

WP22: Image processing pipelines

PI/Lead: Clare Mackay, Ludovica Griffanti

Overview: In this work package we will adapt the UK Biobank brain MRI pipeline to make it applicable to data from DPUK cohorts and accessible from the DPUK remote analysis environment. Moreover, pre-processed data (imaging derived phenotypes, IDPs) will be available on the DPUK Data Portal for (some/three) pilot datasets. Detailed documentation will be available to run the pipeline and derive IDPs from new cohorts as well as facilitate IDPs access and use from non-imaging experts. Particular attention will be given to the aspect of data harmonisation to generate IDPs that are generalisable, comparable across datasets and facilitate cross-cohort analyses.

The close link with UK Biobank in terms of processing and IDPs extraction will enable merging the strengths of UK Biobank in terms of statistical power and representation of the general population, with the specific clinical phenotyping of dementia cohorts. The present work package will focus on brain MRI, while future extension to nuclear imaging (PET, SPECT) will be considered in a future project.

Aim: To provide rapid and standard brain MRI pre-processing for research datasets making imaging datasets more accessible for analysts and facilitating cross-cohort analyses.

Background:

The importance of imaging data for dementia research grows. However, the complexity of imaging data makes it relatively inaccessible for non-specialist analysts.

Led from the Oxford Centre for Human Brain Activity (OHBA) and building on work with the Wellcome Centre for Integrative Neuroimaging (WIN), this work package will establish pre-processing pipelines for imaging data, specifically brain MRI data. These pipelines derive widely used phenotypes from raw imaging data. Image-derived phenotypes (IDPs) can be readily integrated with data from other modalities in order to test complex multi-modal hypotheses. These pipelines will be made available to the wider research community via the Data Portal.

Objective 1: Adapt the UK Biobank processing pipeline to make it applicable to data from the DPUK cohorts

Deliverable 1: Recruit postdoctoral researcher

Deliverable 2: Adapt UK Biobank processing pipeline to different number of modalities and sequence parameters

Deliverable 3: Add pipelines and/or IDPs relevant to DPUK cohorts

Deliverable 4: Implementation of the pipeline on the Data Portal

Objective 2: Apply the pipeline to currently available DPUK cohorts and release IDPs

Deliverable 5: Identify suitable pilot dataset and apply the pipeline to derive IDPs

Deliverable 6: Make IDPs available on the Data Portal for the preselected pilot datasets

Deliverable 7: Support application of the pipeline on other datasets

Objective 3: Generate documentation to use the pipeline and interpret imaging derived phenotypes

Deliverable 8: User guides for pipeline users

Deliverable 9: IDPs documentation for non-imaging experts

Deliverable 10: Feedback and support system

Objective 4: Develop and apply harmonisation techniques to generate comparable IDPs across cohorts

Deliverable 11: Develop criteria to evaluate the success of a harmonisation technique (dataset- and modality-dependent)

Deliverable 12: Adopt or incorporate harmonisation techniques in the processing pipeline to produce generalisable IDPs.

Deliverable 13: Develop and/or apply harmonisation techniques on the extracted IDPs to facilitate cross-cohort analyses.

Objective 5: DPUK/MRC project reporting

Deliverable 14: Quarterly reporting

Deliverable 15: Annual reporting

Deliverable 16: End report

Deliverable 17: Financial reporting

Objective 1: Adapt the UK Biobank processing pipeline to make it applicable to data from the DPUK cohorts

Description	How and who	Outcome	Dates
Deliverable 1: Recruit Postdoctoral researcher			
M1.1.1 Advertise Job	Develop job description and advertise role (CM, LG)	Job advertised	Jun-21
M1.1.2 Recruit postdoctoral researcher	Interview applicants and appoint selected candidate (CM, LG)	Postdoctoral researcher appointed	Sep-21
Deliverable 2: Adapt UK Biobank processing pipeline to different number of modalities and sequence parameters			
M1.2.1 Identification of the adaptations needed in terms of input (e.g. flexibility on the number and type of sequences present in a dataset, Gradient Distortion Correction on/off) and which parameters	Review of UKB code to identify dependencies and built-in parameters (Postdoc)	Adaptations identified and documented	Dec-21

