple sclerosis? *Frontiers in Neurology*, 6.
- SCHREPF, A., KAPLAN, C. M., ICHESCO, E., LARKIN, T., HARTE, S. E., HARRIS, R. E., MURRAY, A. D., WAITER, G. D., CLAUW, D. J. & BASU, N. 2018. A multi-modal MRI study of the central response to inflammation in rheumatoid arthritis. *Nature Communications*, 9.
- SCHULTZ, D. H. & COLE, M. W. 2016. Higher Intelligence Is Associated with Less Task-Related Brain Network Reconfiguration. *Journal of Neuroscience*, 36, 8551-8561.
- SCHULZ, M.-A., YEO, B. T. T., VOGELSTEIN, J. T., MOURAO-MIRANADA, J., KATHER, J. N., KORDING, K., RICHARDS, B. & BZDOK, D. 2020. Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nature Communications*, 11.
- SCHWARTING, A., MOCKEL, T., LUTGENDORF, F., TRIANTAFYLLIAS, K., GRELLA, S., BOEDECKER, S., WEINMANN, A., MEINECK, M., SOMMER, C., SCHERMULY, I., FELLGIEBEL, A., LUESSI, F. & WEINMANN-MENKE, J. 2019. Fatigue in SLE: diagnostic and pathogenic impact of anti-N-methyl-D-aspartate receptor (NMDAR) autoantibodies. *Annals of the Rheumatic Diseases*, 78, 1226-1234.
- SCOTT, D. L., WOLFE, F. & HUIZINGA, T. W. J. 2010. Rheumatoid arthritis. *Lancet*, 376, 1094-1108.
- SCOTT, I. A. 2021. Demystifying machine learning: a primer for physicians. *Internal Medicine Journal*, 51, 1388-1400.
- SCOTT, I. C., WHITTLE, R., BAILEY, J., TWOHIG, H., HIDER, S. L., MALLEEN, C. D., MULLER, S. & JORDAN, K. P. 2022. Rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis epidemiology in England from 2004 to 2020: An observational study using primary care electronic health record data. *Lancet Regional Health-Europe*, 23.
- SCOTTI, L., FRANCHI, M., MARCHESONI, A. & CORRAO, G. 2018. Prevalence and incidence of psoriatic arthritis: A systematic review and meta-analysis. *Seminars in Arthritis and Rheumatism*, 48, 28-34.
- SEDAGHAT, A. R. 2019. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. *Otolaryngology-Head and Neck Surgery*, 161, 551-560.

- SEIFERT, O. & BAERWALD, C. 2019. Impact of fatigue on rheumatic diseases. *Best Practice & Research in Clinical Rheumatology*, 33.
- SEJNOWSKI, T. J., POIZNER, H., LYNCH, G., GEPSHTEIN, S. & GREENSPAN, R. J. 2014. Prospective Optimization. *Proceedings of the Ieee*, 102, 799-811.
- SHAH, L. M., CRAMER, J. A., FERGUSON, M. A., BIRN, R. M. & ANDERSON, J. S. 2016. Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and Behavior*, 6.
- SHANG, H., CHU, Q., JI, M., GUO, J., YE, H., ZHENG, S. & YANG, J. 2022. A retrospective study of mortality for perioperative cardiac arrests toward a personalized treatment. *Scientific Reports*, 12.
- SHARP, D. J., BECKMANN, C. F., GREENWOOD, R., KINNUNEN, K. M., BONNELLE, V., DE BOISSEZON, X., POWELL, J. H., COUNSELL, S. J., PATEL, M. C. & LEECH, R. 2011. Default mode network functional and structural connectivity after traumatic brain injury. *Brain*, 134, 2233-2247.
- SHEA, N., BOLDT, A., BANG, D., YEUNG, N., HEYES, C. & FRITH, C. D. 2014. Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18, 186-193.
