Exploring the relationship between multimodal magnetic resonance neuroimaging and cognitive outcomes in children

Applying machine learning algorithms to brain MRI features to predict cognitive scores and performance categories of children living with and without HIV.

Prepared by Isaac Lebogang Khobo
BSc (Eng) Mechatronics, MSc Biomedical Engineering, UCT

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Biomedical Engineering

Supervisor: Dr Frances C Robertson, PhD
Co-supervisor: Prof Ernesta M Meintjes, PhD
Department of Human Biology
Faculty of Health Sciences
University of Cape Town
February 2024

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Declaration

I, Isaac Lebogang Khobo, present this thesis in fulfilment of the requirements for the degree of Doctor of Philosophy in Biomedical Engineering in the Department of Human Biology, Faculty of Health Sciences, University of Cape Town. I hereby declare that this dissertation is my own original work, except where acknowledgements indicate otherwise, and that neither the whole, nor any part of it, has been, or is to be submitted for any other degree in this or any other university. This thesis has been submitted to the Turnitin module and I can confirm that my supervisor has seen the report and there were no plagiarism concerns. I empower the university to reproduce, for the purpose of research, either the whole or part of the contents of this thesis in any manner.

Preface

I conducted the work in this dissertation between September 2020 and December 2023 at the Neuroscience Institute, University of Cape Town, supervised by the intelligent and good-natured Dr Frances Robertson. In my academic journey, Dr Frances Robertson has also played the roles of mentor, editor, and friend so well that she is the second person in the world I am now genuinely terrified of disappointing. Dr Frances Robertson, I hope you continue guiding me in the best way possible and thank you very much for all the support you have given me so far.

The project arose from the need by our research team, led by Prof Ernesta Meintjes, to establish a clear link between multimodal MRI neuroimaging and future cognitive performance of children. We wished to understand whether neuroimaging would be reliable an auxiliary tool for staging and screening for cognitive deficits in children living with and without HIV. Prof Ernesta Meintjes, I am grateful for your co-supervising this project, providing valuable knowledge and resources to conduct the research as well as expert feedback on the consistency and clarity in my writing.

Neuroimaging scans were acquired at Cape Universities Brain Imaging Centre (CUBIC) which was set up in 2007 as a joint initiative between Siemens, Stellenbosch University, University of Cape Town, and South African Medical Research Council (SAMRC). The Faculty of Health Sciences Human Research Ethics Committee approved the methods (HREC ref: 453/2021). A special thanks to you radiographers at CUBIC for your help in the scanning of the children. Although I was not involved in the scanning process, I did have the opportunity to meet most of the children at a symposium in Tygerberg Hospital in 2019. It was inspiring to see you dance and to hear your stories, poetry, goals, and challenges of living with HIV. I would like to thank you for your courage and your parents, caregivers, and guardians for supporting you in your journey.

Cognitive assessments were performed by our collaboration team and licensed psychologists at Family Clinical Research Unit (FAMCRU), Department of Paediatrics and Child

Health, Tygerberg Children's Hospital, Stellenbosch University. I appreciate your help Prof Barbara Laughton and Dr Kaylee van Wyhe for answering my clinical and demographic questions as well as clarifying aspects of the cognitive scores I initially found difficult to follow.

For the course of the research, I received funding from the National Research Foundation (NRF) freestanding PhD scholarship of South Africa's Department of Science and Innovation, Department of Human Biology scholarships via the NRF grant for the South African Chairs Initiative, University of Cape Town's international conference travel scholarships as well as the Accelerated Transformation of the Academic Program (ATAP) fellowship. I am incredibly grateful to you my funders and supervisors who ensured I always had financial support throughout my research.

This PhD dissertation is written by publication format, with Chapters 2, 3, and 4 comprising the main articles. Chapter 1 introduces the work and Chapter 5 synthesises the overall results, discusses implications of them, details the limitations of the study, recommends future improvements, and concludes the work. None of the main chapters are yet submitted to journals. However, results for Chapter 3 were presented at the annual Organisation of Human Brain Mapping meeting (OHBM) in Glasgow (2022), and those for Chapter 4 at the International Society of Magnetic Resonance in Medicine (ISMRM) African chapter meeting in Accra (2023). The work was also presented at the university's research day in 2021 and departmental joint symposiums in 2021 and 2022.

Thank you, Prof André van der Kouwe at Harvard Medical School for the scanning sequences and your input and suggestions for conference abstracts and poster presentations. Finally, I would like to acknowledge the experienced advice my academic mentor Prof Thomas Franz has given me regarding managing my time, workload, and preparing for an academic career. I hope the research community, educators, and healthcare professionals can use this dissertation as a valuable reference.

—Isaac Lebogang Khobo, February 2024

This dissertation is dedicated to my beloved family and friends who are always there when I need them the most. And to the memory of mama Lobisa Khobo.

Synopsis

Background: A constantly growing body of literature shows that children from low-socioeconomic status (SES) backgrounds are at risk of cognitive developmental delays, poor health outcomes, and cognitive difficulties which lead to high rates of school drop-outs and struggles in other areas of life. In sub-Saharan Africa, where the human immunodeficiency virus (HIV) is the most prevalent, low-SES households and communities are disproportionately affected by the disease and its effects on neurodevelopment. The ability to predict cognitive abilities or deficits from neuroimaging or other methods could make it easier to identify at-risk children who may benefit the most from targeted interventions. This is of relevance in low-SES populations with relatively high rates of childhood HIV that may affect neurodevelopment. Magnetic resonance (MR) imaging (MRI) is a versatile tool that can be used to measure a broad range of brain tissue properties giving rise to cognitive functions. For example, structural MRI (sMRI) can quantify brain volumes and other morphometrics, diffusion tensor imaging (DTI) can estimate the amount of nerve fibre damage, and proton MR spectroscopy (^{1}H-MRS) can characterise the biochemical profile of grey and white matter (GM, WM).

Research aims: The aims of this study were: First, collect evidence for what is known about the relationship between cognitive performance assessed by a comprehensive set of cognitive test batteries and brain changes measured with neuroimaging in children, adolescents, and youth living with HIV. Second, compare the predictive performance of penalised linear models (PLMs), support vector machines/regression (SVM/R), and decision tree ensembles (DTEs) in predicting continuous scores on cognitive tests, as well as categories of cognitive performance from multimodal neuroimaging in a cohort comprising both children living with and without HIV. Third, determine whether multimodal MRI offers any predictive advantage compared to predicting future performance using cognitive scores at a younger age.

Methods and materials: To address these aims, we first conducted a systematic literature review and secondly a multimodal MRI neuroimaging and cognitive testing study of 132 children from low-SES backgrounds. For the review, we searched PubMed, Scopus, Web of Science, CINAHL, APA Psych Info & Psych Articles, and Academic Search Premier for studies published between 1 January 2006 and 31 October 2022. Inclusion criteria were studies that investigated a relationship between neuroimaging brain measures and cognitive test scores and included children (0–14 years), adolescents (15–18), and youth (19–26) living with HIV.

For the neuroimaging and cognitive study, structural MRI, DTI, and ^{1}H-MRS were acquired at ages 7 and 9 years. Cognitive performance was assessed using the Kaufman assessment battery for children, Beery-Buktenica developmental test of visual-motor integrations, test of variables of attention, Purdue pegboard test, the Peabody picture vocabulary test, and semantic fluency test at both ages. PLMs, SVMs/R, and DTEs prediction models were implemented with Bayesian optimization and assessed with 10-fold cross validation (CV) and compared for their ability to predict continuous scores (regression) or categories of cognitive performance (classification). Poorer and better cognitive performance categories were identified with a hierarchical clustering algorithm. Regression performance was assessed via 10-fold CV errors, coefficient of determination (R^2), and Pearson's r between predicted and actual values. For the classification models, 10-fold CV sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) were obtained.

Results: Evidence from the literature suggests that HIV may lead to alterations in the brain's structure, function, neurometabolism, and WM microstructure. Individual brain measures are linked to outcomes of short-term memory, processing speed, working memory, problem solving, and general intelligence quotients in children, adolescents, and youth living with and without. We could not find any studies linking multimodal MRI to cognitive performance in this population of young people.

PLMs, SVMs/SVR, and DTEs performed poorly for the regression problem; the predictive models led to small training and fitting errors but high generalised CV errors. However, using either multimodal MRI data or cognitive scores at age 7, we could predict auditory working memory ($R^2 = 0.45$, $r = 0.75$), short-term memory ($R^2 = 0.43$, $r = 0.62$), visual-motor integration ($R^2 = 0.26$, $r = 0.39$), and executive reasoning ($R^2 = 0.33$, $r = 0.27$) scores at age 9 with moderate to strong Pearson's r. Classification of children into poorer or better performance categories was more successful than regression of the individual scores, with 0.75–0.81 AUC, 70–77% accuracies, 70–81% specificities, 71–79% sensitivities using historic multimodal MRI and cognitive scores. Historic multimodal MRI (AUC = 0.80, accuracy = 76%) was marginally better than cognitive scores (AUC = 0.75, accuracy = 70%) in classifying future overall cognitive performance.

Conclusion: There were multimodal brain measures relevant in the prediction models, these included creatine and glutamate concentrations in midfrontal gray matter region, thalamus volume, diffusivity in the cingulum WM tract, cingulate gyrus area, and gyrification index of the parietal lobe. This suggests that multiple MRI modalities and features should be considered simultaneously to establish correlates of overall cognitive performance. The neural correlates we find could potentially be used to identify biomarkers of cognitive impairment, understand the developmental nature of cognitive plasticity, and enable the development of targeted interventions that can modulate brain networks associated with cognitive functions.

Table of contents

1

INTRODUCTION: exploring the relationship between multimodal magnetic resonance neuroimaging and cognitive outcomes in children

Chapter 1 provides the problem statement, brief background on neuroimaging, cognitive testing, and machine learning, aims and objectives of the study, and finally an outline of the whole dissertation.

1.1 Background

Socioeconomic status (SES), cognitive outcomes and brain development

It is important to understand cognitive abilities of children from low-SES backgrounds, since extensive reviews of the literature show that they experience neurodevelopmental delays (Rakesh et al., 2023), high rates of school dropouts/absenteeism (Sosu et al., 2021), and poor health outcomes (David et al., 2019). Possible reasons are limited access to healthcare resources and their environment being less likely to encourage the development of fundamental skills of language, vocabulary acquisition, working memory, attention, socioemotional processing, and general intellectual functioning. As a consequence, these children are also likely to have low income and poor health in adulthood. Although SES is a complex construct that encompasses factors such as household income, educational level, job security, and individual perceptions of social status and class, it has a measurable impact on cognitive development, brain structure, and the interplay between brain function and cognitive abilities.

A recent comprehensive umbrella review shows that children from higher-SES backgrounds consistently perform better in multiple cognitive areas, such as language, memory, and executive functions, compared to their lower-SES peers (Tan, 2024). This disparity is also attributed to access to educational resources and parental involvement. This poor performance in low-SES communities affects academic achievement and results

in the reduced graduation rates, reduced propensity to seek higher education opportunities, and eventual inability to complete school (Sosu et al., 2021).

SES also plays an important role in how the human brain develops. For example, studies have found associations between neighbourhood poverty, prefrontal and hippocampal volume, and cognitive performance (Taylor et al., 2020); children from low-SES backgrounds often exhibit smaller amygdala and hippocampal volumes as well as reduced cortical thickness (Abo Hamza et al., 2024). Additionally, recent findings indicate that SES circumstances of children and their families influence the response to early-life brain injuries in children born premature (Benavente-Fernandez et al., 2020). Furthermore, SES has also been found to have mediating effects on brain-cognition relationships (Thomas & Coecke, 2023). For example, associations between SES, cortical thickness, and surface area mediate differences in cognitive skills of language and executive function between (ibid.).

Perinatal HIV, cognitive and brain development

In post-apartheid South Africa, which carries the largest share of the global HIV burden, low-SES households and communities are disproportionately affected by the disease (Cleary et al., 2011; Wabiri & Taffa, 2013). Therefore, for children residing in low-SES South African communities, normal brain development is potentially hindered by perinatally-acquired HIV (Musielak & Fine, 2016; van den Hof et al., 2019), neighbourhood poverty and household SES (McDermott et al., 2019; Taylor et al., 2020), or a combination of these factors. Further, HIV has been shown to affect selective developmental/cognitive domains of learning, executive function, and hearing (Musindo et al., 2022; Rowe et al., 2021). Although there is a wealth of data from high-income countries, it is important to investigate the relationship between brain structure and cognitive performance in a population representative of a typical low- or middle-income country setting.

Neuroimaging and cognitive testing are two widely-used methods for understanding paediatric cognitive development/performance in order to monitor treatment efficacy and establish effective intervention strategies. Ultimately, the ability to predict cognitive abilities or deficits from neuroimaging and/or cognitive testing could make it easier to identify at-risk children living in low-SES communities who may benefit the most from earlier targeted interventions.

Paediatric magnetic resonance neuroimaging

Over the past decade, neuroimaging has shown its usefulness in the identification of infant cognitive impairment (Schadl et al., 2018), neurodevelopmental delays (Ouyang et al., 2020), as well as clinical diagnosis and treatment monitoring of attention-deficit/hyperactivity disorder (ADHD) in children with attention problems (Firouzabadi et al., 2022). Furthermore, neuroimaging has been used to characterise HIV-associated cognitive deficits from adolescence to adulthood (Nichols, 2022), although there are still no clear neuroimaging or cognitive criteria for diagnosing HIV-related cognitive impairment in children (Hoare et al., 2014). Nevertheless, neuroimaging has uncovered previously unknown effects of neurological and environmental factors on paediatric brain structure and function.

Magnetic resonance (MR) imaging (MRI) is a valuable brain imaging tool, because of its excellent soft tissue contrast, resolution, signal-to-noise ratio, absence of ionizing radiation, and the ability to examine different properties of tissue using the same scanner. Various non-invasive MR neuroimaging modalities exist: for example, structural MRI (sMRI), which may be used for quantifying the brain's morphometric measures like regional and total volumes, cortical thicknesses, gyrification, areas, depths, and curvatures (Symms et al., 2004). Diffusion MRI or diffusion tensor imaging (DTI) is sensitive to grey and white matter (GM and WM) microstructural alterations, and can detail the distribution and orientation of nerve fibres and myelin using water diffusion (Martinez-Heras et al., 2021; Tournier, 2019). And proton magnetic resonance spectroscopy (^{1}H-

MRS) can detect metabolic and chemical changes in brain regions (Bertholdo et al., 2013; Novotny et al., 1998). There is little doubt that MRI is one of the most versatile neuroimaging tools preferred by clinicians and researchers alike. It is utilised to understand paediatric brain development, specifically in terms of GM reductions from cellular pruning and WM increases from myelination and connectivity that typically occur in the first 2 decades of life, namely: childhood, adolescence, and early adulthood. These MRI modalities are also increasingly used to find neural correlates of cognitive processes—the brain structures whose functional activity, interconnections, chemistry, and metabolic activity may explain variations in mental processes such as attention, working memory, and general intelligence that are impacted by diseases such as perinatal HIV, and environmental factors, for example, household SES.

However, MRI neuroimaging remains expensive and less accessible to children from low-SES backgrounds (Geethanath & Vaughan, 2019). Clinically, children presenting with severe symptoms from gross neurological disorders, for example, traumatic brain injury and tumours, would likely be able to obtain an MRI scan at a public hospital in South Africa. However, large-cohort research studies using more advanced imaging can detect more subtle metabolic or microstructural brain alterations from diseases such as HIV. But these research studies do not currently provide information that is useful to the individual, e.g., diagnosis for developmental disorders like ADHD. Therefore, to be of most benefit, MRI neuroimaging should provide information not readily available through less expensive means. In general, such information may include understanding biomarkers of cognitive impairment, the developmental course of cognitive plasticity, and how to modulate brain networks associated with cognitive function.

Cognitive assessment batteries, scales, or neuropsychological testing

Relative to neuroimaging, cognitive assessment tests are easier to administer, and remain the most accessible and widely used means for evaluation of typical development and diagnosis of paediatric cognitive disorders for children in low to middle-income countries.

The tests are designed to index mental functions that are required for success in educational attainment and goal achievement in life, and are therefore, more closely related to life outcomes than neuroimaging measures. Various tests assessing aspects of cognitive development exist. For example, Beery-Buktenica developmental test of visual-motor integrations is a non-verbal assessment of the extent to which individuals can integrate their visual and motor abilities (Beery & Beery, 2004). Test of variables of attention is a diagnostic measure of attention deficits in children (Leark et al., 2004). Purdue pegboard test is a measurement of unimanual and bimanual finger and hand dexterity (Spreen & Strauss, 2006). Peabody picture vocabulary test is an untimed test of receptive vocabulary (Dunn & Dunn, 2007). And the Kaufman assessment battery for children with subtests for general knowledge, memory, reasoning, and others, is a measure of overall cognitive development for children aged 3 to 18 years (Kaufman & Kaufman, 2014).

There are, however, several limitations of paediatric cognitive assessment testing. For example, the tests have been shown to have poor sensitivity and specificity to learning disability or ADHD (Fletcher & Miciak, 2017). Some tests can be failed due to linguistic and cultural differences of children from different backgrounds and regions than those of which the normative data is based (Howieson, 2019), and an incorrect administration and scoring is likely to skew results (Ellingsen, 2016). Moreover, since these tests assess the central nervous system indirectly via behaviour and language acquisition, they cannot provide the neurobiological basis of cognitive deficits or normal cognitive processes. As such, using these cognitive assessment batteries alone can possibly make interpretations and interventions unclear. Neuroimaging may be a reliable complement to cognitive testing.

Machine learning algorithms

Multimodal MRI neuroimaging and cognitive assessment data can be high dimensional in nature. The data often consist of many predictors, such as multiple brain measures

and test scores, in addition to a limited number of observations from people who have been scanned and cognitively assessed. The high dimensionality of the data can present challenges for analysis; however, machine learning (ML) techniques can be used to identify patterns and relationships in the data. An ML model is any algorithm that takes a known set of input data and known responses to the data, learns information directly from the data without relying on a predetermined equation, then generates reasonable predictions for the response to new/unseen data. There are many ML techniques, some of which have been applied to understand brain development across all ages. ML techniques are broadly categorised into four main paradigms: supervised, unsupervised, semi-supervised, and reinforcement learning (Vu et al., 2018). In supervised learning, the algorithm is trained using labelled data, i.e., input-output pairs that make it easier to predict outcomes for unseen data. For example, an fMRI study may use a supervised learning classifier to diagnose autism spectrum disorder (Santana et al., 2022). But for unsupervised learners, the data is unlabelled, and the algorithm tries to find hidden patterns in the input data without any explicit guidance on the output. An example would be tumour identification and tissue segmentation (annotation) in MRI images using an unsupervised learning method with a clustering approach (Vishnuvarthanan et al., 2016). Semi-supervised learners are hybrid algorithms that use both labelled and unlabelled data. For example, segmentation of large number of three-dimensional brain images by first labelling a relatively small number of images and leveraging co-registration and supervised learning to attach "pseudo" labels to images that were originally unlabelled (Ito et al., 2019). Lastly, reinforcement learning is an extension of supervised learning, where the algorithm is trained to make decisions through a dynamic trial-and-error process to maximise a desired outcome. Instead of creating an input-output pair, a "scoring function" is created to tell the algorithm how good its previous outcome was, and subsequent iterations the algorithm maximises the outcome (Vu et al., 2018). This type of learning is common in the field of gaming (e.g., AlphaZero models trained to

play chess and go), natural language processing, robotics, and computer vision (Fan et al., 2023).

Each paradigm has known models or algorithms to solve complex problems, e.g., dimensionality reduction, classification, regression, and clustering of high-dimensional multimodal MRI and multidomain cognitive data. These sub-algorithms include classifiers and regression models such as support vector machines, decision trees, k-nearest neighbour algorithm, naïve bayes algorithm, and logistic regression used to predict a response that is continuous or belongs to a set of categories. Lastly, ML algorithms offer trade-offs between predictive performance and their ability to produce interpretable and explainable predictions (Linardatos et al., 2020). Interpretable techniques reveal how the ML model makes predictions by uncovering how various features contribute, or do not contribute, to the predictive accuracy. As such, the choice of ML algorithm depends on the data at hand and the specific outcomes, e.g., feature selection; therefore, it can be challenging to generalise for all data types.

Supervised learning and predictive modelling for cognitive development

Nevertheless, there has been a growing interest in recent years to use supervised ML modelling to solve complex problems in neuroscience, such as to understand the relationship between early brain structural development and cognitive functioning (Scheinost et al., 2022). For example, support vector machines have been used to predict categories of neurodevelopmental outcomes at age 2 years (Ouyang et al., 2020); random forest regression has been used to predict future executive function of children aged 5 to 14 years (Khundrakpam et al., 2022); and multi-task multi-linear regression has been used to predict future language scales of infants (Adeli et al., 2019). All these applications of ML on neuroimaging data have demonstrated the possibility to predict current and future cognition of healthy children.

The relative efficacy of different ML algorithms for different types of multimodal neuroimaging and multi-domain cognitive assessment data in predicting cognitive performance of children has not been thoroughly investigated. Moreover, there are no studies investigating the prediction of cognitive performance of school-aged children from low-SES backgrounds or in populations from Africa comprising children living with conditions such as HIV.

1.2 Research aims

Against this background, there are areas of research in the literature that remain unexplored. In this study, we aimed to predict cognitive performance of children from low-SES backgrounds disproportionately affected with HIV using multimodal MRI neuroimaging. The cognitive outcomes we sought to predict covered the crucial domains of attention, processing speed, short-term memory, working memory, verbal memory, executive function, problem solving, visual perception, language, motor coordination, visual-motor integration, and general intellectual functioning. These cognitive domains have been shown to share a bidirectional link with children's academic skills or achievements (Lovden et al., 2020; Peng & Kievit, 2020).

The brain measures we used as predictors were derived from a multimodal MRI combination of *in vivo* sMRI volumes, thickness, curvature, area, and gyrification; DTI fractional anisotropy, axial, radial, and mean diffusivities; as well as ^{1}H-MRS concentrations of choline, creatine, N-acetylaspartate, and other metabolites obtained from a voxel in the midfrontal gray matter region. *In vivo* structural, functional, and diffusion MRI have been demonstrated to be useful metrics for understanding human intelligence or overall cognitive performance in cognitively normal subjects (Vieira et al., 2022).