need to be flexible/customisable within each sequence (e.g. dMRI parameters)			
M1.2.2 Code modification	Python code modification (Postdoc)	Adapted basic UKB pipeline	Dec-22
Deliverable 3: Add pipelines and/or IDPs relevant to DPUK cohorts			
M1.3.1 Identification of additional pipelines and/or IDPs to include in the DPUK Data Portal which could be particularly useful for dementia cohorts (e.g. ASL metrics, sub-classification of vascular lesions)	Literature review and inventory of available sequences in DPUK cohorts (Postdoc)	List of desired IDPs	Dec-21
M1.3.2 Inclusion of the new pipelines/IDPs	Python code modification (Postdoc)	Adapted advanced pipeline	Dec-22
Deliverable 4: Implementation of the pipeline on the Data Portal			
M1.4.1 Specification and set-up of the necessary software, dependencies, environment to support the pipeline on the Data Portal	Specification developed and liaison with Swansea team during development (Postdoc)	Environment set-up and tested	Jun-23
M1.4.2 Pipeline implementation on the Data Portal	Liaison with Swansea team during implementation (Postdoc)	Pipeline released on the DPUK Data Portal	Dec-23

Objective 2 Apply the pipeline to currently available DPUK cohorts and release IDPs			
Description	How and who	Outcome	Dates
Deliverable 5: Identify suitable pilot dataset and apply the pipeline to derive IDPs			
M2.5.1 Identification of the most suitable datasets to be used as pilot, depending on availability on the Data Portal and amount and type of data and MRI modalities. Potential candidates: NHSD, Whitehall, CamCAN	Review of datasets and liaison with Oxford and Swansea teams (Postdoc)	Pilot datasets defined	Dec-21
M2.5.2 Apply the pipeline to pilot datasets	Liaison with Swansea team (Postdoc)	IDPs calculated for the pilot datasets	Jun-23
Deliverable 6: Make IDPs available on the Data Portal			
M2.6.1 Setup automated organisation and storage of the IDPs produced by the pipeline in the Data Portal	Specification of storage and liaison with Swansea team to implement storage capability (Postdoc)	Automated storage and release of IDPs on the Data Portal available as	Jun-23

		functionalities on the platform	
M2.6.2 Release IDPs for the pilot datasets on the Data Portal	Liaison with Swansea team and testing (Postdoc)	IDPs for the pilot datasets available on the Data Portal	Dec-24
Deliverable 7: Support the application the pipeline on other datasets			
M2.7.1 Define workflow of data input, processing and IDPs output so that the adapted UKB pipeline (and non-UKB pipelines developed by DPUK users) can be easily applied to new cohort datasets	Specification of workflow and development of documentation (Postdoc)	Internal documentation on the data workflow for image processing	Dec-24
M2.7.2 Identify datasets minimum requirements to run the adapted UKB pipeline (e.g. resolution range, minimum number of diffusion directions etc.)	Test generalisability of the pipeline on different data parameters on simulations and real datasets (Postdoc)	Data validation checkpoints included in the pipeline	Jun-25

Objective 3 Generate documentation to use the pipeline and interpret imaging derived phenotypes

Description	How and who	Outcome	Dates
Deliverable 8: User guides for pipeline users			
M3.8.1 Write user guides for the adapted pipeline for users accessing raw data and running the pipeline	Development of user guides and liaison with Swansea team to add to a GitLab repository on the DPUK Data Portal (Postdoc)	Pipeline documentation released online	Jun-25
M3.8.2 Detail minimum requirements to run the pipeline on a new dataset (see M2.6.2)	Development of requirements and liaison with Swansea team to add to the GitLab repository on the DPUK Data Portal (Postdoc)	Dataset requirements documentation released online	Jun-25
Deliverable 9: IDPs documentation for non-imaging experts			
M3.9.1 Write documentation on the IDPs generated by the pipeline for users accessing IDPs only	Development of documentation and liaison with the Swansea team to add to the GitLab repository on the DPUK Data Portal (Postdoc)	IDPs documentation released online	Jun-25
Deliverable 10: Feedback and support system			