- SHIMODA, A., ICHIKAWA, D. & OYAMA, H. 2018. Prediction models to identify individuals at risk of metabolic syndrome who are unlikely to participate in a health intervention program. *International Journal of Medical Informatics*, 111, 90-99.
- SHU, N., DUAN, Y., XIA, M., SCHOONHEIM, M. M., HUANG, J., REN, Z., SUN, Z., YE, J., DONG, H., SHI, F.-D., BARKHOF, F., LI, K. & LIU, Y. 2016. Disrupted topological organization of structural and functional brain connectomes in clinically isolated syndrome and multiple sclerosis. *Scientific Reports*, 6.
- SIDDIQI, S. H., TAYLOR, S. F., COOKE, D., PASCUAL-LEONE, A., GEORGE, M. S. & FOX, M. D. 2020. Distinct Symptom-Specific Treatment Targets for Circuit-Based Neuromodulation. *The American journal of psychiatry*, appiajp201919090915-appiajp201919090915.
- SIM, K., GOPALKRISHNAN, V., ZIMEK, A. & CONG, G. 2013. A survey on enhanced subspace clustering. *Data Mining and Knowledge Discovery*, 26, 332-397.
- SIRIMONGKOLKASEM, T. & DRIKVANDI, R. 2019. On regularisation methods for analysis of high dimensional data. *Annals of Data Science*, 6, 737-763.
- SJOGARD, M., WENS, V., VAN SCHEPENDOM, J., COSTERS, L., D'HOOGHE, M., D'HAESELEER, M., WOOLRICH, M., GOLDMAN, S., NAGELS, G. & DE TIEGE, X. 2021. Brain dysconnectivity relates to disability and cognitive impairment in multiple sclerosis. *Human Brain Mapping*, 42, 626-643.

- SMOLEN, J. S., LANDEWE, R. B. M., BERGSTRA, S. A., KERSCHBAUMER, A., SEPRIANO, A., ALETAHA, D., CAPORALI, R., EDWARDS, C. J., HYRICH, K. L., POPE, J. E., DE SOUZA, S., STAMM, T. A., TAKEUCHI, T., VERSCHUEREN, P., WINTHROP, K. L., BALSА, A., BATHON, J. M., BUCH, M. H., BURMESTER, G. R., BUTTGEREIT, F., CARDIEL, M. H., CHATZIDIONYSIOU, K., CODREANU, C., CUTOLO, M., DEN BROEDER, A. A., EL AOUFY, K., FINCKH, A., FONSECA, J. E., GOTTENBERG, J.-E., HAAVARDSHOLM, E. A., IAGNOCCO, A., LAUPER, K., LI, Z., MCINNES, I. B., MYSLER, E. F., NASH, P., POOR, G., RISTIC, G. G., RIVELLESE, F., RUBBERT-ROTH, A., SCHULZE-KOOPS, H., STOILOV, N., STRANGFELD, A., VAN DER HELM-VAN MIL, A., VAN DUUREN, E., VLIET VLIELAND, T. P. M., WESTHOVENS, R. & VAN DER HEIJDE, D. 2023. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. *Annals of the Rheumatic Diseases*, 82, 3-18.
- SMOLEN, J. S., LANDEWE, R. B. M., BIJLSMA, J. W. J., BURMESTER, G. R., DOUGADOS, M., KERSCHBAUMER, A., MCINNES, I. B., SEPRIANO, A., VAN VOLLENHOVEN, R. F., DE WIT, M., ALETAHA, D., ARINGER, M., ASKLING, J., BALSА, A., BOERS, M., DEN BROEDER, A. A., BUCH, M. H., BUTTGEREIT, F., CAPORALI, R., CARDIEL, M. H., DE COCK, D., CODREANU, C., CUTOLO, M., EDWARDS, C. J., VAN EIJK-HUSTINGS, Y., EMERY, P., FINCKH, A., GOSSEC, L., GOTTENBERG, J.-E., HETLAND, M. L., HUIZINGA, T. W. J., KOLOUMAS, M., LI, Z., MARIETTE, X., MUELLER-LADNER, U., MYSLER, E. F., DA SILVA, J. A. P., POOR, G., POPE, J. E., RUBBERT-ROTH, A., RUYSSSEN-WITRAND, A., SAAG, K. G., STRANGFELD, A., TAKEUCHI, T., VOSHAAR, M., WESTHOVENS, R. & VAN DER HEIJDE, D. 2020. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Annals of the Rheumatic Diseases*, 79, 685-699.