We further aimed to test the ability of support vector machines/regression, decision tree ensembles, and penalised linear models to predict cognitive performance at the same age as well as 2 years subsequently. ML predictive modelling is a powerful tool that may help overcome the limitations of cognitive testing and multimodal neuroimaging when

each is used alone. In other words, ML algorithms allow interactions and commonalities between multimodal MRI predictors and cognitive scores to be accounted for.

1.3 Research questions and outline of the thesis

In Chapter 2, we present a systematic literature review to answer whether neuroimaging alterations in cohorts of children, adolescents, and youth including individuals living with HIV are associated socioeconomic status and/or with cognitive deficits measured using cognitive testing. We also investigate which cognitive domains may be assessed using neuroimaging, and which neuroimaging modalities or measures are most indicative of cognitive performance.

In Chapter 3, we compare the regression and classification performances of support vector machines, penalised linear models, and decision tree ensembles in predicting cognitive scores of 7-year-old children from low-SES disproportionately affected with HIV using multimodal neuroimaging at the same age. We ask whether multimodal neuroimaging is (a) accurate in predicting cognitive scores and (b) accurate at predicting categories of cognitive performance of children, and (c) if there is a subset of neuroimaging measures that can predict cognitive performance in children.

In Chapter 4, we investigate the possibility of predicting cognitive performance at age 9 using neuroimaging or cognitive data from age 7 of these cohort of children. We ask whether multimodal MRI neuroimaging is better than cognitive scores at predicting subsequent cognitive performance. Finally, we determine whether using current and historic multimodal MRI neuroimaging (at 2 timepoints) improves prediction of cognitive performance.

Lastly, Chapter 5 synthesises and discusses the significance of the overall results, explains the limitations of the study, offers recommendations for future research, then summarises and concludes the work.

2

Associations between neuroimaging measures and cognitive performance in children, adolescents, and youth living with HIV—a systematic review

Chapter 2:

- Are neuroimaging alterations in cohorts of children, adolescents, and youth living with HIV associated with/indicative of cognitive deficits measured using neuropsychological testing?
- Are the brain-cognition relationships influenced by socioeconomic status?
- Which cognitive domains may be assessed using neuroimaging?
- Which neuroimaging modalities or measures are most indicative of cognitive performance?

Abstract

Neurodevelopmental delays and neuroimaging abnormalities continue to be found in children (0–14 years), adolescents (15–18 years), and youth (19–26 years) living with HIV despite the administration of effective combination antiretroviral therapy (cART) in the last two decades. However, there are no clear criteria for the diagnosis of HIV-related cognitive impairment in children and adolescents (0–18 years) as there are for HIV-associated cognitive disorders (HAND) in youth and adults (>18 years). Neuroimaging could be a useful tool for assessing the extent of HIV-related effects on the brain, however, it is not clear how subtle deficits seen on neuroimaging may be related to cognitive performance or impairments in children, adolescents and youth. We therefore aimed to investigate what is known about the relationship between cognitive outcomes assessed by cognitive testing and brain changes measured with neuroimaging.

We searched the electronic databases PubMed/Medline, Scopus, Web of Science (WoS core collection), and via EBSCOhost CINAHL, APA Psych Info & Psych Articles, and

Academic Search Premier for studies published between 1 January 2006 and 18 March 2024, and conducted the review according to the PRISMA 2020 27-item checklist. The protocol is registered in PROSPERO under the number: CRD42021284780. Two reviewers identified studies that investigated a relationship between neuroimaging brain measures and cognitive test scores and included a population of children and/or youth with HIV aged between 0 and 26 years.

After screening 526 full-text articles, we identified 33 eligible studies. The most used neuroimaging modalities were structural magnetic resonance imaging (sMRI) and diffusion tensor imaging (DTI); the cognitive domains widely assessed were attention, executive function, and general intellectual functioning. Meta-analyses were not performed due to considerable heterogeneity in methodology of the studies. Methodological limitations of the studies were cross-sectional analysis (91%), small sample sizes (82% had less than 80 subjects overall), and lack of quantitative neuroimaging assessment in a small subset (9%). In 97% of the studies, the population living with HIV acquired it perinatally. Evidence suggests that perinatal HIV in children, adolescents, and youth may lead to alterations in regional and total brain volumes and diffusivity measures in various white matter tracts, and that these changes are linked to poorer outcomes in short-term memory, processing speed, working memory, problem solving, and general intelligence quotients. While the studies mainly considered sMRI and DTI measures, other modalities such as perfusion MRI and functional MRI remain rarely assessed. Furthermore, these modalities were analysed separately, while multimodal neuroimaging analyses have the potential to enhance the understanding of the neural mechanisms and networks underlying cognitive performance in this population of young people. In conclusion, neuroimaging shows the potential of understanding cognitive performance in children, adolescents, and youth living with and without HIV.

2.1 Introduction

Perinatally-acquired HIV with its neurotropic properties continues to pose a major threat to typical brain development of infected children, adolescents, and youth in the combination antiretroviral therapy (cART) era. This has led to scaled up interventions targeting HIV prevention and care, disclosure, ART adherence, cognitive and academic function, and psychosocial health in sub-Saharan Africa where the burden of HIV is greatest (Muheriwa Matemba et al., 2021; Nalwanga & Musiime, 2022). There is evidence that children and adolescents living with HIV have impairments in selective developmental or cognitive domains of hearing (Ensink & Kuper, 2017), working memory, processing speed and executive function (Phillips et al., 2016; Rowe et al., 2021), as well as learning and semantic fluency (Musindo et al., 2022). These systematic reviews and meta-analyses also show that the extent, severity, and prevalence of cognitive impairment in children and adolescents with HIV may vary depending on the definitions used and the populations studied. Possible reasons are (1) the overreliance on cognitive testing with limited cross-cultural validity in low-resource settings, (2) possible milder forms of HIV-associated neurological impairments in children and adolescents since the introduction of cART, and (3) no HIV-associated cognitive impairment screening and staging criteria for paediatric HIV like the Frascati criteria for diagnosing HIV-associated cognitive disorders (HAND) in adult HIV.

The goal of screening and early diagnosis of HAND is to minimise the progression of asymptomatic neurological impairment (initial infection stage) and mild cognitive disorder (the progression stage) to HIV-associated dementia, the late stage involving severe memory loss and deficits in problem-solving and language—also the point at which there is a lower chance of complete recovery (Vastag et al., 2022). The Frascati criteria, which have been useful toward this goal, require cognitive testing in at least 5 cognitive domains. However, because they have a high false-positive rate for milder forms of HAND, there has been a need to revise these diagnostic criteria to make them more relevant for

diverse global populations (Meyer, 2022). Attempts have been made to validate the criteria in children and adolescents and create a spectrum of HIV-associated cognitive impairment, most notably a study by Hoare and colleagues (2014), which found a false-positive rate in paediatric HIV (ages 0–18 years) comparable to that observed in adults. Surprisingly, children and adolescents with a clinical diagnosis of HIV encephalopathy were found not to have impairment according to the Frascati criteria. Since HIV encephalopathy is an important part of the spectrum of cognitive dysfunction in children and adolescents, this presents a problem with using the criteria in paediatric HIV. Therefore, more investigations are needed on methods and criteria to detect and define the spectrum of cognitive disorders in children and adolescents to better support cognitive interventions in sub-Saharan Africa.

Neuroimaging is one method likely to play a role in this endeavour, beyond its traditional clinical use for identifying cerebral atrophy, basal ganglia calcification, and HIV-associated encephalopathy, as well as excluding secondary central nervous system (CNS) complications such as tumours in young people with HIV (Hoare et al., 2014). Recently, neuroimaging has been utilised extensively in the characterization of HIV neuropathogenesis and in longitudinal assessments, showing that the impact on the CNS of early HIV disease progression persists into adolescence and young adulthood; but developmental trajectories in adolescence seem mostly stable (Nichols, 2022). While various paediatric neuroimaging cohort studies continue to report alterations in cortical and subcortical brain structure (Nwosu et al., 2021; Wade et al., 2019), white matter integrity (Ackermann et al., 2019; Jankiewicz et al., 2017), functional connectivity (Yadav et al., 2018), and neurometabolite concentrations (Mbugua et al., 2016; Robertson et al., 2018) in paediatric HIV, it remains unclear how these changes translate to neurodevelopmental outcomes such as motor speed and coordination, psychomotor integration, language processing, visual-spatial processing, socio-personal abilities, executive function, and general

intellectual functioning as well as how these changes can be used to establish a spectrum of cognitive disorders in children and adolescents.

For neuroimaging to be useful as an auxiliary tool for staging or screening for cognitive deficits in children and adolescents, a clearer link needs to be established between neuroimaging findings and cognitive performance in children. As far as we know, the last narrative and systematic reviews to provide an overview of brain structure and function of children and young adults with perinatally-acquired HIV were Nichols (2022) and van den Hof et al. (2019) respectively. Neuroimaging evidence focused primarily on children and adolescents living with HIV was last reviewed by Musielak and Fine (2016) as an update to Hoare et al. (2014). Hoare and colleagues (2014) also aimed to find associations between neuroimaging and cognitive performance but concluded that we have an incomplete understanding of the relationship between cognitive impairment and HIV-related CNS damage as seen by neuroimaging. There has been substantial research conducted in paediatric HIV since 2014, warranting a review of cognitive deficits in children and adolescents and their associations with neuroimaging findings. We thus decided to examine the existing evidence on cognitive outcomes in children and adolescents living with HIV in the cART era and their association with underlying neuroimaging changes. Notably, the youth age group (19–26 years) sometimes coincides with late adolescence (18–21 years) in paediatric HIV or young adulthood (22–26 years) in adult HIV research. Given that this is a period of continued brain development (Mills et al., 2021), where young individuals are transitioning out of child or adolescent-focused systems into adult-focused ones, we included studies comprising peoples aged between 19 and 26 years.

The purpose of this review is to answer the following questions: (i) Are neuroimaging alterations in children, adolescents, and youth with HIV associated with cognitive deficits measured using cognitive testing? (ii) Are the brain-cognition relationships influenced by socioeconomic status? (iii) Which cognitive domains may be assessed using neuroimaging in this population of young people? (iv) Which neuroimaging modalities

or measures are most indicative of cognitive performance in children, adolescents, and youth living with HIV?

2.2 Material and methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, which consists of an evidence-based minimum checklist of 27 items (Page et al., 2021). This systematic review has been registered in the international prospective register of systematic reviews (PROSPERO) under the registration number: CRD42021284780.

Search strategy

We utilised the electronic databases PubMed/Medline, Scopus, Web of Science, and via EBSCOhost—CINAHL, APA Psych Info, APA Psych Articles, and Academic Search Premier to search for published journal articles. The following words, abbreviations, phrases as text words and MeSH terms were used as an advanced Boolean search string with the appropriate formatting for each database:

["HIV" **or** "human immunodeficiency virus" **or** "AIDS" **or** "acquired immunodeficiency syndrome"] **AND** ["neuroimaging" **or** "brain imaging" **or** "magnetic resonance imaging" **or** "MRI" **or** "CT" **or** "computerized tomography" **or** "PET" **or** "positron emission tomography" **or** "EEG" **or** "electroencephalogram" **or** "DTI" **or** "diffusion tensor imaging" **or** "magnetic resonance spectroscopy" **or** "MRS" **or** "fMRI" **or** "functional MRI" **or** "single-photon emission computerized tomography" **or** "SPECT" **or** "PET-CT" or "TMS" **or** "Transcranial Magnetic Stimulation"] **AND** ["children" **or** "pediatric" **or** "adolescent" **or** "youth"] **AND** ["cognitive" **or** "neurocognitive" **or** "neuropsychological" **or** "neurodevelopmental" **or** "neurobehavioral" **or** "KABC" **or** "PPT" **or** "PPVT" **or** "semantic fluency" **or** "RAKIT" **or** "Beery-VMI" **or** "WISC" **or** "WAIS" **or** "MoCA" **or** "GMDS"] **AND** ["association" **or** "correlation" **or** "link" **or** "relationship" **or** "connection"]

This query indexed studies that explored associations between neuroimaging modalities and cognitive outcomes of children, adolescents, or youth living with HIV and included all the possible combinations of these keywords in the title, abstract, or main text. We

wanted to ensure that the studies were conducted in the cART era in which antiretrovirals were widely accessible especially to developing countries (Forsythe et al., 2019), therefore the date limit was years 2006 to present. The databases were last accessed on 18 March 2024.

Eligibility criteria

The inclusion criteria were as follows: the study included a population of children, adolescents, or youth living with HIV, and there was an investigation of the relationship between neuroimaging brain measures and cognitive test scores. We excluded studies that (1) did not include participants with HIV, (2) were conducted primarily in adults or the sample's mean age was greater than 26 years, (3) reported neuroimaging and cognitive outcomes separately without investigating their associations, and (4) reviews, editorials, conference proceedings, case reports, book chapters, or protocols for prospective trials.

There were no explicit exclusion criteria based on language of publication, whether the children were on antiretroviral treatment, or how the cohort was infected (i.e., vertical or horizontal transmission), text availability online, study design (e.g., cohort, cross-sectional, or longitudinal), or the socioeconomic status of the studied cohort.

Data extraction and quality assessment

We imported all the records into Rayyan (Ouzzani et al., 2016) systematic review web application tool, removed duplicates, and conducted the screening process. Two investigators (ILK & FCR) systematically assessed the titles, abstracts, and full texts of the indexed journal articles against the inclusion and exclusion criteria. Disagreements or ambiguities were resolved in consultation with another investigator (EMM).

For each eligible study, the investigators extracted (a) administrative data: author, year of publication, title, location of study/country of research, cohort name, and study design; (b) sample characteristics: total size, number of participants with HIV and controls, age range/mean age of the sample, HIV transmission route, and cART details; (c)

neuroimaging information: modalities and brain regions of interest; (d) cognitive testing information: cognitive test batteries or tests used and cognitive domains assessed; (e) general methodology: description of inclusion/exclusion criteria, ethical approval, funding sources, declaration of potential conflicts of interest, strengths, and limitations; (f) outcome data: statistical technique implemented, association findings, p-values, strength of correlation r, or regression coefficients β. Additionally, we used the Quadas-2 criteria to evaluate the methodological quality of the studies and identify potential risks of bias.

Synthesis

We did not perform meta-analyses due to the heterogeneity of neuroimaging modalities, cognitive assessment batteries, and types of statistical analyses employed. To prepare the data for a narrative synthesis, we tabulated administrative data and sample characteristics, neuroimaging modalities and cognitive test batteries, the strengths, and limitations of each study as well as quality assessment scores, and associations findings for each study grouped by cognitive domains.

2.3 Results

2.3.1. Eligibility, administrative data, and critical appraisal

Eligible studies identified during systematic search

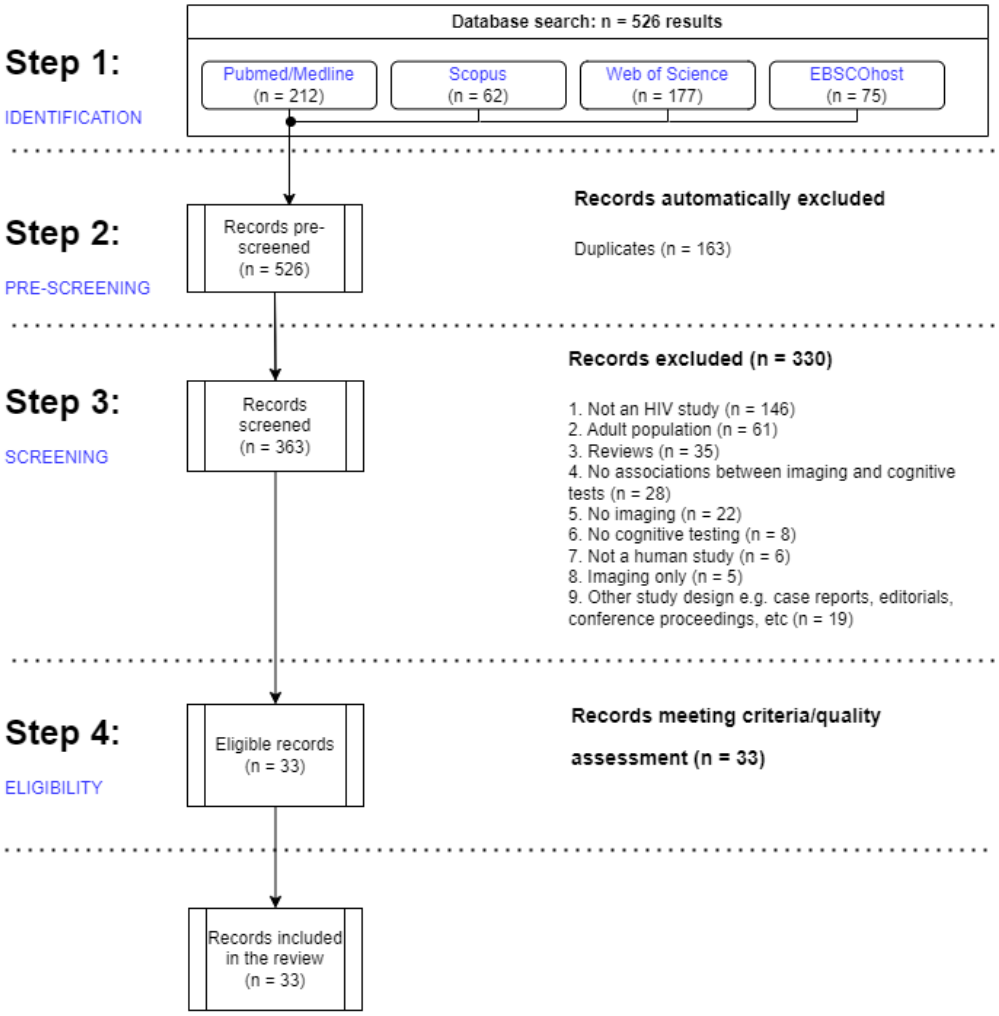


Figure 2.1| PRISMA flow diagram documenting the selection process by which we arrived at the final sample of studies that met the eligibility criteria.

We identified 526 studies during the initial database searches (Figure 2.1) and screened 363 studies after duplicates were automatically removed by the Rayyan software. One hundred and forty-six studies were removed for not studying children, adolescents, and youth living with HIV. An additional 61 studies were rejected because the cohort under study consisted primarily of adults—had a mean age > 26 years. Thirteen studies were excluded due to no cognitive assessment findings being presented. These were predominantly neuroimaging studies, where the associations were with clinical variables. Similarly, we rejected 22 studies investigating only cognitive assessment without neuroimaging.

There were 28 removed studies that had both neuroimaging and cognitive assessment, however no associations between them were reported. Furthermore, we excluded 35 reviews, and 6 studies in non-human subjects. Lastly, we excluded 19 reports with formats that did not meet the inclusion criteria: these included book chapters, conference proceedings, case reports, editorials, and protocols for prospective randomised trials.

At the end of the selection process, thirty-three ($n = 33$) studies were eligible for inclusion (Ackermann et al., 2014; Ackermann et al., 2020; Andronikou et al., 2014; Ashby et al., 2015; Blokhuis et al., 2017; Caceres et al., 2024; Cohen et al., 2016; Dean et al., 2020; Gabis et al., 2006; Heany et al., 2020; Herting et al., 2015; Hoare et al., 2021; Hoare et al., 2015; Hoare et al., 2012; Hoare et al., 2019; Lewis-de los Angeles et al., 2017; Li et al., 2021; Martin Bejarano-Garcia et al., 2021; Nagarajan et al., 2012; Nozyce et al., 2006; R. Paul et al., 2018; Sarma et al., 2021; Schneider et al., 2020; Uban et al., 2015; van Dalen et al., 2016; van den Hof et al., 2021; van Genderen et al., 2022; van Genderen et al., 2021; Wang et al., 2018; Yadav et al., 2017; Yadav et al., 2018; Yadav et al., 2020; Yu et al., 2019).

Administrative data, sample characteristics, and study design

We summarise the administrative data: short details of the studies, sample characteristics, and HIV transmission information of the infected children, adolescents, and youth in each study in Table 1. Ten studies (30%) were conducted in Africa (the majority in South Africa, n = 8, the remainder in Zambia), 8 (24%) in North America (USA), 7 (21%) in Asia, and 8 (24%) in Europe (Netherlands n = 6, England n = 1, Spain = 1). Seventeen studies (52%) described the socioeconomic status (SES) within which the cohorts lived, and it was usually assessed by looking at annual household income and parental education levels. The 17 studies described the cohort in their studies either as residing in low-SES households or their SES being lower than that of the general population. Some of the studies included well-characterised cohorts, namely: the neurological, cognitive, and visual performance in perinatally HIV-infected children cohort (NOVICE, n = 6), Cape Town adolescent antiretroviral cohort (CTAAC, n = 3), children on early antiretroviral therapy (CHER) trial (n = 3), adolescent master protocol (AMP) of the paediatric HIV/AIDS cohort study (PHACS, n = 3), and the paediatric randomized early versus deferred initiation in Cambodia and Thailand cohort (PREDICT, n = 1).

Considering the mean age of the studied participants in each study, there were in total 20 (61%) studies in children (0–14 years), 10 (30%) in adolescents (15–18 years), and 3 (9%) in youth (19–26 years). Therefore, most of the eligible studies (91%) we found were conducted in children and adolescents. For the 3 studies in youth, one study included some participants aged between 26 and 30 years but had an overall mean age of 23 years (Sarma et al., 2021). The other 2 studies by Ashby et al. (2015) and Martin Bejarano-Garcia et al. (2021) had mean ages of 19.0 and 19.6 years, respectively, and included participants in their late adolescence. Ackermann et al. (2014) and Andronikou and colleagues (2014) studied the youngest population, aged less than 4 years.