M3.10.1. Beta testing of the pipeline and documentation internally and potentially with selected users of the DPUK analysis/Data Platform (potentially through a datathon event)	Testing, debugging and improving the robustness of the pipelines (Postdoc)	Issues identified and resolved at the end of the beta testing period	Jun-25
M3.10.2 Enable a system for reporting feedback and issues with the documentation and/or the pipeline	Specification of required feedback and liaison with Swansea team to add to the GitLab repository on the DPUK Data Portal (Postdoc)	Feedback system active on the Gitlab repository	Dec-25
Objective 4 Develop and apply harmonisation techniques to generate comparable IDPs across cohorts			
Description	How and who	Outcome	Dates
Deliverable 11: Develop criteria to evaluate the success of a harmonisation technique (dataset- and modality-dependent)			
M4.11.1 Define go/no-go criteria in relation to the possibility of a dataset or modality to be harmonised	Development of criteria (Postdoc)	Criteria specified and approved	Jun-22
M4.11.2 Define success criteria for datasets and modalities that can be harmonised (define outcome measure – e.g. volume difference within the range of scan-rescan variability)	Development of success criteria (Postdoc)	Success criteria specified and approved	Jun-22
Deliverable 12: Adopt or incorporate harmonisation techniques in the processing pipeline to produce generalisable IDPs			
M4.12.1 Identify sources of variation in the images (with specific focus on the pilot datasets) and identify approaches to remove or compensate non-biological variability (e.g. bias field, contrast)	Analysis of IDPs (Postdoc)	Sources of variation identified and approach to remove variability defined	Dec-22
M4.12.2 Implement harmonisation techniques in the processing pipeline.	Specification of required harmonisation techniques and liaison with Swansea team to implement (Postdoc)	Harmonisation steps included in the pipeline	Dec-25
Deliverable 13: Develop and/or apply harmonisation techniques on the extracted IDPs to facilitate cross-cohort analyses			

M4.13.1 Identify sources of variation in the IDPs (with specific focus on the pilot datasets) and identify approaches to remove or compensate non-biological variability (e.g. deconfounding)	Analysis of IDPs (Postdoc)	Sources of variation identified and approach to remove variability defined	Dec-22
M4.13.2 Implement harmonisation techniques in the processing pipeline or as additional postprocessing steps.	Specification of required harmonisation techniques and liaison with Swansea team to implement (Postdoc)	Harmonisation (post-) processing steps included in the pipeline	Dec-25

Objective 5: DPUK/MRC project reporting				
Specifics	Description	How and who	Outcome	Dates
Deliverable 14: Produce quarterly reports by the required dates				
M5.14.1	Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Mar 2021
M5.14.2	Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Jun 2021
M5.14.3	Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Sep 2021
M5.14.4	Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Mar 2022
M5.14.5	Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Jun 2022
M5.14.6	Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Sep 2022

M5.14.7 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Mar 2023
M5.14.8 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Jun 2023
M5.14.9 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Sep 2023
M5.14.10 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Mar 2024
M5.14.11 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Jun 2024
M5.14.12 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Sep 2024
M5.14.13 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Mar 2025
M5.14.14 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Jun 2025
M5.14.15 Provide Quarterly reports detailing project deliverables and outcomes.	Online quarterly form to be completed for DPUK for MRC meetings	Quarterly report submitted	Sep 2025
Deliverable 15: Produce an annual report by the required dates			
M5.15.1 Annual reports to oversee project status and updates of deliverables and outputs. Information to assess completion criteria as part of payment schedule.	Annual form to be completed for DPUK for MRC meetings. Annual reports submitted to confirm milestone completion for next payment scheduled	Annual report submitted	Dec 2021

M5.15.2 Annual reports to oversee project status and updates of deliverables and outputs. Information to assess completion criteria as part of payment schedule.	Annual form to be completed for DPUK for MRC meetings. Annual reports submitted to confirm milestone completion for next payment scheduled	Annual report submitted	Dec 2022
M5.15.3 Annual reports to oversee project status and updates of deliverables and outputs. Information to assess completion criteria as part of payment schedule.	Annual form to be completed for DPUK for MRC meetings. Annual reports submitted to confirm milestone completion for next payment scheduled	Annual report submitted	Dec 2023
M5.15.4 Annual reports to oversee project status and updates of deliverables and outputs. Information to assess completion criteria as part of payment schedule.	Annual form to be completed for DPUK for MRC meetings. Annual reports submitted to confirm milestone completion for next payment scheduled	Annual report submitted	Dec 2024
Deliverable 16: Provide annual financial reporting against the specified budget by the required dates			
M5.16.1 Financial reports submitted at the end of each year. Information to assess completion criteria as part of payment schedule.	Yearly financial statement on spending	Financial report submitted	Dec 2021
M5.16.2 Financial reports submitted at the end of each year. Information to assess completion criteria as part of payment schedule.	Yearly financial statement on spending	Financial report submitted	Dec 2022
M5.16.3 Financial reports submitted at the end of each year. Information to assess completion criteria as part of payment schedule.	Yearly financial statement on spending	Financial report submitted	Dec 2023
M5.16.4 Financial reports submitted at the end of each year. Information to assess completion criteria as part of payment schedule.	Yearly financial statement on spending	Financial report submitted	Dec 2024