- SNOEK, L., MILETIC, S. & SCHOLTE, H. S. 2019. How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*, 184, 741-760.
- SNOW, N. J., WADDEN, K. P., CHAVES, A. R. & PLOUGHMAN, M. 2019. Transcranial Magnetic Stimulation as a Potential Biomarker in Multiple Sclerosis: A Systematic Review with Recommendations for Future Research. *Neural Plasticity*, 2019.
- SOLIVEN, B. & ALBERT, J. 1992. TUMOR-NECROSIS-FACTOR MODULATES THE INACTIVATION OF CATECHOLAMINE SECRETION IN CULTURED SYMPATHETIC NEURONS. *Journal of Neurochemistry*, 58, 1073-1078.
- SONE, D. & BEHESHTI, I. 2022. Neuroimaging-Based Brain Age Estimation: A Promising Personalized Biomarker in Neuropsychiatry. *Journal of Personalized Medicine*, 12.
- SREE, K. & BINDU, C. 2018. Data analytics: Why data normalization. *International Journal of Engineering and Technology (UAE)*, 7, 209-213.
- STANSFELD, S. A., ROBERTS, R. & FOOT, S. P. 1997. Assessing the validity of the SF-36 General Health Survey. *Quality of Life Research*, 6, 217-224.

- STAUD, R. 2012. Peripheral and Central Mechanisms of Fatigue in Inflammatory and Noninflammatory Rheumatic Diseases. *Current Rheumatology Reports*, 14, 539-548.
- STEINLEY, D. 2004. Properties of the Hubert-Arabie adjusted rand index. *Psychological Methods*, 9, 386-396.
- STEKHOVEN, D. J. & BUEHLMANN, P. 2012. MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28, 112-118.
- STERNE, J. A. C., WHITE, I. R., CARLIN, J. B., SPRATT, M., ROYSTON, P., KENWARD, M. G., WOOD, A. M. & CARPENTER, J. R. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj-British Medical Journal*, 339.
- STEVENS, M. C. 2016. The contributions of resting state and task-based functional connectivity studies to our understanding of adolescent brain network maturation. *Neuroscience and Biobehavioral Reviews*, 70, 13-32.
- STOECKEL, L. E., GARRISON, K. A., GHOSH, S. S., WIGHTON, P., HANLON, C. A., GILMAN, J. M., GREER, S., TURK-BROWNE, N. B., DEDETTENCOURT, M. T., SCHEINOST, D., CRADDOCK, C., THOMPSON, T., CALDERON, V., BAUER, C. C., GEORGE, M., BREITER, H. C., WHITFIELD-GABRIELI, S., GABRIELI, J. D., LACONTE, S. M., HIRSHBERG, L., BREWER, J. A., HAMPSON, M., VAN DER KOUWE, A., MACKKEY, S. & EVINS, A. E. 2014. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *Neuroimage-Clinical*, 5, 245-255.