Many of the studies (91%) were of cross-sectional design, except for 3 (9%) which were longitudinal (Dean et al., 2020; van den Hof et al., 2021; Yu et al., 2019). Most of the

studies (82%) had relatively small sample sizes, i.e., less than 80 subjects overall, the lowest number of subjects being n = 8 (Gabis et al., 2006) and largest n = 374 (Lewis-de los Angeles et al., 2017). However, since 76% of the studies included both children with HIV and controls or children living without HIV, large sample sizes were mostly due to a large sample of the control children. The 8 studies (24%) that included no controls were Nozyce et al. (2006), Gabis et al. (2006), Ackermann et al. (2014), Andronikou et al. (2014), Uban et al. (2015), Herting et al. (2015), Hoare et al. (2015), and Hoare et al. (2019). The largest number of children and adolescents living with HIV studied were by Nozyce et al. (2006) (n = 274), followed by Hoare and colleagues (2019) (n = 125), the other studies had HIV sample sizes less than 65.

The HIV population in 97% of the studies (32 out of 33 articles) acquired HIV perinatally, i.e., via vertical transmission. Nagarajan et al. (2012) had one out of 16 adolescents living with HIV horizontally infected, specified to be via blood transfusion at age 1. Therefore, the studied HIV populations generally lived with HIV from infancy.

Furthermore, with one exception, the sample populations were either all on cART or more than 71% of participants were on cART. One study specifically investigated outcomes of ART-naïve children (Hoare et al., 2012). The study samples had varying treatment regimens, the most common being a combination of thymidine, zidovudine or stavudine with cytidine, and efavirenz or nevirapine. In sixty-seven percent (n = 22) of the studies, the children were on cART or ART, but the drugs/regimen were not specified by the researchers.

Table 2.1| Administrative data of the studies, showing the location where the research was conducted and the study population lives, a short description of the articles, cohort name, study design, age range of the cohort, sample size, and the ART treatment regimen for the HIV sample.

Country /Location	Study	Title/short description	Low SES⁺ community?	Cohort name	Study design#	Age range (mean age*)	Sample size	Vertical HIV transmission	% ART Treated (regimen)
South Africa (Cape Town, Western Cape province)	Hoare et al. (2012)	A diffusion tensor imaging and neurocognitive study of HIV-positive children who are HAART-naïve "slow progressors"	Yes	NOT SPECIFIED	Cross-sectional	8—12 (10.4)	24 (12 HIV, 12 Controls)	Yes	0% (studied ART naïve children)
	Ackermann et al. (2014)	White matter signal abnormalities in children with suspected HIV-related neurologic disease	Yes	CHER	Cross-sectional	0—5 (2.6)	22 (22 HIV, 0 Controls)	Yes	100% (lopinavir-ritonavir, lamivudine, and zidovudine; one child didanosine)
	Andronikou et al. (2014)	Correlating brain volume and callosal thickness with clinical and laboratory indicators of disease severity in children with HIV	Yes	CHER	Cross-sectional	0—4 (2.5)	33 (33 HIV, 0 Controls)	Yes	100% (variable regimens, details not specified)
	Hoare et al. (2015)	Clinical associations of white matter damage in cART-treated HIV-positive children in South Africa	Yes	NOT SPECIFIED	Cross-sectional	6—15 (9.6)	50 (50 HIV, 0 Controls)	Yes	100% (SA standard child regimen: Abacavir, lamivudine, and efavirenz)
	Hoare et al. (2019)	Initiation of antiretroviral therapy after the critical neuronal developmental period	Yes	CTAAC	Cross-sectional	9—11 (10.5)	125 (125 HIV, 0 controls)	Yes	100% (personalised ART programme, drugs not specified)
	Ackermann et al. (2020)	DTI points to ongoing functional impairment in HIV-infected children at age 5, undetectable using standard neurodevelopmental assessments	Yes	CHER	Cross-sectional	5—6 (5.4)	49 (38 HIV, 11 Controls)	Yes	100% (lopinavir-ritonavir, lamivudine, and zidovudine; one child didanosine)
	Heany et al. (2020)	Neural correlates of maintenance working memory, as well as relevant structural qualities	Yes	CTAAC	Cross-sectional	9—13 (10.4)	84 (64 HIV, 20 Controls)	Yes	100% (personalised ART programmes, drugs not specified)
	Hoare et al. (2021)	Alcohol use is associated with mental health problems & brain structural alterations in adolescents with perinatally acquired HIV infection on ART	Yes	CTAAC	Cross-sectional	11—13 (12.5)	78 (52 HIV, 26 Controls)	Yes	100% (personalised ART programmes, drugs not specified)
Zambia (Lusaka)	Dean et al. (2020)	Brain MRI Findings Associated with Cognitive Impairment in Children and Adolescents with Human Immunodeficiency Virus in Zambia	Yes	HANDZ	Longitudinal	8—17 (13.5)	51 (34 HIV, 17 Controls)	Yes	100% (regimens not specified)
	Schneider et al. (2020)	Cerebrovascular disease in children perinatally infected with HIV in Zambia	Yes	HANDZ	Cross-sectional	8—17 (13.5)	51 (34 HIV, 17 Controls)	Yes	100% (regimens not specified)
United States of America (Illinois & California)	Nozyce et al. (2006)	A Behavioral and Cognitive Profile of Clinically Stable HIV-Infected Children	NOT SPECIFIED	PACTG	Cross-sectional	2—17 (7.2)	274 (274 HIV, 0 Controls)	Yes	100% (ARV not cART, either zidovudine or lamivudine)
	Gabis et al. (2006)	Clinical and imaging study of human immunodeficiency virus-1-infected youth receiving highly active antiretroviral therapy	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	10—16 (11.6)	8 (8 HIV, 0 Controls)	Yes	100% (specified only that the regimen consisted of 2 to 3 drugs from more than 2 ART categories)
	Nagarajan et al. (2012)	Neuropsychological function and cerebral metabolites in HIV-infected youth	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	13—25 (17.0)	30 (16 HIV, 14 Controls)	All except 1, infected via blood transfusion at age 1	100% (drugs not specified, 25% before age 1)
	Herting et al. (2015)	Default Mode Connectivity in Youth with Perinatally Acquired HIV	NOT SPECIFIED	AMP	Cross-sectional	11—21 (16.5)	31 (31 HIV, 0 Controls)	Yes	100% (variable, drugs not specified)
	Uban et al. (2015)	White matter microstructure among youth with perinatally acquired HIV is associated with disease severity	NOT SPECIFIED	AMP	Cross-sectional	11—21 (16.5)	31 (31 HIV, 0 Controls)	Yes	100% (variable, drugs not specified)

--Table 2.1 continued--

	Lewis-de los Angeles et al. (2017)	Lower Total and Regional Grey Matter Brain Volumes in Youth with Perinatally-Acquired HIV Infection	NOT SPECIFIED	AMP & PING	Cross-sectional	14—18 (16.7)	374 (40 HIV, 334 Controls)	Yes	92% on cART 5% no ART 3% on ARVs but not cART (drugs not specified)
	Sarma et al. (2021)	White matter of perinatally HIV infected older youths shows low frequency fluctuations that may reflect glial cycling	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	18—30 (22.5)	27 (11 HIV, 16 Controls)	Yes	100% (regimens not specified)
	Caceres et al. (2024)	Relationship between brain structural network integrity and emotional symptoms in youth with perinatally-acquired HIV	NOT SPECIFIED	AMP & PING	Cross-sectional	11—21 (16.7)	254 (40 HIV, 214 Controls)	Yes	100% (variable, drugs not specified)
Spain (Madrid)	Martín-Bejarano et al. (2021)	Brain activity in well-controlled perinatally HIV-infected young adults: a fMRI pilot study	NOT SPECIFIED	CoRISpe	Cross-sectional	16—25 (19.0)	20 (10 HIV, 10 Controls)	Yes	100% (variable, drugs not specified)
Cambodia & Thailand	Paul et al. (2018)	Structural Neuroimaging and Neuropsychologic Signatures of Vertically Acquired HIV	Yes	PREDICT	Cross-sectional	6—17 (11.4)	101 (51 HIV, 50 controls)	Yes	71% on cART > 6 months (zidovudine, lamivudine, nevirapine, efavirenz or lopinavir-ritonavir)
Netherlands (Amsterdam)	van Dalen et al. (2016)	Neurometabolite Alterations Associated with Cognitive Performance in Perinatally HIV-Infected Children	Yes	NOVICE	Cross-sectional	8—18 (13.2)	62 (26 HIV, 36 Controls)	Yes	100% (regimens not specified)
	Cohen et al. (2016)	Cerebral injury in perinatally HIV-infected children compared to matched healthy controls	Yes	NOVICE	Cross-sectional	8—18 (13.8)	72 (35 HIV, 37 Controls)	Yes	90% on cART (regimens not specified)
	Blokhuis et al. (2017)	Higher subcortical and white matter cerebral blood flow in perinatally HIV-infected children	Yes	NOVICE	Cross-sectional	8—18 (13.5)	62 (28 HIV, 34 Controls)	Yes	89% on cART (regimens not specified)
	van den Hof et al. (2021)	Normal structural brain development in adolescents treated for perinatally acquired HIV: a longitudinal imaging study	Yes	NOVICE	Longitudinal	8—18 (13.3–1st MRI, 18–2nd MRI)	43 (20 HIV, 23 Controls)	Yes	100% (regimens not specified)
	van Genderen et al. (2021)	A Longitudinal Analysis of Cerebral Blood Flow in Perinatally HIV Infected Adolescents as Compared to Matched Healthy Controls	Yes	NOVICE	Cross-sectional	8—18 (17.4)	72 (35 HIV, 37 Controls)	Yes	95% (regimens not specified)
	van Genderen et al. (2022)	Brain Differences in Adolescents Living With Perinatally Acquired HIV Compared With Adoption Status Matched Controls	Yes	NOVICE	Cross-sectional	8—18 (14.9)	73 (35 HIV, 38 Controls)	Yes	91% (regimens not specified)
China (Wuhan, Hubei province)	Wang et al. (2018)	Asymptomatic HIV Vertical Transmitted Adolescents' Brain Functional Changes: Based on rs-fMRI	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	12—18 (15.1)	35 (13 HIV, 22 Controls)	Yes	100% (regimens not specified)
	Yu et al. (2019)	Neuroanatomical Changes Underlying Vertical HIV Infection in Adolescents	NOT SPECIFIED	NOT SPECIFIED	Longitudinal	11—17 (13.6)	41 (16 HIV, 25 Controls)	Yes	100% (drugs not specified)
	Li et al. (2021)	Study of Brain Structure in HIV Vertically Infected Adolescents	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	12—18 (15.0)	55 (25 HIV, 33 Controls)	Yes	100% (regimens not specified)
India & Qatar	Yadav et al. (2017)	Altered structural brain changes and neurocognitive performance in pediatric HIV	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	8—15 (10.2)	66 (34 HIV, 32 Controls)	Yes	100% (drugs not specified)
	Yadav et al. (2018)	Changes in resting-state functional brain activity are associated with waning cognitive functions in HIV-infected children	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	6—12 (9.3)	72 (49 HIV, 23 Controls)	Yes	100% (thymidine, zidovudine or stavudine with cytidine, efavirenz or Nevirapine)
	Yadav et al. (2020)	Brain microstructural changes support cognitive deficits in HIV uninfected children born to HIV infected mothers	NOT SPECIFIED	NOT SPECIFIED	Cross-sectional	9—13 (10.0)	48 (22 HIV, 26 controls)	Yes	100% (regimens not specified)

--Table 2.1 continued--

England (London)	Ashby et al. (2015)	Cerebral function in perinatally HIV-infected young adults and their HIV-uninfected sibling controls	NOT SPECIFIED	HYPNet	Cross-sectional	16—25 (19.6)	47 (33 HIV and 14 Controls)	Yes	71% on cART > 6 months (zidovudine, lamivudine, nevirapine, efavirenz or lopinavir-ritonavir)

*Mean age of the participants with HIV. Studies generally included aged-matched controls or participants without HIV.

#Based on the type of analysis performed by the specific study. The studied cohort may have been followed at multiple time points, e.g., CHER is a longitudinal cohort study, but studies included in this table involved a cross-sectional analysis.

+SES—socioeconomic status, quantified by most of the studies using annual household income and maternal educational level

Cohorts:

CHER—Children with HIV Early antiRetroviral randomised trial, CTAAC—Cape Town Adolescent Antiretroviral Cohort, HANDZ—HIV Associated Neurocognitive Disorders in Zambia, PACTG—Pediatric AIDS Clinical Trials Group, AMP—Adolescent Master Protocol, PING—Pediatric Imaging, Neurocognition & Genetics, NOVICE—Neurological, cOgnitive, and VIsual performance in perinatally HIV-infected ChildrEn, PREDICT—Pediatric Randomized Early vs. Deferred Initiation in Cambodia and Thailand, HYPNet - HIV in Young Persons' Network, CoRISpe—Cohort of the Spanish Pediatric HIV Network

Methodological quality of the studies: strengths and limitations

We present strengths and limitations in Table 2.2 as well as the assessment of the risk of bias in Figure 2. The leading limitations that the authors specify are the relatively small sample sizes and the cross-sectional analyses of the data. All studies obtained ethical approval from their institutional regulatory boards. We obtained funding sources and potential conflicts of interest from all but 2 studies (Gabis et al., 2006; Andronikou et al., 2014). Three co-authors for van den Hof et al. (2021) declared potential conflicts, but these were not related to the work presented. We considered the abovementioned factors (limitations, strengths, reporting, sources of bias) along with other Quadas-2 checklist items during the assessment of quality of each study.

One study did not sufficiently detail inclusion/exclusion criteria of the sample population (Wang et al., 2018) and another one (Li et al., 2021) did not describe subscales chosen/cognitive domains assessed with the test. These missing details made their risks of bias, as assessed by Quadas-2 criteria (Figure 2), unclear. One study did not justify the exclusion of left-handed participants in their investigation (Li et al., 2021).

Furthermore, we identified 3 studies (Nozyce et al., 2006; Ackermann et al., 2014; Schneider et al., 2020) that used qualitative, rather than quantitative neuroimaging assessment. The scans in these studies were anonymised and reviewed by one or more neuroradiologists blinded to HIV and cognitive status. The neuroradiologist then identified lesions ('lesion load'), calcifications, and anomalies in the images and provided a diagnosis of cortical atrophy, white matter signal abnormalities, or cerebrovascular disease, respectively. This diagnosis was then associated with cognitive outcomes. Ackermann and colleagues (2014) and Nozyce et al. (2006) used only one paediatric neuroradiologist blinded to the clinical findings at the time of referral, while Schneider et al. (2020) used 2 neuroradiologists. Moreover, there was also one study (Herting et al., 2015) where the cognitive outcome (adaptive functioning) was based on a caregiver report rather than direct assessment.

Overall, there was a potential risk of bias in the studies that used qualitative rather than quantitative neuroimaging or cognitive assessment. We had applicability concerns in the domain of patient selection for the studies that included no controls as these studies did not avoid inappropriate exclusions. Furthermore, there were 2 studies (Wang et al., 2018; Li et al., 2021) were the cognitive assessment battery used for the children is commonly applied in adult populations, so there were applicability concerns in the index test domain for these studies. The remaining studies were judged as having low risks of bias.

Table 2.2| The strengths and limitations, ethical approval, description of inclusion/exclusion criteria, and conflicts of interests for each study.

Country /Location	Study	Limitations	Strengths	Ethical approval/consent	Description of Inclusion/Exclusion criteria	Conflicts of interest & source(s) of funding
South Africa (Cape Town, Western Cape province)	Hoare et al. (2012)	Small sample size, cross-sectional study investigation	Studied relatively obscure cART-naïve "slow progressors" children	Approved by UCT HREC	Yes	None declared by authors; NRF and SAMRC funded
	Ackermann et al. (2014)	Small sample size, qualitative imaging assessment, cross-sectional study, 1.5T magnet	Well characterised cohort with clearly specified treatment regimen	Approved by SU HREC	Yes	None declared; Harry Crossley foundation, NIH grant
	Andronikou et al. (2014)	Small sample size, cross-sectional design, 1.5T magnet	Studied very young children (less than 4 years)	Approved by IRB SAMRC and Wits HREC	Yes	Information not disclosed
	Hoare et al. (2015)	Small sample size, no controls, cross-sectional study	Linear regression analysis used	Approved by UCT HREC	Yes	None declared; SAMRC and NRF funded
	Hoare et al. (2019)	No controls, cross-sectional study	Large sample of HIV participants	Approved by UCT HREC	Yes	None declared; SAMRC, NRF, and NICHD
	Ackermann et al. (2020)	Small sample size, cross-sectional design	Well characterised cohort - followed from birth and ensured adherence, measured personal social quotient	Approved by SU HREC	Yes	None declared; SAMRC, NIAID, NRF grants
	Heany et al. (2020)	Cross-sectional design	Linear mix model analysis, cognitive scores used as covariates in the model	Approved by UCT HREC	Yes	None declared; SAMRC, NRF, and NICHD
	Hoare et al. (2021)	Small sample size, cross-sectional study design	Assessed many aspects of cognition	Approved by UCT HREC	Yes	None declared; SAMRC and NICHD
Zambia (Lusaka)	Dean et al. (2020)	Small sample size, cross-sectional design, 1.5T magnet	Controlled for age and sex, longitudinal analysis	Approved by IRB URMC	Yes	None declared; Funded by URMC, NIH, and McGowan foundation
	Schneider et al. (2020)	Small sample size, qualitative imaging assessment, cross-sectional study, 1.5T magnet	One of the few studies to investigate cerebrovascular disease in perinatal HIV	Approved by IRB URMC	Yes	None declared; funded by NIH, Rochester Center for AIDS Research
United States of America (Illinois & California)	Nozyce et al. (2006)	No controls, qualitative imaging assessment, cross-sectional study	Large sample of HIV participants	Approved by IRB Bronx-Lebanon Hospital Center	Yes	None declared; mainly NIH funded
	Gabis et al. (2006)	No controls, small sample size, cross-sectional study, 1.5 T magnet	Well characterised cohorts, routine MRI follow-ups	Approved by Stony Brook University IRB	Yes	Information not disclosed
	Nagarajan et al. (2012)	Small sample size, cross-sectional study investigation, 3 domains of cognition	Used 2D MRS with localised correlated spectroscopy (L-COSY)	Approved by IRB Harbor-UCLA Medical Center	Yes	None declared; NINDS funded
	Herting et al. (2015)	Small sample size, no controls, cross-sectional study, adaptive functioning based on caregiver report rather than direct assessment	Controlled for age at peak HIV RNA, Used regression analysis	Approved by IRB Harvard School of Public Health	Yes	None declared; NICHD, NIDA
	Uban et al. (2015)	Small sample size, no controls, cross-sectional study design	Controlled for sex, age at peak HIV RNA, and age at cognitive testing; studied youth	Approved by IRB Harvard and NU	Yes	None declared; NICHD, NIDA
	Lewis-de los Angeles et al. (2017)	Only 10 ROIs used in the analysis, cross-sectional study design	Large sample size, linear regression used and adjusted for age and sex	Approved by IRB NU	Yes	None declared; NIH (NICHD, NIDA) funded
	Sarma et al. (2021)	Small sample size, cross-sectional study design	Various cognitive test batteries used to assess all cognitive domains, studied youth	Approved by IRBs Harbor-UCLA Medical Center and UCLA	Yes	None declared; NINDS funded
	Caceres et al. (2024)	Cross-sectional study design	Studied psychosocial cognitive domain	Approved by IRB NU	Yes	None declared; NIH funded
Spain (Madrid)	Martín-Bejarano et al. (2021)	Small sample size, cross-sectional study design	Used the Frascati criteria to assess cognition	Approved by Ethics committees of participating hospitals	Yes	None declared; funding information not available

--Table 2.2 continued--

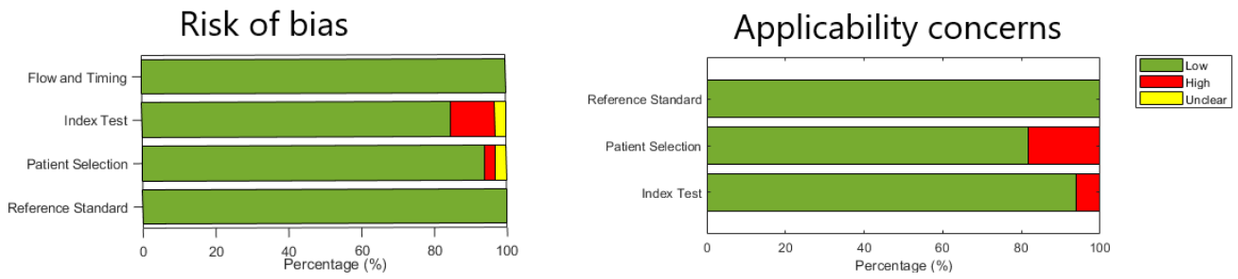
Cambodia & Thailand	Paul et al. (2018)	Cross-sectional study	Large sample size, children enrolled in early treatment	Approved by IRB UMSL	Yes	None declared; mainly funded by NIAID
Netherlands (Amsterdam)	van Dalen et al. (2016)	Small sample size, cross-sectional study	Well characterised cohort	Approved by Amsterdam UMC Ethics committee	Yes	None declared; funding Emma foundation
	Cohen et al. (2016)	Small sample size, cross-sectional study	HIV and controls matched for age, sex, ethnicity, and SES	Approved by UvA Academic Medical Center	Yes	None declared; Emma Foundation funded
	Blokhuis et al. (2017)	Small sample size, cross-sectional study investigation	One of the only studies to use ASL to assess CBF	Approved by Amsterdam UMC Ethics committee	Yes	None declared; funding Emma foundation
	van den Hof et al. (2021)	Small sample size, cross-sectional study	Well characterised cohort, longitudinal analysis, controlled for sex & age	Approved by Amsterdam UMC Ethics committee	Yes	Declared, not related to current work; AIDSfonds funding
	van Genderen et al. (2021)	Small sample size, cross-sectional study	Multiple modalities; well characterised cohort	Approved by UvA Academic Medical Center	Yes	None declared; AIDSfonds funded
	van Genderen et al. (2022)	Small sample size, cross-sectional study	Multiple modalities; well characterised cohort	Approved by UvA Academic Medical Center	Yes	None declared; AIDSfonds funded
China (Wuhan, Hubei province)	Wang et al. (2018)	Small sample size, cross-sectional study, age-inappropriate tests used	Clearly specified confounders	Approved by Medical Ethics committee of Zhongnan Hospital	No	None declared; Government funded
	Yu et al. (2019)	Small sample size	Longitudinal analysis of GM volume and cortical thickness	Approved by Medical Ethics committee of Zhongnan Hospital	Yes	None declared; Government funded
	Li et al. (2021)	Small sample size, cross-section study, right-handed participants only, age-inappropriate tests (domains not described), correlation strength not reported	Controlled for age, sex, and intracranial volume	Approved by Medical Ethics committee of Zhongnan Hospital	Yes	None declared; Declared no funding received for the study
India & Qatar	Yadav et al. (2017)	Small sample size, cross-sectional analysis	Graph-theoretical analysis, age & gender confounding, multiple comparison correction	Approved by IRB Sidra medicine	Yes	None declared; Funded by government's DST
	Yadav et al. (2018)	Small sample size, cross-sectional study investigation	Well characterised sample population	Approved by IRB Sidra medicine	Yes	None declared; Government DST and Sidra Medicine funded
	Yadav et al. (2020)	Small sample size, cross-sectional design	Well characterised cohort - Provided with psychological support, monitoring of nutritional intake, etc.	Approved by IRB Sidra medicine	Yes	None declared; Government DST and Sidra Medicine funded
England (London)	Ashby et al. (2015)	Small sample size, cross-sectional study, 1.5T magnet	Computerised cognitive testing, tests less time consuming	Approved by St Mary's Hospital IRB	Yes	None declared; British HIV association funded, NHS trust

Institutions:

UCT—University of Cape Town, SU—Stellenbosch University, Wits—University of the Witwatersrand, HREC—human research ethics committee, NRF—SA National Research Foundation, DST—Department of Science and Technology, URMC—University of Rochester Medical Center, NIH—National Institutes of Health, SAMRC—South African Medical Research Council, NIAID—NIH National Institute of Allergy and Infectious Diseases, NINDS—NIH National Institute of Neurological Disorders and Stroke, NICHD—NIH National Institute of Child Health and Human Development, UMSL—University of Missouri-St. Louis, NIDA—NIH National Institute on Drug Abuse, NHS—UK National Health Service, Institutional regulatory board (IRB), UvA—University of Amsterdam, NU—Northwestern University, UCLA—University of California, Los Angeles, Amsterdam UMC—Amsterdam University Medical Centers

Abbreviations:

SES—socioeconomic status, ASL—arterial spin labelling, CBF—cerebral blood flow, GM—grey matter.



Study	RISK OF BIAS				APPICABILITY CONCERNS		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Hoare et al. (2012)	+	+	+	+	+	+	+
Ackermann et al. (2014)	+	−	+	+	+	+	+
Andronikou et al. (2014)	+	+	+	+	+	+	+
Hoare et al. (2015)	+	+	+	+	−	+	+
Hoare et al. (2019)	+	+	+	+	−	+	+
Ackermann et al. (2020)	+	+	+	+	+	+	+
Heany et al. (2020)	+	+	+	+	+	+	+
Hoare et al. (2021)	+	+	+	+	+	+	+
Dean et al. (2020)	+	+	+	+	+	+	+
Schneider et al. (2020)	+	−	+	+	+	+	+
Nozyce et al. (2006)	+	−	+	+	−	+	+
Gabis et al. (2006)	+	+	+	+	−	+	+
Nagarajan et al. (2012)	+	+	+	+	+	+	+
Herting et al. (2015)	+	−	+	+	−	+	+
Uban et al. (2015)	+	+	+	+	−	+	+
Lewis-de los Angeles et al. (2017)	+	+	+	+	+	+	+
Sarma et al. (2021)	+	+	+	+	+	+	+
Caceres et al. (2024)	+	+	+	+	+	+	+
Martín-Bejarano et al. (2021)	+	+	+	+	+	+	+
Paul et al. (2018)	+	+	+	+	+	+	+
van Dalen et al. (2016)	+	+	+	+	+	+	+
Cohen et al. (2016)	+	+	+	+	+	+	+
Blokhuis et al. (2017)	+	+	+	+	+	+	+
van den Hof et al. (2021)	+	+	+	+	+	+	+
van Genderen et al. (2021)	+	+	+	+	+	+	+
van Genderen et al. (2022)	+	+	+	+	+	+	+
Wang et al. (2018)	?	+	+	+	+	−	+
Yu et al. (2019)	+	+	+	+	+	+	+
Li et al. (2021)	−	?	+	+	+	−	+
Yadav et al. (2017)	+	+	+	+	+	+	+
Yadav et al. (2018)	+	+	+	+	+	+	+
Yadav et al. (2020)	+	+	+	+	+	+	+
Ashby et al. (2015)	+	+	+	+	+	+	+

Figure 2.2| Methodological quality and risk of bias assessment of each study using the Quadas-2 checklist.

2.3.2. Study methodology: neuroimaging modalities and cognitive domains

Table 2.3 summarises the neuroimaging modalities and measures of interest, cognitive test batteries and cognitive domains of interest, as well as the statistical technique used to find the associations between them.

Neuroimaging

The neuroimaging used in these studies was MRI, except for the oldest study Nozyce et al. (2006) in which 91% of the children were scanned with computerized tomography (CT) and 9% with MRI. It is worth noting this was one of the three studies where neuroimaging was used qualitatively and found no association between neuroimaging and behavioural problems. Seven of these studies (21%) exclusively used diffusion tensor imaging (DTI) to evaluate white matter microstructure in their investigations; 8 (24%) only used structural MRI (sMRI) to quantify regional or total brain volumes, surface area, and cortical thickness. Three studies used single-voxel and 1 used multi-voxel spectroscopy to quantify the concentrations of neurometabolite concentrations, and 1 study exclusively used arterial spin labelling (ASL) to measure cerebral blood flow. Four studies (12%) used resting-state and 2 (6%) task-based functional MRI (fMRI) to measure brain activity and functional connectivity. There were 4 (12%) studies that utilised multiple modalities (Hoare et al., 2021; van den Hof et al., 2021; van Genderen et al., 2022; van Genderen et al., 2021). Hoare et al. (2021) and van den Hof et al. (2021) acquired sMRI and DTI, while van Genderen et al. (2022) and van Genderen et al. (2021) additionally obtained ASL-measured cerebral blow flow (CBF).

Cognitive testing

The studies assessed cognitive outcomes using various cognitive test batteries, the most frequently used test being the Wechsler intelligence scale for children (WISC, n = 16, 48%), followed by Wechsler adult intelligence scale (WAIS, n = 12, 36%) and the revised

Amsterdam child intelligence test (RAKIT, n = 5, 15%). However, since different tests can assess the same cognitive domains, we identified the following broad domains and counted how many studies assessed each of them: (1) attention, processing speed, or short-term memory—n = 28, 85%; (2) executive function, working memory, problem solving, or abstract thinking—n = 21, 64%; (3) visual perception/spatial ability—n = 13, 39%; (4) language, verbal memory, verbal fluency, or verbal IQ—n = 16, 48%; (5) locomotor or motor coordination—n = 10, 30%; (6) psychomotor functioning or visual-motor integration—n = 13, 39%; (7) general intellectual functioning, full scale intelligent quotient (IQ), performance IQ, practical reasoning, or learning—n = 20, 61%; and (8) behavioural problems, or personal-social cognition—n = 4, 12%.

Table 2.3| Imaging techniques, cognitive assessment batteries, and statistical analysis used to find associations

Type of imaging assessment	Study	Imaging technique/ Sequences*	Measures used for associations	Cognitive assessment battery	Cognitive domains	Statistical technique for associations	Were associations found?
Morphometry (sMRI)	Andronikou et al. (2014)	1.5T MRI (T_1, T_2)	Volumes (GM, WM) and cortical thickness	GMDS	Intelligence (FSIQ, VIQ, locomotor IQ)	Pearson's/Spearman's correlation	Yes
	Yadav et al. (2017)	3T MRI (T_1, T_2)	Volumes (GM, WM) and cortical thickness	RAKIT	Visual, short-term memory, processing speed, language, psychomotor	Pearson's correlation, linear regression	Yes
	Lewis-de los Angeles et al. (2017)	3T MRI (T_1)	Volumes (GM)	WISC-IV, WAIS-IV, CPI	Working memory, processing speed, FSIQ	Linear regression	Yes
	Paul et al. (2018)	3T MRI (T_1)	Volumes (GM, WM)	WAIS-III, SBT, CCTT 1&2, Beery-VMI, PPT	VIQ, PIQ, short-term memory, working memory	Pearson's correlation	Yes
	Yu et al. (2019)	3T MRI (T_1, T_2-FLAIR)	Volumes (GM, WM) and cortical thickness	WST, WAIS, WCST, PMT, MpTLS	Executive function, language, attention, visual, motor coordination	Nonparametric correlation	Yes
	Dean et al. (2020)	1.5T MRI (T_1, T_2, and T_2-FLAIR, DTI, MRA)	Volumes (GM, WM)	RAKIT	Attention, language, learning, visual, motor coordination	Linear regression	Yes
	Li et al. (2021)	3T MRI (T_1)	Volumes (GM and WM)	MMSE, MoCA	Attention, executive function, language, visual	Pearson's correlation	Yes
	Caceres et al. (2024)	3T MRI (T_1)	Volumes (GM, WM), Cortical thickness, curvature, LGI	BASC-2	Psychosocial (Emotional Symptoms Index: social stress, anxiety, depression, sense of adequacy)	Linear regression	Yes
Diffusion tensor imaging (DTI)	Hoare et al. (2012)	3T MRI (DTI)	Regional FA, MD, AD, and RD values	WASI, NEPSY-II, GPT, WISC-IV, BNT-SF-SA, RCFT	Intelligence, executive function, attention, motor function, visual	Pearson's correlation	Yes
	Uban et al. (2015)	3T MRI (DTI)	Whole-brain FA, MD, AD, and RD maps	WISC-IV, WAIS-IV	Working memory, processing speed, FSIQ	Linear regression	Yes
	Hoare et al. (2015)	3T MRI (DTI)	Regional FA, MD, AD, and RD values	WISC-IV, CCTT, NEPSY-II	processing speed, executive function, learning, and visual	Linear regression	Yes
	Cohen et al. (2016)	3T MRI (DTI, T_1, T_2-FLAIR)	Whole-brain FA MD, and RD maps	WISC-III, WAIS-III, TMT, RAVLT, Beery-VMI	Intelligence (IQ), processing speed, executive function, psychomotor integration	Linear regression Age-Sex confounders	Yes
	Hoare et al. (2019)	3T MRI (DTI, T_1)	Regional FA, MD, AD, and RD values	WASI, NEPSY-II, GPT, WISC-IV, BNT-SF-SA, RCFT	Intelligence, Executive function, Attention	Pearson's correlation	Yes
	Yadav et al. (2020)	3T MRI (T_1, T_2, DTI)	Regional FA and MD values	RAKIT	Visual, short-term memory, processing speed, language, psychomotor	Pearson's correlation	Yes
	Ackermann et al. (2020)	3T MRI (T_1, DTI)	Whole-brain FA, MD, AD, and RD maps	GMDS, Beery-VMI	locomotor, personal-social, hearing and language, psychomotor, IQ (practical reasoning), processing speed	Spearman's correlation	Yes
Spectroscopy	Gabis et al. (2006)	1.5T MRI (T_1, T_2, T_2-FLAIR, SVS ^{1}H-MRS)	4 metabolite ratios to creatine: N-acetyl aspartate, choline, myo-inositol, lactate	WISC-III, PPT	Intelligence (FSIQ, VIQ, PIQ), language, motor coordination	Pearson's correlation	Yes
	Nagarajan et al. (2012)	3T MRI (T_1, L-COSY)	2D ^{1}H-MRS (20 metabolites)	CPT II, PPT, DKEFS	Processing speed, psychomotor, executive function (problem solving)	Pearson's correlation	Yes
	Ashby et al. (2015)	1.5T sMRI (T_1, T_2, SVS ^{1}H-MRS)	3 metabolite ratios to creatine: N-acetyl aspartate, choline, myo-inositol	CogState, IHDS, PRMQ	Executive function, attention	Pearson's correlation	No
	van Dalen et al. (2016)	3T MRI (T_1, CSI)	CSI MRS—33 metabolites	WAIS-III, WISC, TMT, RAVLT, Beery-VMI	Intelligence, attention, psychomotor function, executive function	Linear regression	Yes

Functional MRI	Herting et al. (2015)	3T MRI (T_1, rs-fMRI)	Regional functional connectivity values	WISC-IV, WAIS-IV	Working memory, processing speed, FSIQ	Linear regression	Yes
	Wang et al. (2018)	3T MRI (T_1, rs-fMRI)	ReHo	MoCA	Attention, executive function, language, visual	Pearson's correlation	No
	Yadav et al. (2018)	3T MRI (T_1, T_2, and T_2-FLAIR rs-fMRI)	Whole-brain ALFF maps; regional functional connectivity showing group differences	RAKIT	Attention, language, learning, visual, motor coordination	Pearson's correlation	Yes
	Heany et al. (2020)	3T MRI (T_1, T_2-star, tb-fMRI)	Whole-brain BOLD signal changes during a maintenance working memory task	WASI, NEPSY-II, GPT, WISC-IV, BNT-SF-SA, RCFT	FSIQ, verbal memory, visual-spatial ability/memory, motor coordination, language, processing speed, and executive function	General linear model	Yes
	Sarma et al. (2021)	3T MRI (T_1, rs-fMRI)	ReHo and ALFF in regions showing group differences	BACS, SFT, TMT, WMS-III, BVMT-R, NAB, MSCEIT, ROCFT, GPT, PSQI, BDI, Stroop, PANSS, WTAR, CPT-IP, HVLT-R, LNS	Executive functioning, motor coordination, social cognition, psychomotor functioning visual, processing speed, abstract thinking	Pearson's correlation	Yes
	Martín-Bejarano et al. (2021)	3T MRI (T_1, tb-fMRI)	Whole-brain BOLD signal changes during a verbal fluency task	K-BIT, WAIS, Luria DNA, TMT, FTT	Attention, processing speed, executive function, language, motor coordination	General linear model	Yes
Radiological assessment (qualitative)	Nozyce et al. (2006)	CT (91% of the children) MRI (9%)	Basal ganglia calcification, WM signal abnormalities, cortical atrophy, focal mass lesions	CPRS, BSID-II, WISC-III, WPPSI-R	VIQ, PIQ, FSIQ, psychomotor function, behavioural problems (learning, psychosomatic, impulsive-hyperactivity, anxiety, hyperactivity)	Wilcoxon exact test	No
	Ackermann et al. (2014)	1.5T MRI (T_1, T_2, T_2-FLAIR, DWI)	WM signal abnormalities, lesion load	GMDS	Intelligence (FSIQ, VIQ, locomotor IQ)	Spearman's correlation	Yes
	Schneider et al. (2020)	1.5T MRI (T_1, T_2, T_2-FLAIR, DTI, MRA)	WM signal abnormalities suggesting cerebrovascular disease	RAKIT	Visual, short-term memory, language (verbal fluency), psychomotor, processing speed	Mann- Whitney U/chi square test/Student's T test	Yes
Cerebral blood flow	Blokhuis et al. (2017)	3T MRI (T_1, ASL, DTI, CSI ^{1}H-MRS)	CBF maps	WISC-III, WAIS-III, TMT, RAVLT, Beery VMI	Intelligence (IQ), processing speed, executive function, psychomotor integration	Linear regression	Yes
Multiple modalities	van den Hof et al. (2021)	3T MRI (DTI, T_1, T_2-FLAIR)	sMRI (WM and GM volumes), DTI (whole-brain FA, MD, AD, and RD maps)	WISC-III, WAIS-III, TMT, RAVLT, Beery VMI	Intelligence (IQ), processing speed, executive function, psychomotor integration	Linear mixed effects	No
	Hoare et al. (2021)	3T MRI (DTI, T_1)	sMRI (GM and WM volumes, cortical thickness, and area) DTI (regional FA, MD, AD, and RD values)	WASI, WISC-IV, CTT2, NEPSY-II Behavioral (BHR, BYI-II, CBCL), BDBSY	FSIQ, processing speed, executive function, learning, visual, socio-personal	Pearson's correlation	Yes
	van Genderen et al. (2021)	3T MRI (T_1, T_2-FLAIR, ASL, DTI)	WM hypointensities volume DTI (regional FA, MD, AD, and RD), CBF maps	WISC-III, WAIS-III, TMT, RAVLT, Beery VMI	Intelligence (IQ), processing speed, executive function, psychomotor integration	Linear mixed effects	Yes
	van Genderen et al. (2022)	3T MRI (T_1, T_2-FLAIR, ASL, DTI)	sMRI (WM and GM volumes), DTI (whole-brain FA, MD, AD, and RD maps), CBF maps	WISC-III, WAIS-III, TMT, RAVLT, Beery VMI	Intelligence (IQ), processing speed, executive function, psychomotor integration	Linear regression	No

*Imaging sequences reported by the authors to have been acquired for the study. Some studies acquired images that were not used in the analysis for associations with cognitive performance.

Imaging technique & sequences:

3T—3 Tesla, T_1 & T_2—T_1-weighted and T_2-weighted contrasts; FLAIR—fluid attenuated inversion recovery, DWI—diffusion weighted imaging, DTI—diffusion tensor imaging, ADC—apparent diffusion coefficient, MRA—magnetic resonance angiography, ASL—arterial spin labelling, CBF—cerebral blood flow, rs-fMRI—resting state functional MRI, tb-fMRI—task-

--Table 2.3 continued--

based functional MRI, ReHo—regional homogeneity, ALFF—amplitude of low frequency fluctuations, MRS—MR Spectroscopy, SVS—single voxel spectroscopy, CSI—Chemical shift imaging, FA—Fractional anisotropy, AD—Axial diffusivity, MD—Mean diffusivity, RD—radial diffusivity, CT—computerised tomography, BOLD - Blood-Oxygen-Level-Dependent

Cognitive test batteries:

RAKIT—"Revisie Amsterdamse Kinder Intelligentie Test" (revised Amsterdam child intelligence test), K-BIT—Kaufman Brief Intelligence Test, WAIS-IV—Wechsler Adult Intelligence Scale 4th edition, TMT—Trail Making Test A, SVT—Semantic verbal fluency, Luria DNA—Luria Neuropsychological Diagnostic for Adults, FTT—Finger tapping test, VIQ—Verbal Intelligence Quotient, PIQ—Performance Intelligence Quotient, FSIQ—Full scale IQ, WAIS-III—Wechsler Adult Intelligence Scale 3rd ed, SBT—Stanford-Binet Test, CCTT 1&2—Child Color Trains Test 1 and 2, PPT—Perdue pegboard test, IHDS—International HIV Dementia scale, WASI—Wechsler Abbreviated Scale of Intelligence, RCFT—REY Complex Figure Test, BNT-SF-SA—Boston naming test-short form-South Africa, CPT II—Conners' Continuous Performance Test II version 5, DKEFS—Delis-Kaplan Executive Function System, WST—Word Semantic Test, VWT—Verbal Working Memory Test, WCST—Wisconsin Card Sorting Test, PMT—Picture Memory Test, MpTLS—Midpoint test of the Line Segment, MoCA—Montreal Cognitive Assessment, BACS—Brief Assessment of Cognition in Schizophrenia, CPT-IP—Continuous Performance Test—Identical Pairs, WMS-III—Wechsler Memory Scale-3rd Ed, HVLT-R—Hopkin's Verbal Learning Test-Revised, LNS—Letter-Number Span, BVMT-R—Brief Visuospatial Memory Test-Revised, NAB—Neuropsychological Assessment Battery, MSCEIT—Mazes, Mayer-Salovey-Caruso Emotional Intelligence Test, ROCFT—Rey–Osterrieth Complex Figure Test, PSQI—Pittsburgh Sleep Quality Index, BDI—Beck Depression Inventory, Stroop—Stroop Color Word Test, PANSS—Positive and Negative Syndrome Scale, WTAR—Wechsler Test of Adult Reading, RAVLT—Rey Auditory Verbal Learning Test, BSID-II—Bayley Scales of Infant Development 2nd ed, WPPSI-R—Wechsler Preschool and Primary Scales of Intelligence-Revised, CPRS—Conners' Parent Rating Scale, BHR—Behavioural health risks, BYI-II—Beck Youth Inventories, CBCL—Child Behaviour Checklist, MMSE—Mini-Mental State Examination, GMDS—Griffith mental development scales, CogState—computerised cognitive testing, BASC-2—Behaviour Assessment System for Children 2nd edition

2.3.3. Synthesis: Associations between neuroimaging and cognitive tests by cognitive domain

A summary of the associations between neuroimaging and cognitive outcomes and their strengths is presented below in Table 2.4 and an illustration of the associations is provided in Figure 2.3.

Attention, processing speed, or short-term memory

The attention domain was associated with volume of the anterior corpus callosum in both HIV and control children (Yadav et al., 2017); better processing speed was correlated with higher grey and white matter volumes (GM and WM; Cohen et al. 2016), as well as bilateral precentral gyrus, left rostral middle frontal, and total GM volumes (Lewis-de los Angeles et al., 2017). The study by van den Hof and colleagues (2021) found that the rate of GM volume decline from 13 ± 3 years to 18 ± 3 years was not associated with cognitive outcomes, including attention, among all the participants. Although van Genderen et al. (2021) and Paul et al. (2018) did not find any associations in the HIV group of attention-related cognitive performance with either GM or WM volumes and cerebral blood flow, poorer short-term memory has been correlated with higher GM choline to creatine ratio (van Dalen et al., 2016). However, increased cerebral blood flow in the basal ganglia was associated with better attention scores of both adolescents with and without HIV in a subsequent study (van Genderen et al., 2022). Moreover, among adolescents living with HIV, stronger temporal correlations of blood-oxygen level-dependent (BOLD) signals, i.e., functional connectivity between the posterior cingulate cortex and left occipital cortex (i.e., default mode network—visual connectivity), and between medial frontal cortex and the right inferior frontal gyrus (i.e., default mode network—executive connectivity), predicted better processing speed scores (Herting et al., 2015). Poorer short-term memory has also been associated with reduced amplitude of low frequency fluctuations in the left middle temporal region and functional connectivity of the left precentral gyrus of the motor network with right precuneus region

(Yadav et al., 2018). Finally, attention scores of children with HIV correlated with mean diffusivity in the right medial frontal region (Yadav et al., 2018), axial diffusivity in the left cerebral peduncle, and fractional anisotropy (FA) in the right superior fronto-occipital fasciculus tracts (Hoare et al., 2019). In contrast, thalamus volume (Paul et al., 2018), FA in left insula (Yadav et al., 2020), WM glutamate to creatine ratio (van Dalen et al. 2016) and ratio of N-acetylaspartate to total creatine (Nagarajan et al., 2012) correlated with attention only in control adolescents, but not in adolescents living with HIV.