- STRAATHOF, M., SINKE, M. R. T., DIJKHUIZEN, R. M., OTTE, W. M., BUITELAAR, J., DE RUITER, S., NAAIJEN, J., AKKERMANS, S., MENNES, M., ZWIERS, M., ILBEGI, S., HENNISSEN, L., GLENNON, J., VAN DE VONDERVOORT, I., KAPUSTA, K., BIELCZYK, N., AMIRI, H., HAVENITH, M., FRANKE, B., POELMANS, G., BRALTEN, J., HESKES, T., SOKOLOVA, E., GROOT, P., WILLIAMS, S., MURPHY, D., LYTHGOE, D., BRUCHHAGE, M., DUD, I., VOINESCU, B., DITTMANN, R., BANASCHEWSKI, T., BRANDEIS, D., MECHLER, K., BERG, R., WOLF, I., HAEGE, A., LANDAUER, M., HOHMANN, S., BOECKER-SCHLIER, R., RUFF, M., DIJKHUIZEN, R., BLEZER, E., VAN DER MAREL, K., PULLENS, P., MOL, W., VAN DER TOORN, A., OTTE, W., VAN HEIJNINGEN, C., DURSTON, S., MENSEN, V., ORANJE, B., MANDL, R., JOEL, D., CRYAN, J., PETRYSHEN, T., PAULS, D., SAITO, M., HECKMAN, A., BAHN, S., SCHWALBER, A., FLOREA, I. & CONSORTIUM, T. 2019. A systematic review on the quantitative relationship between structural and functional network connectivity strength in mammalian brains. *Journal of Cerebral Blood Flow and Metabolism*, 39, 189-209.
- STREINER, D. L. 2015. Best (but oft-forgotten) practices: the multiple problems of multiplicity-whether and how to correct for many statistical tests. *American Journal of Clinical Nutrition*, 102, 721-728.
- STRICK, P. L., DUM, R. P. & FIEZ, J. A. 2009. Cerebellum and Nonmotor Function. *Annual Review of Neuroscience*, 32, 413-434.

- STROBL, C., BOULESTEIX, A.-L., KNEIB, T., AUGUSTIN, T. & ZEILEIS, A. 2008. Conditional variable importance for random forests. *Bmc Bioinformatics*, 9.
- STROMINGER, N. L., DEMAREST, R. J. & LAEMLE, L. B. 2012. *Noback's human nervous system: structure and function*, Springer Science & Business Media.
- STUDENIC, P., ALETAHA, D., DE WIT, M., STAMM, T. A., ALASTI, F., LACAILE, D., SMOLEN, J. S. & FELSON, D. T. 2023. American College of Rheumatology/EULAR remission criteria for rheumatoid arthritis: 2022 revision. *Arthritis & Rheumatology*, 75, 15-22.
- SUMPTON, D., KELLY, A., TUNNICLIFFE, D. J., CRAIG, J. C., HASSETT, G., CHESSMAN, D. & TONG, A. 2020. Patients' Perspectives and Experience of Psoriasis and Psoriatic Arthritis: A Systematic Review and Thematic Synthesis of Qualitative Studies. *Arthritis Care & Research*, 72, 711-722.
- SWAINSON, R., CUNNINGTON, R., JACKSON, G. M., RORDEN, C., PETERS, A. M., MORRIS, P. G. & JACKSON, S. R. 2003. Cognitive control mechanisms revealed by ERP and fMRI: Evidence from repeated task-switching. *Journal of Cognitive Neuroscience*, 15, 785-799.
- SZUCS, D. & IOANNIDIS, J. P. A. 2020. Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990-2012) and of latest practices (2017-2018) in high-impact journals. *Neuroimage*, 221.
- TAYLOR, W., GLADMAN, D., HELLIWELL, P., MARCHESONI, A., MEASE, P., MIELANTS, H. & GROUP, C. S. 2006. Classification criteria for psoriatic arthritis - Development of new criteria from a large international study. *Arthritis and Rheumatism*, 54, 2665-2673.
- TEAM, R. C. 2021. R: A Language and Environment for Statistical Computing.
- THACH, W. T., GOODKIN, H. P. & KEATING, J. G. 1992. THE CEREBELLUM AND THE ADAPTIVE COORDINATION OF MOVEMENT. *Annual Review of Neuroscience*, 15, 403-442.
- THAKUR, A. 2020. *Approaching (almost) any machine learning problem*, Abhishek Thakur.
- TOWNSEND, M. J. 2014. Molecular and cellular heterogeneity in the Rheumatoid Arthritis synovium: Clinical correlates of synovitis. *Best Practice & Research in Clinical Rheumatology*, 28, 539-549.
- TRACEY, I., WOOLF, C. J. & ANDREWS, N. A. 2019. Composite Pain Biomarker Signatures for Objective Assessment and Effective Treatment. *Neuron*, 101, 783-800.

- TREHARNE, G. J., LYONS, A. C., HALE, E. D., GOODCHILD, C. E., BOOTH, D. A. & KITAS, G. D. 2008. Predictors of fatigue over 1 year among people with rheumatoid arthritis. *Psychology, health & medicine*, 13, 494-504.
- UDDIN, L. Q., YEO, B. T. T. & SPRENG, R. N. 2019. Towards a Universal Taxonomy of Macro-scale Functional Human Brain Networks. *Brain Topography*, 32, 926-942.
- UK, A. R. & . 2014. *Self-help and daily living. Fatigue and arthritis* [Online]. Available: <https://www.versusarthritis.org/media/1269/fatigue-and-arthritis-information-booklet.pdf> [Accessed 14/01/2021].
- URBANOWICZ, R. J., MEEKER, M., LA CAVA, W., OLSON, R. S. & MOORE, J. H. 2018. Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, 85, 189-203.
- VAESSEN, M. J., JANSEN, J. F. A., BRAAKMAN, H. M. H., HOFMAN, P. A. M., DE LOUW, A., ALDENKAMP, A. P. & BACKES, W. H. 2014. Functional and Structural Network Impairment in Childhood Frontal Lobe Epilepsy. *Plos One*, 9.
- VALLADARES-NETO, J. 2018. Effect size: A statistical basis for clinical practice. *Revista Odonto Ciência*, 33, 84-90.
- VALSASINA, P., DE LA CRUZ, M. H., FILIPPI, M. & ROCCA, M. A. 2019. Characterizing Rapid Fluctuations of Resting State Functional Connectivity in Demyelinating, Neurodegenerative, and Psychiatric Conditions: From Static to Time-Varying Analysis. *Frontiers in Neuroscience*, 13.
- VAN BUUREN, S. 2018. *Flexible imputation of missing data*, CRC press.
- VAN BUUREN, S. & GROOTHUIS-OUDSHOORN, K. 2011. mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 45, 1-67.
- VAN HOOGMOED, D., FRANSEN, J., BLEIJENBERG, G. & VAN RIEL, P. 2010. Physical and psychosocial correlates of severe fatigue in rheumatoid arthritis. *Rheumatology*, 49, 1294-1302.
- VAN LOO, H. M., CAI, T., GRUBER, M. J., LI, J., DE JONGE, P., PETUKHOVA, M., ROSE, S., SAMPSON, N. A., SCHOEVERS, R. A., WARDENAAR, K. J., WILCOX, M. A., AL-HAMZAWI, A. O., ANDRADE, L. H., BROMET, E. J., BUNTING, B., FAYYAD, J., FLORESCU, S. E., GUREJE, O., HU, C., HUANG, Y., LEVINSON, D., MEDINA-MORA, M. E., NAKANE, Y., POSADA-VILLA, J., SCOTT, K. M., XAVIER, M., ZARKOV, Z. & KESSLER, R. C. 2014. MAJOR DEPRESSIVE DISORDER SUBTYPES TO PREDICT LONG-TERM COURSE. *Depression and Anxiety*, 31, 765-777.
- VAN ZANTEN, J. J. C. S. V., ROUSE, P. C., HALE, E. D., NTOUMANIS, N., METSIOS, G. S., DUDA, J. L. & KITAS, G. D. 2015. Perceived Barriers, Facilitators and Benefits for Regular Physical Activity and Exercise in Patients with Rheumatoid Arthritis: A Review of the Literature. *Sports Medicine*, 45, 1401-1412.