Executive function, working memory, problem solving, or abstract thinking

Across all participants, better executive function was associated with increased GM cerebral blood flow (van Genderen et al., 2021), larger volumes of the bilateral cerebellum GM (Yu et al., 2019), left rostral middle frontal gyrus, bilateral precentral gyrus, and total GM (Lewis-de los Angeles et al., 2017), as well as right cerebral white matter amplitude of low frequency fluctuations (Sarma et al, 2021) and FA in the regions of the superior fronto-occipital fasciculus, right superior corona radiata, and corpus callosum (Hoare et al., 2021). In the HIV group, poorer working memory was associated with higher choline to creatine ratio in GM and WM (van Dalen et al., 2016), lower FA in right external capsule and axial diffusivity in left cerebral peduncle (Hoare et al., 2019), and lower cortical thickness of the right rostral middle frontal and right superior frontal gyri; however, while van Genderen et al. (2022) did not find an association between cerebral blood flow and executive function in this group. Better working memory/executive function was associated with higher GM cerebral blood flow (Blokhuis et al., 2017), higher glutamate to creatine ratio (van Dalen et al., 2016), and thalamic volume only in the control children. Comparatively, Ashby and colleagues (2015) reported no associations between metabolite concentrations and scores of attention and executive function among the participants. However, the study did not investigate associations in the controls and HIV group separately, likely due to the small sample size.

Visual perception/spatial ability

In both controls and HIV groups, poor visual perception has been associated with decreased anterior corpus callosum volume (Yadav et al., 2017), reduced FA in the splenium region of the corpus callosum (Hoare et al., 2012), and reduced radial diffusivity in the right temporal superior-longitudinal fasciculus and left putamen region of the inferior-longitudinal fasciculus (Ackermann et al., 2020). Visual perception was positively correlated with mean diffusivity in right cerebellum only in the control group (Yadav et al., 2020). The study by van den Hof and colleagues (2021) found no associations between GM and WM volumes with visual perception/spatial ability.

Language, verbal memory, verbal fluency, or verbal IQ

Four studies have demonstrated an association of verbal ability in children with imaging measures (Yadav et al., 2020; Hoare et al., 2012; Gabis et al., 2006; Ackermann et al., 2014). While Hoare and colleagues (2012) found that poor semantic fluency was associated with decreased FA in the genu and splenium of the corpus callosum, Yadav and colleagues (2020) found that verbal scores correlated with FA in the left postcentral region, right corpus callosum, left insula, and left hypothalamus only in the HIV group; and additionally found that these scores correlated with FA in right fusiform and left pulvinar in the control group only. While Gabis and colleagues (2006) found that verbal intelligent quotient (IQ) in subtests of arithmetic and comprehension was associated with N-acetylaspartate to choline ratio in subcortical white matter and basal ganglia, Ackermann et al. (2014) found no association between verbal IQ and WM lesion load in the children living with HIV. In youth with and without HIV, Bartin-Bejarano and colleagues (2021) found strong correlations between BOLD signals in a cluster in the left inferior frontal gyrus and phonological fluency outcomes.

Motor, locomotor, psychomotor functioning, or visual-motor coordination

Psychomotor function has been associated with amplitude of low frequency fluctuations in the right cerebral WM/right medial orbital gyrus region among youth with and

without HIV (Sarma et al., 2021). In the HIV group, no correlation with metabolite concentrations were found (Nagarajan et al., 2012), rather higher GM myo-inositol to creatine ratio was associated with poorer psychomotor function (van Dalen et al., 2016) in adolescents without HIV. In children living with HIV, it was only motor coordination that was associated with AD in brainstem corticospinal tract (Ackermann et al., 2020). Similarly, lesion load showed no correlation with developmental quotients such as locomotor (Ackermann et al., 2014).

General intellectual functioning, and social cognition

Higher general intellectual functioning (general quotients, IQ) has been associated with increased total brain volumes (Dean et al., 2020) and both GM and WM volumes (Lewis-de los Angeles et al., 2017; Cohen et al., 2016). In the HIV group, poorer scores on processing speed, attention and executive function were associated with decreased AD in the left inferior cerebellar peduncle (Hoare et al., 2015), IQ positively correlated with thickness of the motor corpus callosum (Andronikou et al., 2014), and there was a weak correlation of N-acetylaspartate to choline ratio in frontal subcortical WM with basal ganglia and full-scale IQ (Gabis et al., 2006). In youth, social cognition was associated with regional homogeneity in right central operculum and amplitude of low frequency fluctuations in left cerebral WM (Sarma et al., 2021), and emotional symptoms with global efficiency in salience/midcingulo-insular network (Caceres et al., 2024). However, Wang and colleagues (2018) found no correlation between altered regional homogeneity values in various brain regions and cognitive performance, nor did van Genderen et al (2021) between multiple brain measures and general intellectual functioning in children and adolescents. Furthermore, childhood behavioural problems were associated with the surface area of the right middle temporal and left isthmus cingulate regions (Hoare et al., 2021). However, Ackermann and colleagues (2014) found no correlation between lesion load and developmental quotients, and Nozyce et al. (2006) found no associations between behavioural problems and neuroimaging in children.

Table 2.4| Associations between cognitive performance and neuroimaging measures grouped by cognitive domain.

COGNITIVE PERFORMANCE		ASSOCIATIONS WITH NEUROIMAGING MEASURES		
Domain	Study	All participants	HIV group	Controls
Attention, processing speed, or short-term memory	Nagarajan et al. (2012)	No associations (r & p not reported)	No associations (r & p not reported)	Right frontal lobe NAA/tCr (r=-0.75, p=0.002)
	Herting et al. (2015)	N/A	RSFC from PCC to L occipital cortex (+ve, DMN-Visual connectivity, p = 0.006) RSFC from mPFC to R IFG (+ve, DMN-Executive connectivity p=0.011)	N/A
	Cohen et al. (2016)	Total GM volume (β=+0.06, p=0.020)	NOT REPORTED	NOT REPORTED
	van Dalen et al. (2016)*	NOT REPORTED	GM Cho/Cre (-ve, p=0.047)	WM Glu/Cre (+ve, p=0.013)
	Yadav et al. (2017)	Anterior CC volume (r=+0.49, p=0.003)	NOT REPORTED	NOT REPORTED
	Lewis-de los Angeles et al. (2017)	Bilateral precentral gyrus, LH rostral middle frontal gyrus, and total GM volumes (+ve, p=0.02 to 0.03)	NOT REPORTED	NOT REPORTED
	Paul et al. (2018)	NOT REPORTED	No correlations with volumes (-0.27 < r < 0.24)	Thalamic volume (rho=+0.368, p<0.01)
	Yadav et al. (2018)	ALFF in the left post central region (+ve, p=0.005)	NOT REPORTED	NOT REPORTED
	Hoare et al. (2019)	N/A	AD cerebral peduncle LH (r=+0.17, p=0.02) FA in R superior fronto-occipital fasciculus (r=+0.17, p=0.04)	N/A
	Yadav et al. (2020)	NOT REPORTED	MD in R medial frontal (r=+0.48)	FA in L insula (r=-0.72)
	Martín-Bejarano et al. (2021)	NOT REPORTED	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2021)	CBF in caudate nucleus, putamen, and thalamus (+ve, p = 0.033; 0.036; 0.003 respectively)	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2022)	NOT REPORTED	No associations (r & p not reported)	NOT REPORTED
Executive function, working memory, problem solving, or abstract thinking	van Dalen et al. (2016)*	NOT REPORTED	Cho/Cre in GM (-ve, p=0.035 & p=0.012), Cho/Cre in WM (-ve, p=0.003 & p=0.017)	WM NAA/Cre (+ve, p=0.042) and Glu/Cre (+ve, p=0.008), Glu/Cre in GM (+ve, p=0.006)
	Lewis-de los Angeles et al. (2017)	Bilateral precentral gyrus, LH rostral middle frontal gyrus, and total GM volumes (p=0.004 to 0.05)	NOT REPORTED	NOT REPORTED
	Blokhuis et al. (2017)	NOT REPORTED	No associations (r & p not reported)	GM CBF (r=+0.11, p=0.029)
	Hoare et al. (2019)	N/A	FA in RH external capsule (r=+0.18, p=0.04); LH cerebral peduncle (r=+0.17, p=0.05)	N/A
	Yu et al. (2019)	Bilateral cerebellar GM volume (r=+0.681, p=0.005)	Thicknesses of R rostral middle frontal (r=-0.646, p=0.009) and R superior frontal gyri (r & p not reported)	NOT REPORTED
	Sarma et al. (2021)	R cerebral WM ALFF (-ve, p=0.001) and ReHo (-ve, p=0.006)	NOT REPORTED	NOT REPORTED
	Hoare et al. (2021)*	FA CC's genu (r= +0.0.267 & +0.273, p=0.006 & 0.005) and splenium (r =+0.292 & +0.320, p=0.002 & 0.001), R superior corona radiata (r=+0.266 & +0.273, p=0.006 & 0.005), superior fronto-occipital fasciculus (r=+0.252 & +0.259, p=0.009 & 0.007)	NOT REPORTED	NOT REPORTED
	Martín-Bejarano et al. (2021)	NOT REPORTED	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2021)	GM CBF (+ve, p = 0.045)	NOT REPORTED	NOT REPORTED

	van Genderen et al. (2022)	NOT REPORTED	No associations (r & p not reported)	NOT REPORTED
Visual perception/spatial ability	Hoare et al. (2012)	FA in CC splenium (r=+0.479, p=0.044)	NOT REPORTED	NOT REPORTED
	Yadav et al. (2017)	Anterior CC volume (r=+0.511, p=0.002)	NOT REPORTED	NOT REPORTED
	Ackermann et al. (2020)	RD in right temporal SLF (r=-0.31, p=0.03)	NOT REPORTED	NOT REPORTED
	Yadav et al. (2020)	NOT REPORTED	NOT REPORTED	MD in right cerebellum (r=-0.72)
Language, verbal memory, verbal fluency, or verbal IQ	Gabis et al. (2006)	N/A	NAA/Cho with verbal arithmetic (r=+0.74, p=0.034) and comprehension (r=+0.77, p=0.025) and both (r=+0.65, p=0.08)	N/A
	Hoare et al. (2012)	CC genu (r=+0.471, p=0.048), splenium (r=+0.474, p=0.047), left SLF (r=+0.579, p =0.010)	NOT REPORTED	NOT REPORTED
	Ackermann et al. (2014)	N/A	No associations (p=0.50)	N/A
	Yadav et al. (2020)	NOT REPORTED	FA in L postcentral (r=+0.50), R CC (r=+0.56), L insula (r=+0.53), & L hypothalamus (r=+0.56)	FA in R fusiform (r=-0.74), L pulvinar (r=-0.86)
	Martín-Bejarano et al. (2021)	L IFG cluster activity (r & p not reported)	L IFG cluster activity (r = +0.648, p = 0.043)	NOT REPORTED
Psychomotor functioning, or visual-motor coordination	van Dalen et al. (2016)*	NOT REPORTED	NOT REPORTED	GM Ins/Cre (-ve, p=0.039).
	Sarma et al. (2021)	ALFF RH CWM/MidOG (-ve, p=0.002)	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2021)	GM CBF (+ve, p = 0.043); CBF in caudate nucleus, putamen, and thalamus (+ve, p = 0.023; 0.045; 0.003 respectively)	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2022)	NOT REPORTED	No associations (r & p not reported)	NOT REPORTED
Motor function, locomotor, or motor skills	Ackermann et al. (2014)	NOT REPORTED	No correlations with lesion load (p = 0.80)	NOT REPORTED
	Ackermann et al. (2020)	NOT REPORTED	AD in the brainstem in the CST (r=-0.33, p=0.05)	NOT REPORTED
	Martín-Bejarano et al. (2021)	NOT REPORTED	NOT REPORTED	NOT REPORTED
General intellectual functioning, full scale IQ, social cognition, overall cognitive performance	Gabis et al. (2006)	N/A	NAA/Cho (r=+0.54, p=0.54)	N/A
	Ackermann et al. (2014)	NOT REPORTED	No correlation with lesion load (p = 0.99)	NOT REPORTED
	Andronikou et al. (2014)	N/A	motor CC maximum (+ve, p = 0.0277)	N/A
	Hoare et al. (2015)	N/A	AD in the left inferior cerebellar peduncle (β=+0.00002, p=0.037)	N/A
	Cohen et al. (2016)	GM (β=+0.07, p=0.009) and WM (β=+0.11, p=0.002) volumes; WM MD (β=-1.63, p=0.038)	WM volume (β=+0.11, p=0.029)	NOT REPORTED
	Lewis-de los Angeles et al. (2017)	Bilateral precentral gyrus, L rostral middle frontal gyrus, and total GM volumes (p=0.01 to 0.05)	NOT REPORTED	NOT REPORTED
	Wang et al. (2018)	NOT REPORTED	No associations (r & p not reported)	No associations (r & p not reported)
	Dean et al. (2020)	Total brain volume (β=25.9, p=0.04)	NOT REPORTED	NOT REPORTED
	Sarma et al. (2021)	ReHo in RH central operculum (p<0.001); ALFF in L CWM/MidOG (-ve, p<0.0009), L CWM (-ve, p<0.002), ReHO in R CWM (-ve, p<0.0009) and ReHo in R central operculum (p=0.002)	NOT REPORTED	NOT REPORTED
	Hoare et al. (2021)*	Surface area of L isthmus cingulate (r=-0.238, p=0.008); surface area of R middle temporal region (r=+0.253, p=0.005)	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2021)	NOT REPORTED	NOT REPORTED	NOT REPORTED
	van Genderen et al. (2022)	NOT REPORTED	No associations (r & p not reported)	NOT REPORTED
	Caceres et al. (2024)	NOT REPORTED	Global efficiency in salient network (β=+205.1, p=0.03)	NOT REPORTED

--Table 2.4 continued--

*van Dalen et al. (2016) and Hoare et al. (2021) defined working memory as an instance of attention and had a separate executive function score from a different NP test; we report both outcomes in the broader domain of executive function, and their processing speed/short-term memory outcomes as instances of attention in line with other studies.

Abbreviations:

BOLD—blood-oxygen-level-dependent, NAA—N-acetylaspartate, Cr/tCr—creatine/total creatine, Glu—glutamate, Cho—choline, PCC—posterior cingulate cortex, DMN—default mode network, mPFC—medial prefrontal cortex, R/L IFG- right/left inferior frontal gyrus, ALFF RH CWM/MidOG—amplitude of low frequency fluctuations in right hemisphere cerebral white matter/medial orbital gyrus region, , ReHo—regional homogeneity, L/R CWM—left/right cerebral WM, CC—corpus callosum, AD—axial diffusivity, MD—mean diffusivity, RD—radial diffusivity, FA—fractional anisotropy, CST—corticospinal tract, SLF—superior longitudinal fasciculus, -ve/+ve—negative/positive associations

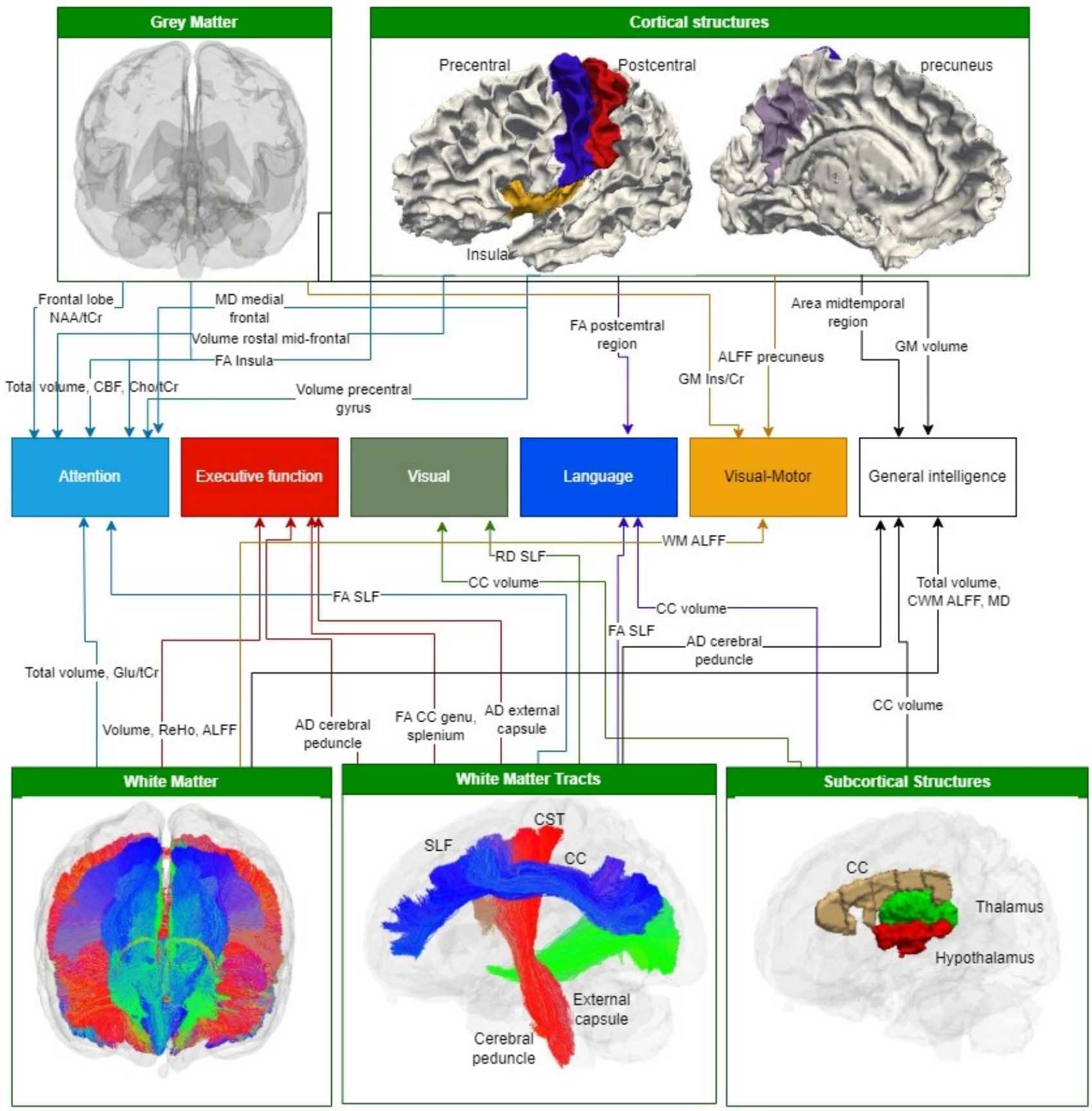


Figure 2.3| Illustration of associations between various brain measures and the cognitive domains presented in Table 2.4.

2.4 Discussion

Given that there are relatively few studies exploring relationships between neuroimaging measures and cognitive outcomes of children, adolescents, and youth living with HIV, caution should be exercised when interpreting and generalising these findings.

HIV research in children, adolescents, and youth

Most of the studies relating neuroimaging and cognitive ability were in children (61%) and adolescents (30%). We did not find studies with a primary focus on young people aged between 19 and 26 years. However, two studies (Ashby et al., 2015; Sarma et al., 2021) had a sample population whose mean age coincided with late adolescence (age range 15–25 years) and another's (Martin-Bejarano et al., 2021) coincided with young adulthood (18–30 years). One study in youth with and without HIV found no relation between neurometabolite levels and executive function or attention (Ashby et al., 2015), while the other found relations between functional connectivity of cerebral white matter, regional homogeneity of operculum and general intellectual ability (Sarma et al., 2021), and Martin-Bejarano et al. (2021) found emotional symptoms were associated with BOLD activation in the inferior gyrus. This suggests there is still more research needed to understand how outcomes in youth may differ from those in childhood and adolescence. Although there are notable developmental changes in all three age groups, the variables influencing these changes can differ substantially. Knowing these developmental trajectories can help us better understand how people grow and change, and the impact of HIV infection during these important life stages. The discussion that follows is therefore mostly applicable to persons living with and without HIV in the childhood and adolescent periods.

Neuroimaging techniques in paediatric HIV research

Most HIV neuroimaging studies we found analysed MRI data, although computed tomography (CT) is more accessible, particularly in resource-limited settings. A similar

result was found by Musielak and Fine (2016) and van den Hof et al. (2019) in their systematic reviews of neuroimaging studies of children and adolescents with perinatally-acquired HIV on long-term treatment. Possible reasons are MRI not using ionizing radiation and being more sensitive and specific for detecting HIV-related brain changes. Additionally, one study (Nozyce et al., 2006) included children scanned with both CT and MRI, but somewhat surprisingly, no associations were found between behavioural problems, or lower developmental and cognitive performance with WM signal abnormalities, basal ganglia calcification, or lesion load. This may be because of the low proportion of children in the study with behavioural or developmental problems.