- VAROQUAUX, G. 2018. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*, 180, 68-77.
- VAROQUAUX, G., RAAMANA, P. R., ENGEMANN, D. A., HOYOS-IDROBO, A., SCHWARTZ, Y. & THIRION, B. 2017. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage*, 145, 166-179.
- VEALE, D. J. & FEARON, U. 2015. What makes psoriatic and rheumatoid arthritis so different? *Rmd Open*, 1.
- VENETSANOPOULOU, A. I., ALAMANOS, Y., VOULGARI, P. V. & DROSOS, A. A. 2023. Epidemiology and Risk Factors for Rheumatoid Arthritis Development. *Mediterranean journal of rheumatology*, 34, 404-413.
- VENKATESH, B. & ANURADHA, J. 2019. A Review of Feature Selection and Its Methods. *Cybernetics and Information Technologies*, 19, 3-26.
- VENTA, A., SHARP, C., PATRIQUIN, M., SALAS, R., NEWLIN, E., CURTIS, K., BALDWIN, P., FOWLER, C. & FRUEH, B. C. 2018. Amygdala-frontal connectivity predicts internalizing symptom recovery among inpatient adolescents. *Journal of Affective Disorders*, 225, 453-459.
- VERHEUL, M. K., FEARON, U., TROUW, L. A. & VEALE, D. J. 2015. Biomarkers for rheumatoid and psoriatic arthritis. *Clinical Immunology*, 161, 2-10.
- VICKERS, A. J. & ALTMAN, D. G. 2001. Statistics notes - Analysing controlled trials with baseline and follow up measurements. *Bmj-British Medical Journal*, 323, 1123-1124.
- VISSER, M. M., MARECHAL, B., GOODIN, P., LILICRAP, T. P., GARCIA-ESPERON, C., SPRATT, N. J., PARSONS, M. W., LEVI, C. R. & BIVARD, A. 2019. Predicting Modafinil-Treatment Response in Poststroke Fatigue Using Brain Morphometry and Functional Connectivity. *Stroke*, 50, 602-609.
- VOGELZANGS, N., BEEKMAN, A. T. F., DORTLAND, A. K. B. V. R., SCHOEVEERS, R. A., GILTAY, E. J., DE JONGE, P. & PENNINX, B. W. J. H. 2014. Inflammatory and Metabolic Dysregulation and the 2-Year Course of Depressive Disorders in Antidepressant Users. *Neuropsychopharmacology*, 39, 1624-1634.
- VREEBURG, S. A., HOOGENDIJK, W. J. G., DERIJK, R. H., VAN DYCK, R., SMIT, J. H., ZITMAN, F. G. & PENNINX, B. W. J. H. 2013. Salivary cortisol levels and the 2-year course of depressive and anxiety disorders. *Psychoneuroendocrinology*, 38, 1494-1502.
- WAHL, B., COSSY-GANTNER, A., GERMANN, S. & SCHWALBE, N. R. 2018. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *Bmj Global Health*, 3.

- WALTER, M. J. M., KUIJPER, T. M., HAZES, J. M. W., WEEL, A. E. & LUIME, J. J. 2018. Fatigue in early, intensively treated and tight-controlled rheumatoid arthritis patients is frequent and persistent: a prospective study. *Rheumatology International*, 38, 1643-1650.
- WANG, H.-T., SMALLWOOD, J., MOURAO-MIRANDA, J., XIA, C. H., SATTERTHWAITE, T. D., BASSETT, D. S. & BZDOK, D. 2018. Finding the needle in high-dimensional haystack: A tutorial on canonical correlation analysis. *arXiv preprint arXiv:1812.02598*.
- WANG, J.-H., ZUO, X.-N., GOHEL, S., MILHAM, M. P., BISWAL, B. B. & HE, Y. 2011. Graph Theoretical Analysis of Functional Brain Networks: Test-Retest Evaluation on Short- and Long-Term Resting-State Functional MRI Data. *Plos One*, 6.