The common MRI modalities used were structural MRI and diffusion tensor imaging (sMRI & DTI), which quantified volumes, cortical thickness, and microstructural properties in white matter tracts. This bias toward sMRI and DTI was found by Musielak and Fine (2016) in the most recent review conducted on neuroimaging in perinatal HIV. Four studies we found examined both sMRI and DTI, and two of these additionally included arterial spin labelling (ASL) measured cerebral blood flow. Magnetic resonance angiography was acquired by 3 studies; however, no associations were reported between MRA and cognitive performance. More studies are needed investigating MRA and other MRI modalities such as spectroscopic and functional MRI (MRS & fMRI). Furthermore, in the studies that used multiple neuroimaging modalities, analyses were performed on each modality separately. There is therefore scope for future research using multimodal and/or multivariate analyses to identify broad relationships, linear or non-linear, between neuroimaging measures and cognitive performance.

Cognitive testing assessments in paediatric HIV research

The commonly assessed cognitive domains were attention (processing speed and short-term memory), executive function (working memory, abstract thinking) and general intellectual functioning (overall intelligence quotients). Deficits in executive function (Rowe et al., 2021), attention (Phillips et al., 2016), and learning (Musindo et al., 2022)

are some of the widely reported domains impacted in children and adolescents living with HIV. This explains the numerous attempts by studies to link these domains to neuroimaging. Other less studied domains were hearing and semantic fluency, which are also implicated in cognitive dysfunction due to perinatal HIV (Ensink & Kuper, 2017; Musindo et al., 2022). Many of the studies assessed other cognitive domains, e.g., visual/spatial ability and psychomotor functioning, but most failed to test for socio-personal performance outcomes such as disruptive behaviour, anger, depression, self-concept, and apathy, which carry more implications for practice and policy related to paediatric HIV in sub-Saharan Africa (Cockcroft & Cassimjee, 2020; Gabbidon et al., 2020; Tuke et al., 2020).

Heterogeneity in study methodology

There was considerable heterogeneity in the studies reviewed due to either clinical or methodological diversity, which was expected because of the broad scope that neuroimaging and cognitive testing cover (Thompson & Jahanshad, 2015). However, heterogeneity makes it challenging to extract commonalities between findings from different studies. For this reason, we did not explicitly perform meta-analyses in this systematic review. Instead, the wide range of associations between brain measures and cognitive outcomes found by the different studies were synthesised narratively. Diverse neuroimaging modalities or cognitive test batteries could lead to divergent results on the same disease, resulting in limited reproducibility of the findings from different studies (Darby et al., 2019). There has been extensive research on challenges and solutions to reproducibility in neuroimaging analysis (Botvinik-Nezer & Wager, 2023). For example, a possible solution to identifying a finding consistently across different methods is to ensure that factors contributing to variability in the results are minimised, these include data smoothness, the analysis software used, and using parametric versus nonparametric statistical tests (ibid.) Similarly, for replicability between datasets, studies should be statistically well-powered. However, even if diverse neuroimaging studies do not converge

on a particular brain region that is associated with a disease or cognitive function, it is possible that the regions identified in different studies may instead be localised to a common brain network (Darby et al., 2019).

Quality assessment of the studies and certainty of evidence

We independently assessed the quality of each study to rate the certainty of their evidence based on their limitations and risk of bias. Our assessment criterion included their full text, since partial screening, especially that conducted by single reviewers, has been found to be problematic (Gartlehner et al., 2020; Stoll et al., 2019; Waffenschmidt et al., 2019). The study populations mostly included young people with HIV and age-matched controls, many under the age of 18. The children with HIV had varied treatment regimens, most of which were not detailed by the authors. Therefore, differences in outcomes across studies may also be related to the type of cART the children were receiving or the stage of the disease. Sufficient reporting of eligibility criteria, age of cART initiation and regimen, interventions/tests, and statistical analyses is important to ensure sound methodology. Only a few failed to detail the exclusion criteria of the participants. Additionally, some of the control groups had uninfected children born to mothers living with HIV, in whom cognitive and neuroimaging changes have also been reported (Tran et al., 2016; Wedderburn et al., 2020). Furthermore, we noted a subset of studies that assessed neuroimaging qualitatively. It is important in these cases for the scans to be anonymised and reviewed by more than one neuroradiologist blinded to HIV and cognitive status, and for the neuroradiologists to follow the revised RECIST guidelines (Eisenhauer et al., 2009; Schwartz et al., 2016). Another limitation was the cross-sectional analyses employed by most of the studies, although for 3 studies (Ackermann et al., 2014; Ackermann et al., 2020; Andronikou et al., 2014) longitudinal measures were available for the cohort. However, none of these methodological factors had an immense impact on the quality appraisal of the studies and general conclusions of the review.

Reporting of findings across participants living with and without HIV

It is notable that 76% of the eligible studies in this review included uninfected controls. Therefore, investigators had the opportunity to look for associations between neuroimaging and cognitive scores across three different groups: the children living with HIV, controls, and all participants. However, there was no consistency in the reporting of findings. Either the studies provided results for associations across all subjects including controls or only in the sample with HIV. Two studies presented results in the control group only. Where associations in the other groups were not reported, it is unclear whether this was because none were found, or the researchers did not look for them. For example, there were 6 instances where associations between neuroimaging measures and cognitive scores were reported in the control group but not in the group with HIV. This may be weak evidence for perinatal HIV affecting the typical relationship between brain structure and performance measures or simply a lack of statistical power. It was unfortunate that studies reporting on associations in the HIV group did not report or analyse the relationship in controls or across all subjects in their sample. This could have provided support for the hypothesis that HIV-related alterations affect the typical brain structure and performance of the studied population. It is important that negative findings in research are reported because they provide a more complete understanding and allow more informed conclusions to be made.

Associations between neuroimaging and cognitive performance

Overall, the studies found associations of regional and total volumes and WM microstructural measures with performance in tests of attention, executive function, and learning ability. This may be in part due to sMRI, DTI, and these cognitive domains being overrepresented in these studies. Due to the challenge of keeping young children still in the scanner for long periods, good quality sMRI and DTI data are more frequently obtained compared to fMRI, MRA and MRS, possibly leading to their overrepresentation in neuroimaging investigations. Similarly, there may be stronger a priori evidence of HIV

impacting the domains of attention and executive function than personal-social cognition in young children (Ensink & Kuper, 2017; Musindo et al., 2022), resulting in the latter being the least commonly assessed.

The corpus callosum (CC) was a structure mostly associated with cognitive outcomes of children and adolescents living with and without HIV. CC volume was associated with outcomes of attention, visual perception, and language, and microstructure of the CC genu and splenium with visual, motor, and executive function in these young people. Among the 7 studies that exclusively used DTI, measures reflecting better white matter integrity and well-organised structure, i.e., higher fractional anisotropy (FA) or axial diffusivity and lower mean or radial diffusivities, were found across different regions and tracts. Yadav et al. (2020), albeit using a small sample size, surprisingly reported better scores of attention being associated with higher mean diffusivity in the HIV group and lower FA in the control group; lower FA was also found to be associated with better language scores in the control group. This underscores the need for more large-cohort studies linking DTI metrics with cognition. Moreover, the relationship between neurometabolite concentrations and cognition is still unclear from the four spectroscopic investigations we reviewed, but consistent results were found. For example: higher concentration of N-acetylaspartate (typically considered a measure of better myelin maintenance, higher neuronal density and viability, and lower axonal damage) to choline ratio were associated with better scores of general intellectual functioning and language in the HIV group, and higher N-acetylaspartate to creatine was linked to better scores of attention and executive function in controls. Similarly, the relationship between functional connectivity and cognition in children and adolescents also remains poorly understood, as the results of the 6 studies included in this review were heterogeneous in measures of connectivity found and in regions implicated.

Although it was common for grey and white matter measures to be associated with outcomes of attention, executive function, and general intellectual functioning, the

studies differed widely in which measures and regions were linked. For example, while some studies considered total volumes of grey and white matter, other studies linked measures of regional homogeneity, cerebral blood flow, FA, metabolite concentrations, and so on, to cognitive outcomes. Other structures, for example, white matter tracts, cortical and subcortical regions, varied from one study to the next. Therefore, no definitive conclusions can be drawn from the current evidence, and more high-quality studies are needed to elucidate the links between brain measures and cognitive outcomes of children and adolescents living with HIV.

Socioeconomic status background and the brain-cognition relationship

Approximately half of the studies specified the cohort to reside in low-SES households; a criterion primarily based on low-household income and educational level of the parent, typically the mother. However, this review has found it challenging to determine the role that SES played in the associations between HIV, brain structure, and cognitive performance. None of the studies performed analyses with SES or discussed how it might have influenced their results. Furthermore, not all studies presented the data used to determine SES described in their methodology.

Recommendations for future work

Future studies should explore using modalities such as perfusion MRI, resting-state and task-based fMRI, sMRI that includes brain measures incorporating gyrification indices, multimodal combinations of MRI techniques, clinical indicators of HIV disease severity, and demographic confounders. Additionally, studies using predictive ability of single modality or multimodal MRI to determine future cognitive outcomes may provide further insight into longitudinal exposure to HIV or socioeconomic-status factors. Understanding the underlying key predictors of cognitive outcomes is of vital importance for improving learning and rehabilitation of this cohort of children and adolescents.

In conclusion, there is a lack of multimodal analysis of MRI data in populations including children and adolescents living with HIV. There is also limited research exploring predictive modelling of future cognitive performance of this cohort of young people.

3

Predicting child cognitive performance from multimodal magnetic resonance neuroimaging—comparing classification and regression performances of penalised linear models, support vector machines, and decision tree ensembles

Chapter 3:

- What is the regression and classification performances of penalised linear models (PLMs), support vector machines (SVMs), and decision tree ensembles (DTEs) in predicting cognitive scores?
- Is multimodal MRI accurate in predicting cognitive scores or categories of overall cognitive performance in children?
- Is there a subset of neuroimaging measures that can predict cognitive performance in children?

Abstract

Children with cognitive difficulties may struggle in school and other areas of life, and early identification and intervention is crucial for their development. The ability to predict cognitive performance from neuroimaging or other measures could make it easier to identify at-risk children who could most benefit from interventions. We aimed to compare the performance of penalised linear models (PLMs), support vector machines/regression (SVMs/R), and decision tree ensembles (DTEs) in predicting individual continuous scores as well as categories of overall cognitive performance in children from a multimodal neuroimaging combination of structural magnetic resonance imaging (sMRI), diffusion tensor imaging (DTI), and proton magnetic resonance spectroscopy (^{1}H-MRS). Our study population comprised 132 7-year-old children from low-socioeconomic status households and communities. Better and poorer overall cognitive performance categories

were constructed with a hierarchical clustering algorithm. The prediction models were implemented with Bayesian optimization and assessed with 10-fold cross validation. Predicting individual cognitive scores led to small training and fitting errors but high generalised errors, with the linear SVM algorithm producing smaller training and validation errors compared to PLMs and DTEs. The classification models of better versus poorer performing children resulted in area under the receiver operating characteristics curve (AUC) of 0.71–0.83, 65–74% accuracies, 70–85% sensitivities, and 65–73% specificities, with linear SVMs being the best performing algorithms and PLMs having the worst performance. Using SVM-recursive feature elimination, we identified 40 relevant predictors of overall cognitive performance out of 489 candidate neuroimaging measures, which included cortical and subcortical volumes and diffusivity in select white matter tracts. More research using multimodal neuroimaging and cognitive assessment is required with larger sample sizes to improve prediction and reliability, especially the prediction of future cognitive outcomes from neuroimaging.

3.1 Introduction

The ability to predict cognitive outcomes, such as measures of intelligence, working memory, and attention from neuroimaging data has the potential to inform the development of new diagnostic methods and interventions for children with cognitive difficulties. Yet, building models with high reliability and prediction performance remains a challenging task (Callaway, 2022; Kong et al., 2022; Sui et al., 2020). Several machine learning algorithms have been applied to neuroimaging data to predict childhood cognitive abilities with reasonable performance, including penalised linear models (PLMs) (Pan et al., 2018; Yao et al., 2023), support vector machines/regression (SVMs/SVR) (Oztekin et al., 2021), and decision tree ensembles (DTEs) (Gaus et al., 2023).

PLMs overcome the challenges faced by standard linear models such as collinearity of the predictors, overfitting of the prediction model, and a non-constant error variance when dealing with multivariate datasets containing many predictors and few observations (p>>n, curse of dimensionality). Penalisation introduces constraints to the model weights, allowing the less contributing predictors to have coefficients close or equal to zero (Hastie et al., 2009). PLMs with the elastic net or least absolute shrinkage and selection operator (lasso) penalties have been shown to improve overall prediction performance (Greenwood et al., 2020). In a recent application to neuroimaging, Yao et al. (2023) could use sulcal depth in left prefrontal cortex and lasso regression to predict verbal working memory of children and adolescents achieving correlation strengths $0.25 < r < 0.40$ between actual and predicted values. Pan et al. (2018) also managed to use lasso regression on sociodemographic, neuroanatomical, and genetic predictors to predict 21% of variance (coefficient of determination $R^2 = 0.21$) in cognitive flexibility scores of children.

Theorised by Cortes and Vapnik (1995), SVM (also SVR) is a different kind of supervised learning algorithm used for continuous value regression or category classification. In the case of classification, the algorithm seeks to identify a hyperplane with the largest margin

(defined as the distance between the nearest data points from each category and the hyperplane) that maximally separates the various categories in the training data. New data points can be categorised based on which side of the hyperplane they fall once it has been determined. Kernel SVM can be implemented when there is no discernible margin of separation in the data or when data points are not linearly separable (Scholkopf & Smola, 2018; Schölkopf et al., 2002). An example of an application of the algorithm is the study by Oztekin et al. (2021) who classified cognitive deficits of children associated with attention deficit hyperactivity disorder (ADHD) from cortical anatomy and cognitive measures of executive function with 61–94% accuracies.

The decision tree paradigm, on the other hand, follows a tree-like structure to understand how a classification, ranking, or a regression problem can be solved. Generally, trees are constructed via an algorithmic approach that identifies ways to split a dataset based on different conditions, typically information gain or mutual information (Costa & Pedreira, 2022). Often a single tree is not accurate for high-dimensional neuroimaging problems (n<<p), as such, ensemble learning methods combine several decision trees to produce better predictive performance (Deshmukh et al., 2020; Rokach, 2019; Zhou, 2012). Gaus and colleagues (2023) using boosted trees on regional volumes and thickness measures could marginally predict psychiatric disorders like ADHD (area under the receiver operating characteristic curve, AUC = 0.57, $p = 0.002$) and bipolar disorder (AUC = 0.55, $p = 0.002$) in children.

Studies linking brain measures to cognitive performance tend to use a single modality or small number of regions. For example, Yao et al. (2023), Paul et al. (2018), Oztekin et al. (2021), and Gaus et al. (2023) used structural MRI (sMRI) measures. Multimodal neuroimaging on the other hand, combines multiple measures and may provide a more comprehensive examination of paediatric brain structure and function as well as offer a better predictive performance of cognitive abilities of children than using one modality alone (Khobo et al., 2022; Rasero et al., 2021).

While each of these algorithms has been applied to neuroimaging data to predict cognitive outcomes, there is currently a lack of comparison studies that investigate the performance of different algorithms on the same dataset to achieve both regression and classification of cognitive outcomes. To our knowledge, there are two studies (Pat et al., 2022; Yuan et al., 2022) comparing machine learning approaches to link brain measures to cognitive performance of children living without any neurodevelopmental disorder. Yuan and colleagues (2022) reported comparable predictive performance for the algorithms PLMs (ridge, lasso, and elastic net), SVM, and DTEs (random forest and gradient boosted trees) in the regression of inhibitory control, a component of executive function, from task-based functional MRI (fMRI). The second study by Pat and colleagues (2022) compared similar models to Yuan et al. (2022) but included kernel SVMs and univariate algorithms like ordinary least squares (OLS) in the prediction of response inhibition measured via task-based fMRI stop-signal reaction time. Since the investigators primarily focused on the cognitive domain of executive function, and a single modality in the case of Pat et al. (2022), it is still unclear how these algorithms would perform on other neuroimaging measures, cognitive domains, or measures of overall cognitive performance. Moreover, it is uncertain how the predictive performance would change for classification of the same scores because only regression of the cognitive scores was considered by these 2 studies.

Therefore, the goal of this study was to compare the performance of PLMs, SVMs, and DTEs in predicting cognitive outcomes in children from multimodal neuroimaging data. Specifically, we aimed to utilize a combination of sMRI, DTI, and proton magnetic resonance spectroscopy (^{1}H-MRS) and apply these three families of algorithms (1) to regress continuous cognitive scores and (2) classify categories of overall cognitive performance. Additionally, beyond enhancing reliability and prediction of cognition from multimodal neuroimaging and predictive modelling, we sought to find multimodal predictors or brain measures of cognitive performance using feature selection in conjunction with these

algorithms. Examining feature importance in the prediction model allows us to know which brain measures contribute the most to cognitive abilities of children.

Among the various machine learning algorithms, Pat et al. (2022) found the elastic net performed similar to or better than nonlinear and interactive algorithms, and Yuan et al. (2022) found that while the elastic net algorithm yielded the most accurate predictions, SVMs and ridge regression generated comparable results. Therefore, we hypothesised that (1) PLMs would outperform SVMs and DTEs in predicting cognitive scores in our cohort, (2) the prediction of categories of cognitive performance would be more accurate than that of individual scores, and (3) we would find a subset of multimodal neuroimaging measures predictive of cognitive performance in children.

3.2 Material and Methods

3.2.1. Participants

The study population comprised 132 children aged 7 ($7.01 - 7.70\ years, mean\ age \pm sd = 7.22\ \pm 0.13\ years$) from low-socioeconomic status (SES) backgrounds in Cape Town, South Africa (demographics: supplementary Table B.1). HIV is a common condition that disproportionately affects low-SES communities (Wabiri & Taffa, 2013) and in urban South Africa, is concentrated amongst the poorer SES quintiles (Cleary et al., 2011). A subset of 69 participants were asymptomatic, virally-suppressed children perinatally-infected with HIV from the Children with HIV Early Antiretroviral Therapy (CHER) trial conducted between 2005 and 2011 to determine when to start combination antiretroviral therapy treatment (Cotton et al., 2013; Laughton et al., 2012; Violari et al., 2008). The rest of the participants were 63 age and community-matched uninfected children, including those from a related vaccine trial (Madhi et al., 2010) and children subsequently recruited for a neurodevelopmental sub-study (Laughton et al., 2018).

In this cohort the classification of children living with and without HIV using multimodal neuroimaging has been previously reported (Khobo et al., 2022). Similarly, group-wise comparisons of spectroscopic measures (Holmes et al., 2017; Mbugua et al., 2016; Robertson et al., 2018), DTI (Ackermann et al., 2016; Jankiewicz et al., 2017; Madzime et al., 2022), resting-state fMRI (Toich et al., 2018), sMRI (Nwosu et al., 2021; Nwosu et al., 2018; Randall et al., 2017), and cognitive scores (Laughton et al., 2013; Laughton et al., 2018; van Wyhe et al., 2021) have previously been presented for this population. This was a well-characterised cohort with three-monthly adherence visits for those living with HIV and six-monthly clinical assessments for children without HIV. All study procedures were performed according to protocols approved by the Human Research Ethics Committee of the University of Cape Town (HREC ref: 453/2021). Parents or guardians provided written consent, and oral assent was given by the children who were first familiarised with the procedures with practice cognitive tests and a mock MRI scanner.

3.2.2. Cognitive assessment

Trained research assistants supervised by a licensed psychologist administered six cognitive test batteries as described by van Wyhe and colleagues (2021):

- Kaufmann assessment battery for children – second edition (KABC-II) with 4 scores, viz., learning (short-term memory), planning (executive reasoning), simultaneous processing (visuospatial processing and problem solving), and sequential processing (auditory working memory) (Kaufman & Kaufman, 2014);
- Beery-Buktenica developmental test of visual-motor integrations (Berry-VMI) with one score, a non-verbal assessment for the extent to which individuals can integrate their visual and motor abilities (Beery & Beery, 2004);
- Peabody picture vocabulary test – fourth edition (PPVT-IV), with a single score, an untimed test of receptive vocabulary or the quick estimate of the examinee's receptive vocabulary (Dunn & Dunn, 2007);
- semantic (or category) fluency test (SFT) with one score (total correct animal naming), a psychometric test for verbal generativity (Spreen & Strauss, 2006);
- test of variables of attention (TOVA) with one summary score (attention performance index, API), a diagnostic measure of attention deficits in children (Leark et al., 2004); and
- Purdue pegboard test (PPT), a measurement of unimanual and bimanual finger and hand dexterity, which included 3 scores, i.e., assessments using preferred, non-preferred, and both hands (Spreen & Strauss, 2006).

We focused on scaled individual cognitive scores, or the raw score if substantive normative data was not available. A scaled score represents where a child's raw score is in the distribution of raw scores among his or her peers[1] at the same age. The mean is 10 and the standard deviation is 3.

[1] The normative data for the tests are primarily based on children residing high-income Western countries, particularly the United States.

3.2.3. Neuroimaging assessment

Neuroimaging took place at the Cape Universities Brain Imaging Centre (CUBIC) using a 3 Tesla Allegra MRI scanner (Siemens Erlangen, Germany) with a single channel head coil. A single scanning session lasted under 60 minutes and was performed without sedation.

We acquired a high-resolution motion-corrected structural T_1-weighted dataset using a 3-dimensional echo planar imaging (EPI) navigated multi-echo magnetization prepared rapid gradient echo (MEMPRAGE) sequence (van der Kouwe et al., 2008) with voxel size = 1.3×1.0×1.0 mm^3, field of view (FOV) = 224×224×144 mm^3, times of echo (TEs) = [1.53, 3.19, 4.86, 6.53] ms, time of repetition (TR) = 2530ms, time of inversion (TI) = 1160ms, flip angle = 7°, bandwidth = 650 Hz/px, matrix size = 224×168, and 144 sagittal slices.

A further acquisition of two diffusion-weighted (DW) datasets used a volumetric navigated (Alhamud et al., 2012) twice refocused spin-echo (TRSE) sequence: 2×2×2 mm^3 voxel size, 220×220×144 mm^3 FOV, 10000ms TR, 86ms TE, 112×112 matrix size, 72 slices, 4 non-DW b_0 = 0 s/mm^2, b_1 = 1000 s/mm^2, 30 diffusion directions, and opposite phase encodings (anterior-posterior and vice versa) for EPI distortion correction during processing.