- WARE, M., O'CONNOR, P., BUB, K., BACKUS, D. & MCCULLY, K. 2023. Investigating Relationships Among Interoceptive Awareness, Emotional Susceptibility, and Fatigue in Persons With Multiple Sclerosis. *International journal of MS care*, 25, 75-81.
- WATSON, D. S., KRUTZINNA, J., BRUCE, I. N., GRIFFITHS, C. E. M., MCINNES, I. B., BARNES, M. R. & FLORIDI, L. 2019. Clinical applications of machine learning algorithms: beyond the black box. *Bmj-British Medical Journal*, 364.
- WENDELKEN, C., FERRER, E., GHETTI, S., BAILEY, S. K., CUTTING, L. & BUNGE, S. A. 2017. Frontoparietal Structural Connectivity in Childhood Predicts Development of Functional Connectivity and Reasoning Ability: A Large-Scale Longitudinal Investigation. *Journal of Neuroscience*, 37, 8549-8558.
- WHITFIELD-GABRIELI, S. & NIETO-CASTANON, A. 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2, 125-41.
- WILLIAMS, L. M. 2017. Defining biotypes for depression and anxiety based on large-scale circuit dysfunction: a theoretical review of the evidence and future directions for clinical translation. *Depression and Anxiety*, 34, 9-24.
- WILLIAMS, L. M., KORGAONKAR, M. S., SONG, Y. C., PATON, R., EAGLES, S., GOLDSTEIN-PIEKARSKI, A., GRIEVE, S. M., HARRIS, A. W. F., USHERWOOD, T. & ETKIN, A. 2015. Amygdala Reactivity to Emotional Faces in the Prediction of General and Medication-Specific Responses to Antidepressant Treatment in the Randomized iSPOT-D Trial. *Neuropsychopharmacology*, 40, 2398-2408.
- WINDLE, M. 2016. *Statistical approaches to gene X environment interactions for complex phenotypes*, MIT press.
- WOLFE, F., CLAUW, D. J., FITZCHARLES, M.-A., GOLDENBERG, D. L., HAEUSER, W., KATZ, R. S., MEASE, P., RUSSELL, A. S., RUSSELL, I. J. & WINFIELD, J. B. 2011. Fibromyalgia Criteria and Severity Scales for Clinical and Epidemiological

- Studies: A Modification of the ACR Preliminary Diagnostic Criteria for Fibromyalgia. *Journal of Rheumatology*, 38, 1113-1122.
- WOLFE, F. & PINCUS, T. 2001. The level of inflammation in rheumatoid arthritis is determined early and remains stable over the longterm course of the illness. *Journal of Rheumatology*, 28, 1817-1824.
- WU, H., RUAN, W., WANG, J., ZHENG, D., LIU, B., GENG, Y., CHAI, X., CHEN, J., LI, K. & LI, S. 2021. Interpretable machine learning for covid-19: An empirical study on severity prediction task. *IEEE Transactions on Artificial Intelligence*.
- WU, Q., INMAN, R. D. & DAVIS, K. D. 2014. Fatigue in Ankylosing Spondylitis Is Associated With the Brain Networks of Sensory Salience and Attention. *Arthritis & Rheumatology*, 66, 295-303.
- WU, X., SHI, L., WEI, D. & QIU, J. 2019. Brain connection pattern under interoceptive attention state predict interoceptive intensity and subjective anxiety feeling. *Human Brain Mapping*, 40, 1760-1773.
- WYLIE, G. R. & FLASHMAN, L. A. 2017. Understanding the interplay between mild traumatic brain injury and cognitive fatigue: models and treatments. *Concussion (London, England)*, 2, CNC50-CNC50.
- WYLIE, G. R., YAO, B., GENOVA, H. M., CHEN, M. H. & DELUCA, J. 2020. Using functional connectivity changes associated with cognitive fatigue to delineate a fatigue network. *Scientific Reports*, 10.
- XIA, M., WANG, J. & HE, Y. 2013. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *Plos One*, 8.
- YAN, L., ZHUO, Y., YE, Y., XIE, S. X., AN, J., AGUIRRE, G. K. & WANG, J. 2009. Physiological Origin of Low-Frequency Drift in Blood Oxygen Level Dependent (BOLD) Functional Magnetic Resonance Imaging (fMRI). *Magnetic Resonance in Medicine*, 61, 819-827.
- YAO, Z., ZOU, Y., ZHENG, W., ZHANG, Z., LI, Y., YU, Y., ZHANG, Z., FU, Y., SHI, J., ZHANG, W., WU, X. & HU, B. 2019. Structural alterations of the brain preceded functional alterations in major depressive disorder patients: Evidence from multimodal connectivity. *Journal of Affective Disorders*, 253, 107-117.
- YESHURUN, Y., NGUYEN, M. & HASSON, U. 2021. The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, 22, 181-192.
- YEUNG, H. W., SHEN, X., STOLICYN, A., DE NOOIJ, L., HARRIS, M. A., ROMANIUK, L., BUCHANAN, C. R., WAITER, G. D., SANDU, A.-L., MCNEIL, C. J., MURRAY, A., STEELE, J. D., CAMPBELL, A., PORTEOUS, D., LAWRIE, S. M., MCINTOSH, A. M., COX, S. R., SMITH, K. M. & WHALLEY, H. C. 2021. Spectral clustering based on structural magnetic resonance imaging and its

- relationship with major depressive disorder and cognitive ability. *European Journal of Neuroscience*, 54, 6281-6303.
- YOVEL, G. & KANWISHER, N. 2004. Face perception: domain specific, not process specific. *Neuron*, 44, 889-898.
- ZAEHLE, T. 2021. Frontal Transcranial Direct Current Stimulation as a Potential Treatment of Parkinson's Disease-Related Fatigue. *Brain Sciences*, 11.
- ZALESKY, A., FORNITO, A., COCCHI, L., GOLLO, L. L. & BREAKSPEAR, M. 2014. Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 10341-10346.
- ZHANG, J., CHENG, W., LIU, Z., ZHANG, K., LEI, X., YAO, Y., BECKER, B., LIU, Y., KENDRICK, K. M., LU, G. & FENG, J. 2016. Neural, electrophysiological and anatomical basis of brain-network variability and its characteristic changes in mental disorders. *Brain*, 139, 2307-2321.
- ZHANG, S. 2012. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, 85, 2541-2552.
- ZHI, D., KING, M., HERNANDEZ-CASTILLO, C. R. & DIEDRICHSEN, J. 2022. Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*.
- ZHOU, F., ZHUANG, Y., GONG, H., ZHAN, J., GROSSMAN, M. & WANG, Z. 2016. Resting State Brain Entropy Alterations in Relapsing Remitting Multiple Sclerosis. *Plos One*, 11.
- ZHOU, Y., MILHAM, M. P., LUI, Y. W., MILES, L., REAUME, J., SODICKSON, D. K., GROSSMAN, R. I. & GE, Y. 2012. Default-Mode Network Disruption in Mild Traumatic Brain Injury. *Radiology*, 265, 882-892.
- ZHU, H., LI, T. & ZHAO, B. 2023. Statistical Learning Methods for Neuroimaging Data Analysis with Applications. *Annual Review of Biomedical Data Science*, 6.
- ZUO, X.-N., XU, T. & MILHAM, M. P. 2019. Harnessing reliability for neuroscience research. *Nature Human Behaviour*, 3, 768-771.
- ZWARTS, M. J., BLEIJENBERG, G. & VAN ENGELEN, B. G. M. 2008. Clinical neurophysiology of fatigue. *Clinical Neurophysiology*, 119, 2-10.