Single-voxel ^{1}H-MRS from the midfrontal gray matter (MFGM), peritrigonal white matter, and basal ganglia acquisition used an EPI navigated point resolved spectroscopy (PRESS) sequence (Hess et al., 2011): 1.5×1.5×1.5 cm^3 voxel size, 2000ms TR, 30ms TE, and 64 averages with chemical shift selective (CHESS) water suppression (Haase et al., 1985). An acquisition without water suppression was performed for water referencing.

3.2.4. Image pre-processing and feature generation

Automated brain extraction, cortical reconstruction, and volumetric segmentation were performed on the structural T1-weighted dataset with FreeSurfer version 6.0 (Dale et al., 1999; Fischl et al., 2002) as described by Nwosu and colleagues (2018). We computed

sMRI cortical measures—thickness, area, mean curvature, volume, and local gyrification index for each of the 68 regions (34 in each cerebral hemisphere) of the Desikan-Killiany (DK) atlas (Desikan et al., 2006). Furthermore, 47 sMRI subcortical and other non-cortical regional and total volumes (e.g., total gray matter, estimated intracranial volume) were included in the analyses.

On the DW datasets, Tortoise v.2.5.2 (Pierpaoli & Walker, 2010) was used for motion and field distortion corrections as well as voxel-wise calculation of the diffusion tensor. We calculated voxel-wise diffusivity measures in the analysis of functional neuroimages (AFNI) software (Cox, 1996) as described by Jankiewicz and colleagues (2017). We calculated average values of fractional anisotropy (FA) as well as axial, mean, and radial diffusivities (AD, MD, and RD) for each of the 20 white matter tracts in the Johns Hopkins University (JHU) atlas (Mori, 2007).

The processing of ^{1}H-MRS data was performed using LCModel version 6.1 (Provencher, 2001) as described by Robertson and colleagues (2018). We calculated the absolute concentrations of 11 metabolites as well as their ratios to total creatine (CrPCr): creatine (Cr), phosphocreatine (PCr), glutamate (Glu), myo-inositol (Ins), N-acetylaspartate (NAA), choline (Cho), phosphocholine (PCh), total choline (GPCPCh), glutamate plus glutamine (GluGln), N-acetylaspartylglutamate plus N-acetylaspartate (NAAGNAA), and CrPCr. We excluded metabolite concentrations from peritrigonal white matter and basal ganglia as the acquisition in many subjects were of poor quality, reducing the overall sample size considerably. Therefore, a total of 489 imaging features (5×68 cortical sMRI + 47 other sMRI + 4×20 DTI + 22 MFGM ^{1}H-MRS) were included as predictors in the models.

3.2.5. Quality control and exclusion criteria

Participants with a history of major medical or psychiatric disorder, learning disability, and the use of psychiatric medication that might impact cognitive performance (medication for ADHD) were excluded. We wanted to include only healthy, virally suppressed

children with no acute illness in the study, so children living with HIV were also excluded if they had detectable viral loads (>400 copies/ml) at age of scan or were symptomatic in any way. Cognitive assessments were video recorded for quality control and administered in the child's preferred language (English, isiXhosa, or Afrikaans), and the translations were conducted under license from test publishers (van Wyhe et al., 2021).

We eliminated T1-weighted and DW datasets with motion artefacts, and poor-quality ^{1}H-MRS spectra (broad linewidth, FWHM > 0.07 ppm and/or low signal-to-noise ratio, SNR < 6). Images with dropout slices, or failed alignment to the JHU and DK templates in standard space were also eliminated. Furthermore, subjects were excluded if they did not complete the full scanning session, missed one or more cognitive assessments or any of the individual neuroimaging modalities.

3.2.6. Construction of cognitive performance categories via hierarchical clustering

Prior to the classification analysis, we concatenated the cognitive scores to form an 11-element vector score representing overall cognitive performance. An agglomerative (bottom-up) hierarchical clustering algorithm (HCA) was implemented to group the children into categories of cognitive performance using this 11-dimensional vector score. We began with no predefined clusters (categories) and sought to find natural divisions in the data. This is an important step, since it allows for a verification of the extent to which the resulting clusters or categories are meaningful and may indicate a departure from typical cognitive functioning.

HCA involves using a similarity/dissimilarity function to measure how related pairs of subjects are, a linkage function to sequentially group pairs into larger clusters to form a hierarchical tree, and cophenet correlation/consistency function to quantify how faithfully the clusters represent dissimilarities between subjects (Murtagh & Contreras, 2012, 2017). We used the Euclidean distance as the similarity metric, i.e., we first calculated the L^2-norm distance between individual subjects' vectors. Clusters can be linked using

single, centroid, median, ward, average, and complete linkage methods. To evaluate the effect of using different linkage methods we calculated the cophenetic coefficient for each method to measure how faithfully the HCA preserved the pairwise distances in the dataset.

The result of HCA can be visualised with a dendrogram (tree) shown in Figure 3.3. To divide the tree into clusters with greatest dissimilarity, we located the links in the tree with the largest inconsistency coefficient (φ), which is the ratio of the height (δ) of the link to the average height of links that lie below it. When a link has approximately the same δ as links below it, the cluster distinction between these two levels is not significant and φ is low, indicating similar clusters which can be merged. A high φ indicates dissimilar clusters, highlighting a natural division in the data. The optimal number of clusters from the clustering solution can also be determined using the Silhouette criterion, measuring how a point is similar to other points in the same cluster compared to points in other clusters (Rousseeuw, 1987); or the Calinksi-Harabasz criterion, which maximises the ratio of the between-cluster and within-cluster variances with respect to the number of clusters (Caliński & Harabasz, 1974).

3.2.7. Predictive modelling algorithms

We chose to compare the following 3 families of algorithms: penalised linear models, support vector machines, and decision tree ensembles.

Penalised linear models (regularisation-based models)

In linear predictive modelling using regression, the response variable (e.g., cognitive outcome) is a weighted sum of its predictor variables (e.g., neuroimaging measures) plus random noise. Ordinary least squares (OLS) method is the simplest regression solution, which estimates the weights/coefficients of the variables in the prediction model by minimising the residual sum of squares (RSS). OLS is unstable and ineffective for high-dimensional datasets (Harrell, 2001). Penalised regression solutions introduce constraints or penalties on the size of the weights while minimising RSS, thereby solving the instability problem posed by OLS. Overall, penalisation/regularisation improves the prediction error by decreasing the variance of the estimated weights. The least absolute shrinkage and selection operator (lasso, L^1-norm penalty) shrinks some of the coefficients to exactly zero (Tibshirani, 1996). Ridge technique (L^2-norm penalty) makes predictors that are irrelevent approach zero (Awad & Khanna, 2015). The elastic net (EN) is the generalisation of lasso and ridge regression (Zou & Hastie, 2005). Penalised regression can also be extended to classification, as such, neuroimaging classification problems often use the EN estimation method (Bowman, Drake, & Huddleston, 2016; Schouten et al., 2016; Tohka et al., 2016).

Support vector machines/regression models

Linear SVMs and kernel-based SVMs, with polynomial, gaussian radial basis function (RBF), and sigmoid kernels, are used to find hyperplanes or decision boundaries used to predict a category or continuous output in n-dimensional space (Scholkopf & Smola, 2018; Schölkopf et al., 2002). The points closest to the hyperplane corresponding to the data are termed support vectors. The goal of SVM/SVR is to fit the best line without a

threshold value, a distance between the hyperplane and the boundary lines (two lines drawn around the hyperplane to create a margin between the data points).

In real world data, it is often difficult to classify or regress data points in lower dimensions. Kernels are useful mathematical functions to transform data from a space where it is not linearly separable to a high dimensional space in which it can be separated by a hyperplane (Scholkopf & Smola, 2018; Schölkopf et al., 2002). They bridge the gap between linearity and non-linearity of a dataset. There are no established principles or general rules to know which kernel will work best for a particular dataset, it depends primarily on the data at hand. For large datasets where the number of predictors exceeds the number of observations, linear SVM/SVR is often suggested, an RBF kernel is otherwise used when dealing with large amounts of data.

Decision tree ensembles (tree-based models)

Bagging (bootstrap aggregating) and boosting are the two popular ensemble methods to combine weak learners or individually trained decision trees to achieved superior predictive performance (Plaia et al., 2022). Bagging takes a random subset of the original dataset, with replacement, and fits either a classifier or regressor to each subset. The predictions for each subset are then aggregated through a majority vote for classification or averaging for regression, thereby increasing accuracy. An example of bagging is the random forest technique. Boosting on the other hand trains decision trees iteratively on the residuals of the previous trees, residuals being errors made by the previous weak learner in classifying or regressing the data. This sequential training by the boosting algorithm corrects the residuals by adjusting the weights of the training instances for the tree next in line, leading to a reduction in the bias of the model by combining the results of multiple weak learners. There are several sub-types of boosting algorithms, such as adaptive, random under-sampling, least squares, and totally corrective boosting, each having its own strengths and weaknesses depending on the data at hand or whether the problem is regression or classification.

3.2.8. Statistical analyses

Analyses were implemented in the MATLAB Statistics and Machine Learning Toolbox, Software Release R2021a (MathWorks, Inc., 2021). Our analysis was split into 3 sub-analyses: (1) regression analysis or predicting continuous values of each of the 11 cognitive scores, (2) classification analysis or predicting the categories resulting from the hierarchical clustering algorithm, and (3) extracting relevant features of the best performing prediction models. After performing each analysis on the set of all children in the study sample we repeated the analysis for the subset of children living with HIV and the subset without HIV separately.

Regression and classification performance

We trained 3 different kinds of predictive models: DTEs which encompassed totally corrective boosted (TotalBoost), adaptive boosted (AdaBoost), random subsampling boosted (RUSBoost), least-squares boosted (LSBoost), and bagged trees; SVMs/SVR with linear, gaussian, quadratic, and cubic kernels; and PLMs which included elastic net, ridge, and lasso penalisations. In essence, we trained optimizable classifiers or regression models in each category. The machine learning toolbox employs a Bayesian optimisation algorithm to optimise model fit. We selected the best-point hyperparameters in each category that minimised cross-validation (CV) loss, i.e., lowest CV error for regression and highest CV accuracy for classifiers. We therefore compared the best DTE, SVM, and PLM models in terms of the regression and classification of cognitive performance scores or categories.

For each prediction model, children were randomised into k = 10 folds for k-fold CV (Kohavi, 1995). In the first iteration, children in Fold 1 were used as a testing (validation) set, and Folds 2 through 10 formed the training set. In the next iteration, those in Fold 2 were tested on and the rest were part of training the prediction model. This was done iteratively until each fold had been used either as a validation or training set. Ten-fold

CV avoids overfitting and selection bias in neuroimaging studies with relatively small sample size (Tohka et al., 2016).

For the regression models, we computed the fitting error (in-sample/resubstitution loss) for each observation in the response using all other training data, and cross-validated root mean square errors (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and correlation coefficient (r). Generalized error (validation RMSE) measures how the models will perform on unseen data, i.e., the score is the RMSE on all observations, counting each observation when it was in a held-out validation fold. MAE is like RMSE, but less sensitive to outliers. R^2 was calculated using $1 - \frac{SSE}{SST}$ (where SSE and SST are sum of squared error and total, respectively). R^2 reflects the variance in the dependent variable that can be predicted by the independent variables. We computed the correlation coefficient r between the actual and predicted values to test how similar they were.

The classification performance was assessed via cross-validation accuracy, percentage of correct predictions, as well as the area under the receiver operating characteristic curve (AUC) or C-statistic, an overall performance measure summarising the diagnostic ability of a classifier as its discrimination threshold (all possible category decision thresholds) is varied. A flowchart of the regression and classification is shown in Figure 3.1.

Feature selection

DTE and PLM select relevant features while the models are being created—this is the embedded feature selection technique part of the regression or classification algorithm (Khaire and Dhanalakshmi, 2022). For DTE, the weight/coefficient of each feature is calculated and used to construct the tree, a feature ranking is then obtained on the sum of the improvements in all tree nodes; redundancy and relevance is quantified using the pairwise mutual information of features and that of a feature and the response, which is similar to the minimum redundancy–maximum relevance (MRMR) algorithm (Ding &

Peng, 2005). PLM regularization methods introduce penalties in the learning process, making coefficients of irrelevant predictors equal to or close to zero.

Having fitted a Linear SVR/SVM model, it is possible to access the classifier or regression coefficients on the trained model. Feature importance is then determined by comparing the size of these coefficients. The coefficients are less interpretable when kernels are used, therefore another method is usually required to determine the most important features in SVM/SVR. For this reason, we obtained SVR/SVM relevant features by implementing wrapper feature selection (Khaire & Dhanalakshmi, 2022), using the output SVM classification model as an objective function to rank the features. We used recursive feature elimination (SVM-RFE). All features were first given weights, e.g., coefficients of the linear SVM model, after which the weakest coefficient was eliminated, and the model refitted to determine the new weakest coefficient. This procedure was repeated until all irrelevant features were removed and the remaining features achieve the same classification accuracy as the SVM classification model.

We included the possible non-imaging confounders age at scan, birth weight, sex, and HIV status as candidate predictors, i.e., estimated them jointly with neuroimaging data in the predictive model (Linn et al., 2016) instead of regressing them out before constructing the models (Adeli et al., 2018). This allowed the predictive model to select them if they were relevant features for predicting cognitive performance.

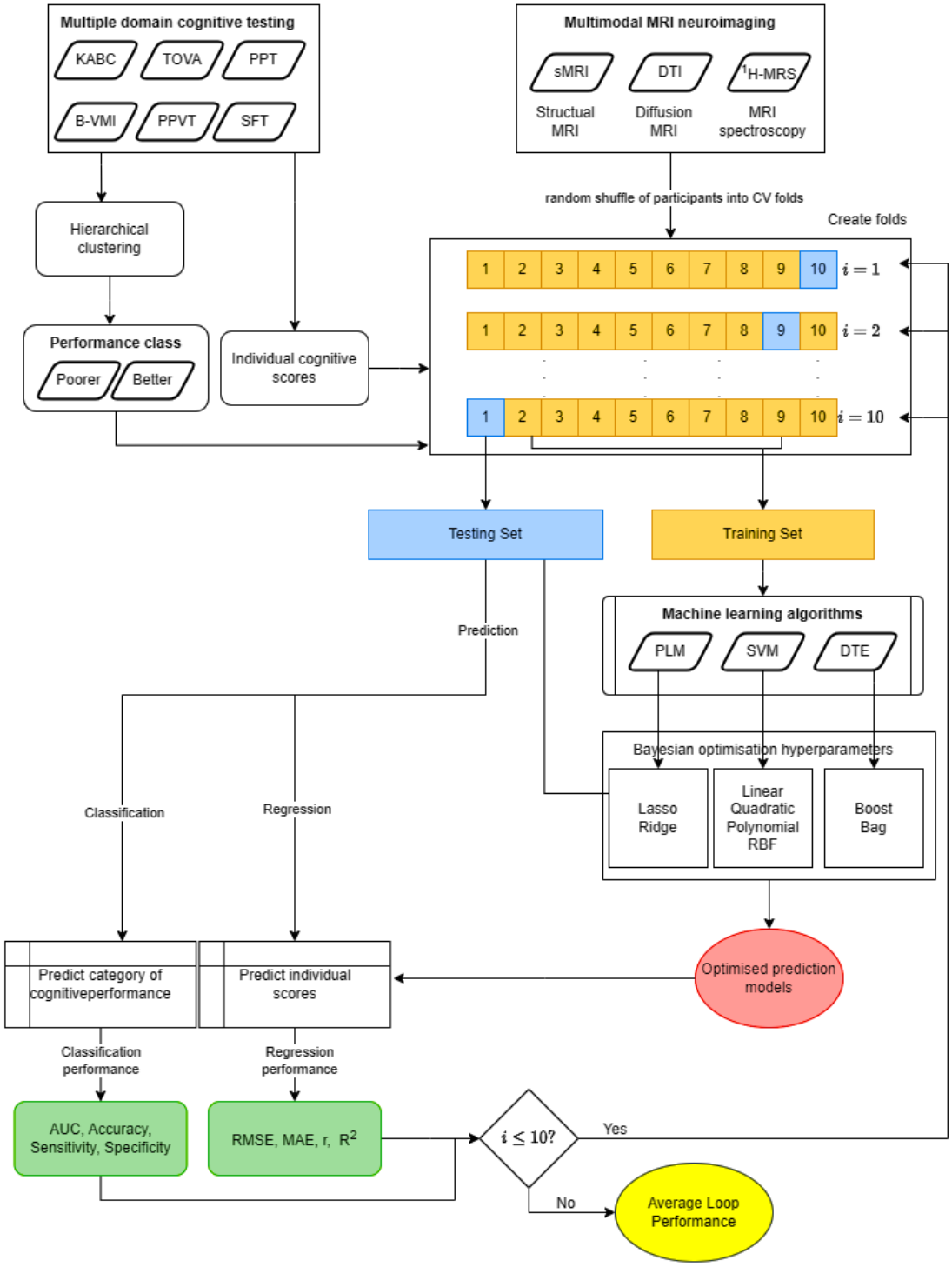


Figure 3.1| Illustration of classification of categories of cognitive performance and regression of continuous individual scores of cognitive performances using 10-fold cross-validated and optimised prediction models PLMs, SVMs, and DTEs.

3.3 Results

3.3.1. Final sample after exclusions

After quality control, 14 children were excluded for missing one or more cognitive scores due to incomplete assessments or failure to pass the practice test numerous times. We excluded 5 children with detectable viral loads at age of scan. A further 2 did not complete the T_1-weighted sMRI acquisition, 10 DW images had motion artefacts, 12 had poor ^{1}H-MRS spectra, therefore these participants were also excluded. Ultimately, a total of 89 subjects were included in the prediction analyses.

3.3.2. Clustering of cognitive scores

Correlation of the cognitive scores

Figure 3.2 shows a pairwise correlation map of the 11 cognitive scores; none of the pairs show strong correlation (Pearson correlation coefficients, $r < +0.51$). The KABC scores have a moderate positive correlation ($0.42 < r < 0.46$), and PPT preferred and non-preferred had the strongest correlation of any pair, with moderate positive correlation of $r = 0.51$.

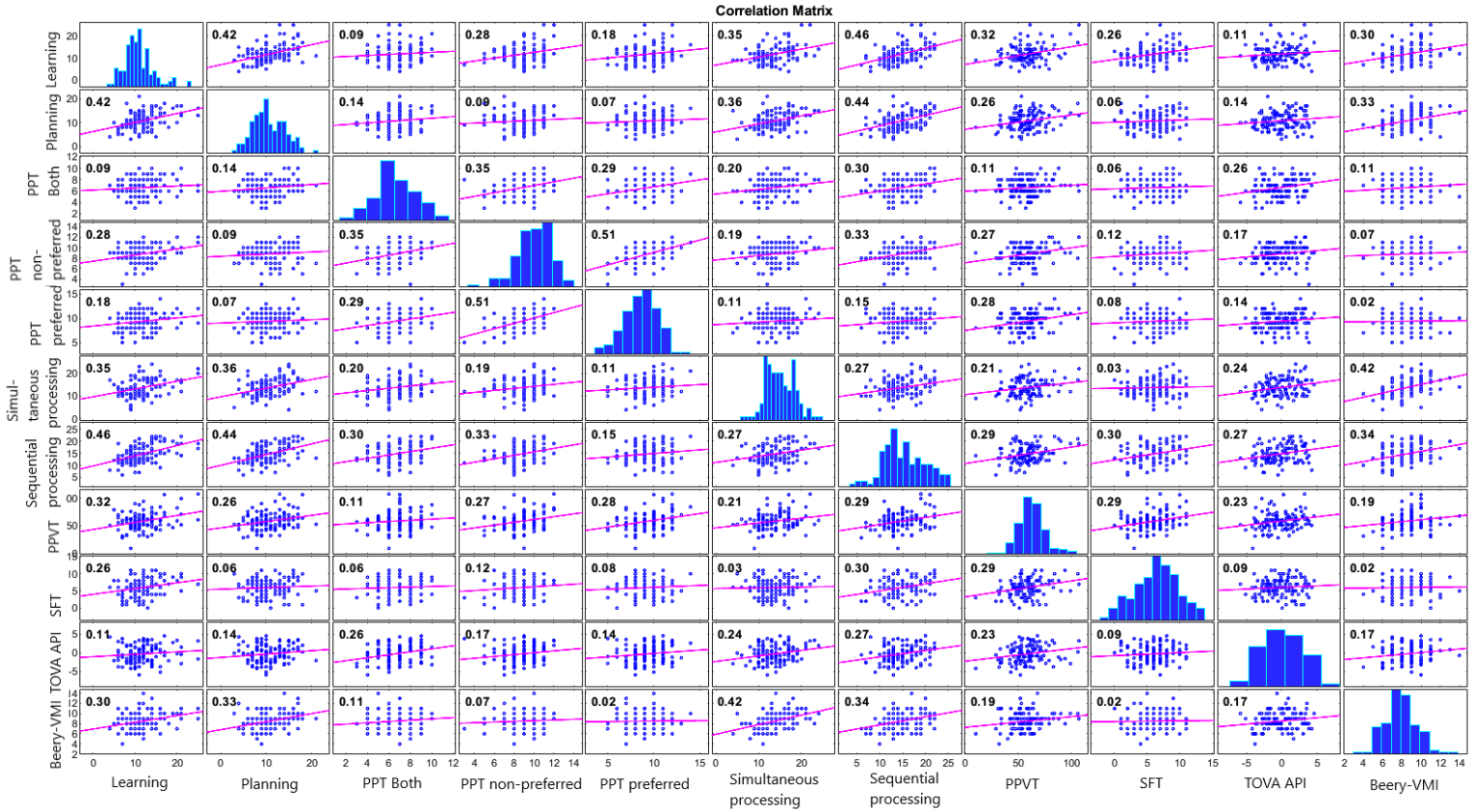


Figure 3.2| Correlation matrix showing the pairwise correlation strengths among the 11 cognitive scores. The pair showing the strongest correlation (moderate positive correlation) is PPT preferred and non-preferred hands (r = +0.51).

Hierarchical clustering

A cophenetic coefficient closer to 1 indicates higher validity of the output from HCA. Complete linkage (the farthest neighbour method) resulted in the highest cophenetic coefficient (c = 0.7311) and a monotonic cluster tree (shown in Figure 3.3), while centroid and median linkages either resulted in non-monotonic clusters, and ward, median, average, and single linkages had lower cophenetic coefficients. The hierarchical tree was therefore constructed using distance between farthest objects in the separate clusters to represent distance between clusters.

Using this method, we identified two main clusters of children with the highest dissimilarity ($\varphi = 1.15$) between cognitive scores. The two main clusters contained identifiable sub-clusters: the lowest scoring children (Figure 3.3, green colour-coded cluster) and highest scoring children (Figure 3.3, blue colour-coded cluster). However, these sub-clusters were not significantly different from their parent clusters. The division of the tree into two clusters was further supported by the Silhouette and Calinski-Harabasz criteria (Figures 3.4A–B), where the highest silhouette value occurs at two clusters, and 2 clusters had the largest between-cluster variance and a smallest within-cluster variance.

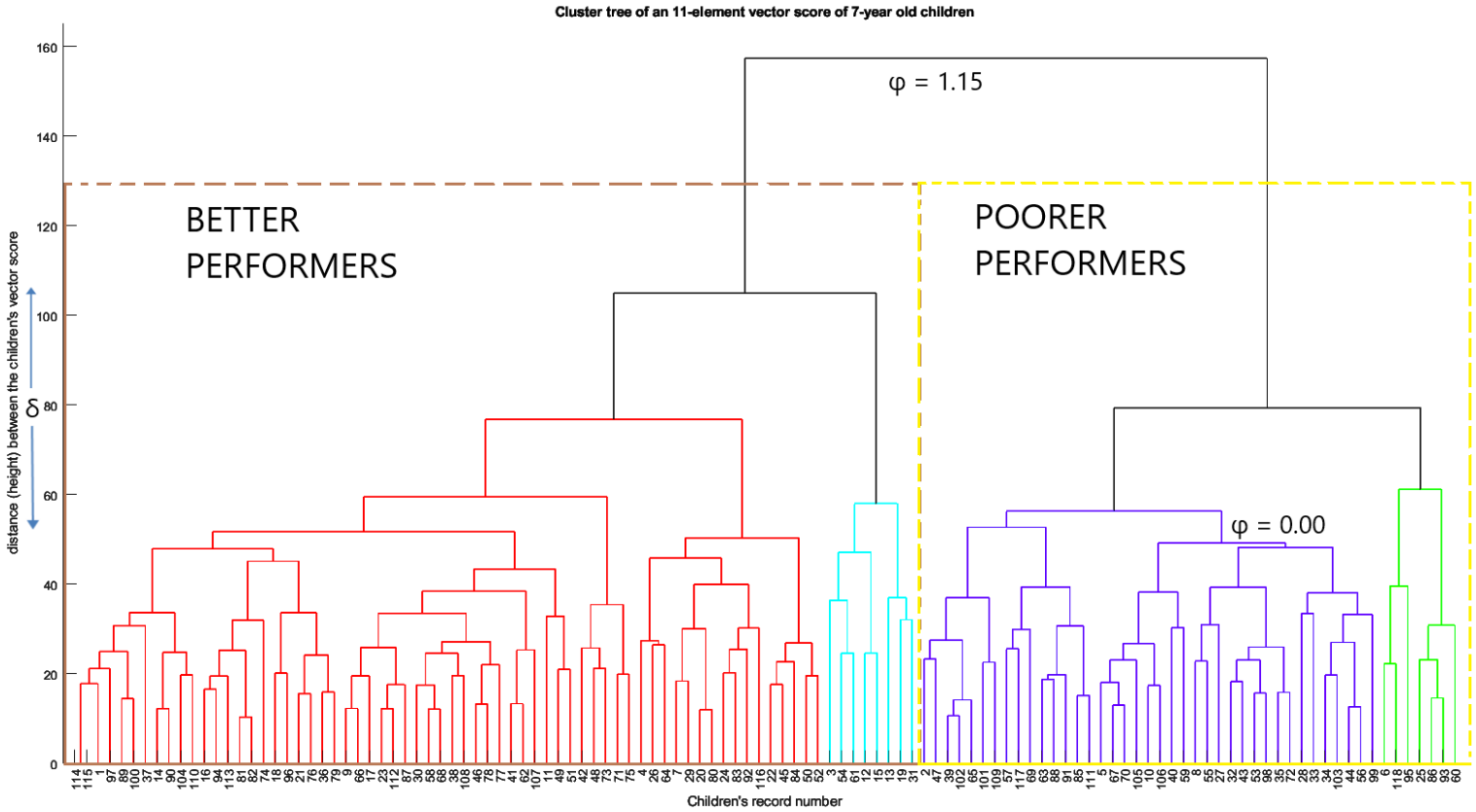


Figure 3.3| A dendrogram (cluster tree) showing the result of hierarchical clustering algorithm on the 11-dimensional cognitive score vector. Shown are two main clusters with high consistency links, which are the groups of better and poorer performing children. The colour-coded subsets are the lowest scorers (green), highest scorers (blue), below average scorers (violet), and average (red) scoring children compared to normative data.

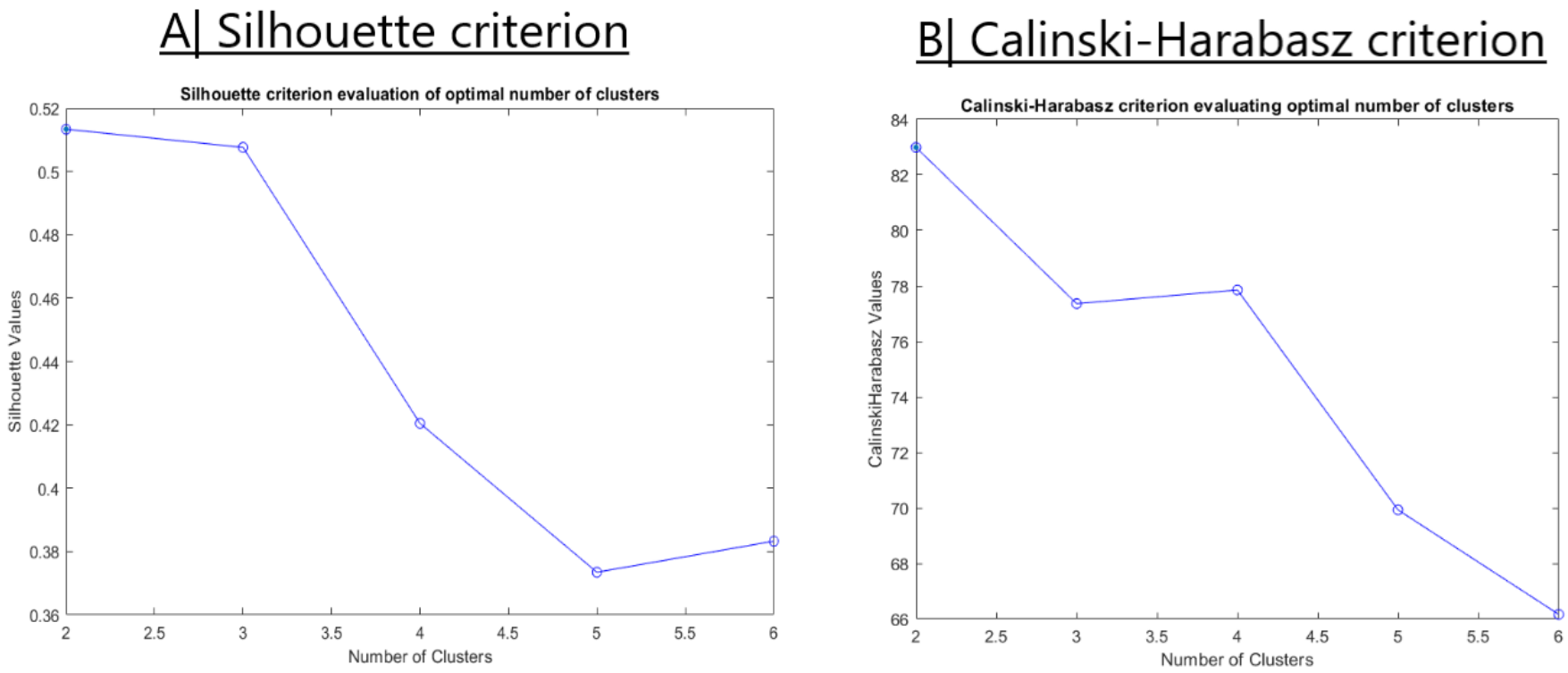


Figure 3.4| Hierarchical clustering evaluation using the silhouette criterion (A) and Calinksi-Harabasz criterion (B). The two criteria show that the optimal number of clusters is 2.

Demographic comparison between clusters

The two clusters did not differ with respect to the demographic or non-imaging variables—age, birth weight, or gender/sex (p's > 0.12, Table 3.1). Similarly, while the poorer performers had 20% more participants with HIV than without, better performers had approximately the same number children living with and without HIV, so HIV status was not statistically different ($p = 0.30$) between clusters i.e., HIV does not explain the differences in cognition between the two clusters.

The first cluster (*poorer performers*) had significantly lower scores (p's < 0.01) than the second cluster (*better performers*) for all scores except Beery VMI ($p = 0.07$) and PPT preferred hand ($p = 0.09$) and both hands (p = 0.63), where the means and medians were lower for *poorer performers*, but not statistically significant (Table 3.1).

Looking at the broader paediatric population taking the KABC-II test around this age (Kaufman and Kaufman, 2014), sequential and simultaneous processing scores for *better performers* are in the upper average range (14–16), while *poorer performers* have average scores (7–13). Both planning and learning are in the average category for the two clusters. For Beery-VMI, *poorer performers* score below average (21–30 percentile), and *better performers* score average (32–42 percentile). While *better performers* have normative (about zero) TOVA's API scores, *poorer performers* score closer to the ADHD sample (API = -1.04). PPT scores for both clusters are below average (10–20 percentile) when compared to the normative data provided by Gardner and Broman (1979). There was no substantive normative data to compare the sample's PPVT-IV and SFT to their out-of-sample peers at the same age.

Table 3.1| Sample characteristics of the two clusters obtained via hierarchical clustering, showing the *poorer performers* with significantly lower scorers than *better performers*.

	Poorer performers (n = 40)	*Better performers* (n = 49)	*p*-values from Wilcoxon rank sum test
Cognitive scores			
KABC-II			
Learning	10 (2.00)	12 (4.25)	6×10^{-5}
Planning	9 (4.75)	12 (5.25)	5×10^{-3}
Sequential processing	13 (3.50)	16 (6.25)	3×10^{-3}
Simultaneous processing	13 (3.75)	15 (5.00)	5×10^{-2}
Beery-VMI	8 (3.00)	9 (1.25)	0.07
PPVT-IV	52 (10.50)	64 (14.00)	1×10^{-8}
SFT	5 (2.00)	7 (3.00)	0.01
TOVA API	-1.04 (3.21)	-0.05 (3.51)	0.03
PPT			
Preferred hand	9 (2.00)	10 (2.25)	0.09
Non-preferred hand	8 (2.00)	9 (2.00)	0.01
Both hands	7 (1.00)	7 (2.00)	0.63
Demographics			
Age (years)	7.20 (0.15)	7.22 (0.16)	0.12
Birth weight (g)	3200 (580)	3173 (460)	0.68
Gender* Female n (%)	23 (58%)	21 (43%)	0.17
Status* HIV n (%)	24 (60%)	24 (49%)	0.30

*A chi-squared (χ^2) test was used to compare percentage of children between the two clusters
Values are median (IQR) or n (%)
We used an alpha significance level of $p = 0.05$

3.3.3. Comparison of regression performance between SVM, DTE and PLM models

After quality control, the sample sizes were small to perform the regression analyses in children with (n = 48) and without HIV (n = 40) separately using 10-fold cross validation, as such we used 5-fold cross validation for these groups of children. However, the prediction of the 11 cognitive scores individually in children with and without HIV

separately resulted in poor regression performance (supplementary Tables B.2 and B.3). None of the PLMs, DTEs, and SVMs models predicted the individual scores with r, $R^2 > 0.18$.

Across all subjects (children with and without HIV, n = 89): when the cognitive scores were predicted individually from neuroimaging measures, the regression models resulted in small resubstitution error rates but high cross-validation errors, suggesting overfitting and poor performance on unseen data. We present the out-of-sample error rates in Table 3.2. Figures 3.5A–D illustrate the training and validation errors of a subset of scores (KABC-II learning, PPVT, SFT, and Beery-VMI) which showed the highest positive associations ($r \approx 0.30$) between predicted and actual values. SVMs tended to have lower RMSE than DTE and PLMs, while DTE tended to have lower MAE than the other 2 models. PLMs tended to be worse than a prediction with a constant mean response of the scores ($R^2 < 0$). Overall, the linear SVM algorithm achieved lower CV errors than the other algorithms, and R^2 for all the models were small.

Table 3.2| Regression performance resulting from the prediction of the 11 cognitive scores individually using multimodal MRI at age 7 for both children living with and without HIV. Lowest RMSE and MAE, and high r and R^2 for the best algorithm among PLMs, SVMs, and DTEs is given in **bold**.

Cognitive test	Model category	RMSE	MAE	r	R^2	Algorithm with lowest CV errors
Learning	PLM	4.18	3.03	-0.44	-0.05	Linear SVM
	SVM	**3.96**	**3.02**	**0.34**	**0.10**	
	DTE	4.16	3.02	0.30	0.01	
Planning	PLM	6.02	5.11	0.03	-1.45	Linear SVM
	SVM	**3.87**	**3.28**	**0.04**	**0.00**	
	DTE	3.96	3.26	0.04	-0.04	
Simultaneous processing	PLM	3.98	3.16	-0.25	-0.02	Bagged DTE
	SVM	4.02	3.20	0.10	0.00	
	DTE	**4.00**	**3.15**	**0.20**	**0.01**	
Sequential processing	PLM	5.17	4.28	-0.10	-0.92	Linear SVM
	SVM	**3.56**	**3.01**	**0.03**	**0.11**	

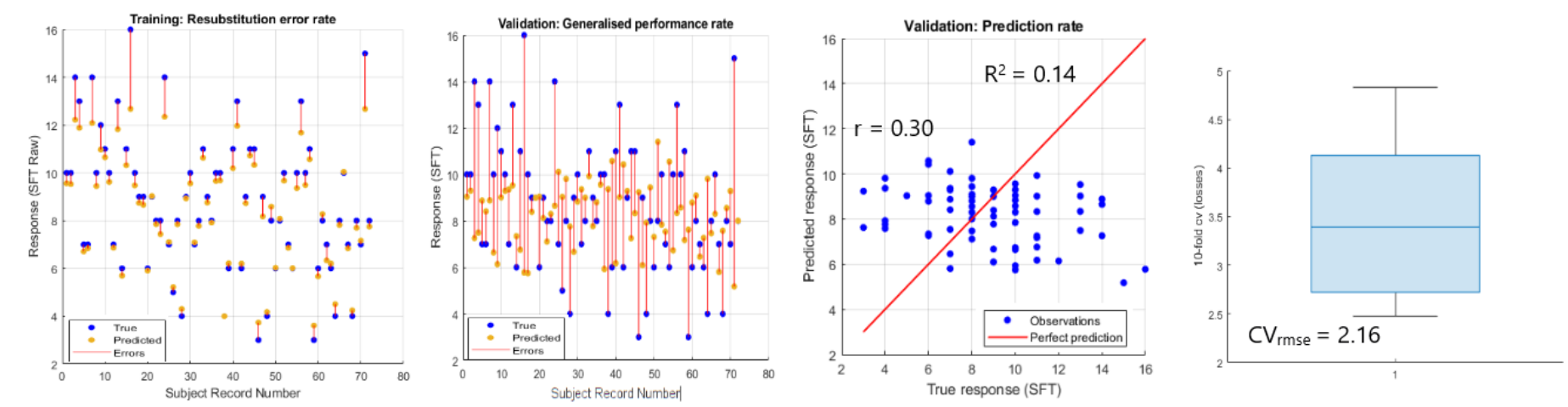
--Table 3.2 continued--

	DTE	3.77	3.17	0.03	0.00	
Beery-VMI	PLM	1.58	1.25	-0.54	-0.07	
	SVM	**1.52**	**1.20**	**0.28**	**0.04**	Linear SVM
	DTE	1.55	1.22	0.20	0.00	
PPVT	PLM	15.50	11.17	0.02	-1.24	
	SVM	**16.28**	**11.57**	**0.31**	**0.00**	Linear SVM
	DTE	16.32	11.49	0.28	0.00	
SFT	PLM	2.61	2.15	0.07	-0.30	
	SVM	2.32	1.87	0.10	0.00	Bagged DTE
	DTE	**2.16**	**1.67**	**0.30**	**0.14**	
TOVA API	PLM	2.40	1.99	-0.24	-0.01	
	SVM	2.37	1.91	0.12	0.03	Bagged DTE
	DTE	**2.32**	**1.91**	**0.11**	**0.07**	
PPT Preferred	PLM	1.77	1.45	-0.28	-0.02	
	SVM	**1.58**	**1.28**	**0.19**	**0.20**	Linear SVM
	DTE	1.70	1.37	0.21	0.09	
PPT non-preferred	PLM	1.64	1.29	-0.41	-0.03	
	SVM	**1.64**	**1.30**	**0.05**	**0.00**	Linear SVM
	DTE	1.64	1.30	0.18	-0.01	
PPT Both hands	PLM	1.69	1.38	-0.22	-0.49	
	SVM	**1.39**	**1.09**	**0.29**	**0.03**	Linear SVM
	DTE	1.41	1.11	0.26	0.00	

A| KABC Learning

B| Purdue picture vocabulary test

C| Semantic fluency test



D|Beery VMI

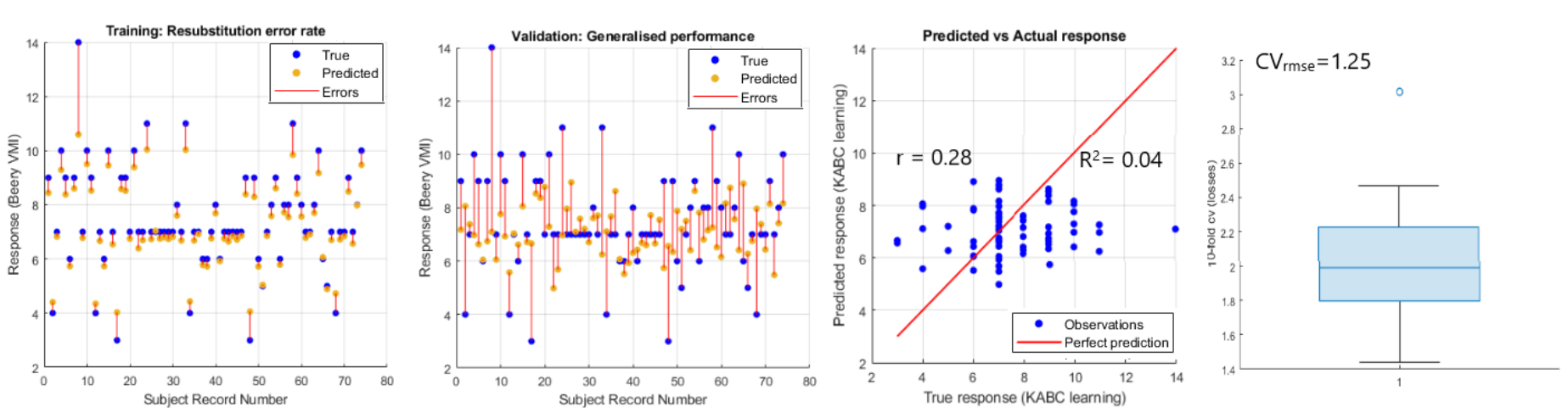


Figure 3.5| The prediction of individual continuous cognitive scores at age 7 using multimodal MRI at age 7 for children living both with and without HIV. The subfigures show the prediction of (A) KABC-II learning, (B) PPVT, (C) SFT, and (D) Beery-VMI. For each sub-figure, the first graph is the fitting/resubstitution validation errors, second figure is cross validation errors, third figure is the prediction rate, and fourth is the 10-fold cross-validation RMSEs. The models show small training (fitting) errors, but high generalised errors.

3.3.4. Comparison of classification performance between SVM, DTE and PLM models

Children were assigned to either the *poorer performers* or *better performers* cluster. These labels were used to train the SVM, DTE, and PLM supervised learning classifiers. Since there were two main clusters, we also computed sensitivity and specificity of the classifiers at the optimum cut-off point, i.e., the point closest to true positive rate of 1 and false positive rate of 0. Similar to the regression models, the classification analyses using the children living with and without HIV separately were performed with 5-fold CV, and the models resulted in poor classification performances. For children living with HIV (n = 24 *poorer* vs 24 *better* performing children), we could only achieve AUC of 0.55–0.61 (supplementary Table B.4). And for children without HIV (n = 16 *poorer* vs 25 *better* performing children), AUC was 0.53–0.57 (supplementary Table B.5).

Across all subjects (n = 40 *poorer* vs 49 *better* performing children) after predicting category memberships using the three techniques, the best performing model was linear SVM with classification performance metrics: 0.83 AUC, 74% accuracy, 70% specificity, 85% sensitivity. RUS boosted trees (DTE) and lasso PLM have generalised performances of 0.78 and 0.71 AUC, respectively (Table 3.3). The diagnostic ability of each classifier with varying discrimination threshold is presented in Figures 3.6A–D.

Table 3.3| Classification performance of SVM, DTE, and PLM. Linear SVM (in bold) has the best generalised performance.

Technique	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	Balanced accuracy (%)*	Relevant Features
SVM	**0.83**	**85**	**70**	**74**	**78**	**40**
DTE	0.78	72	73	71	73	48
PLM	0.71	70	65	65	68	45

*Balanced accuracy was computed due to imbalances in category memberships. It is the arithmetic mean of sensitivity and specificity.

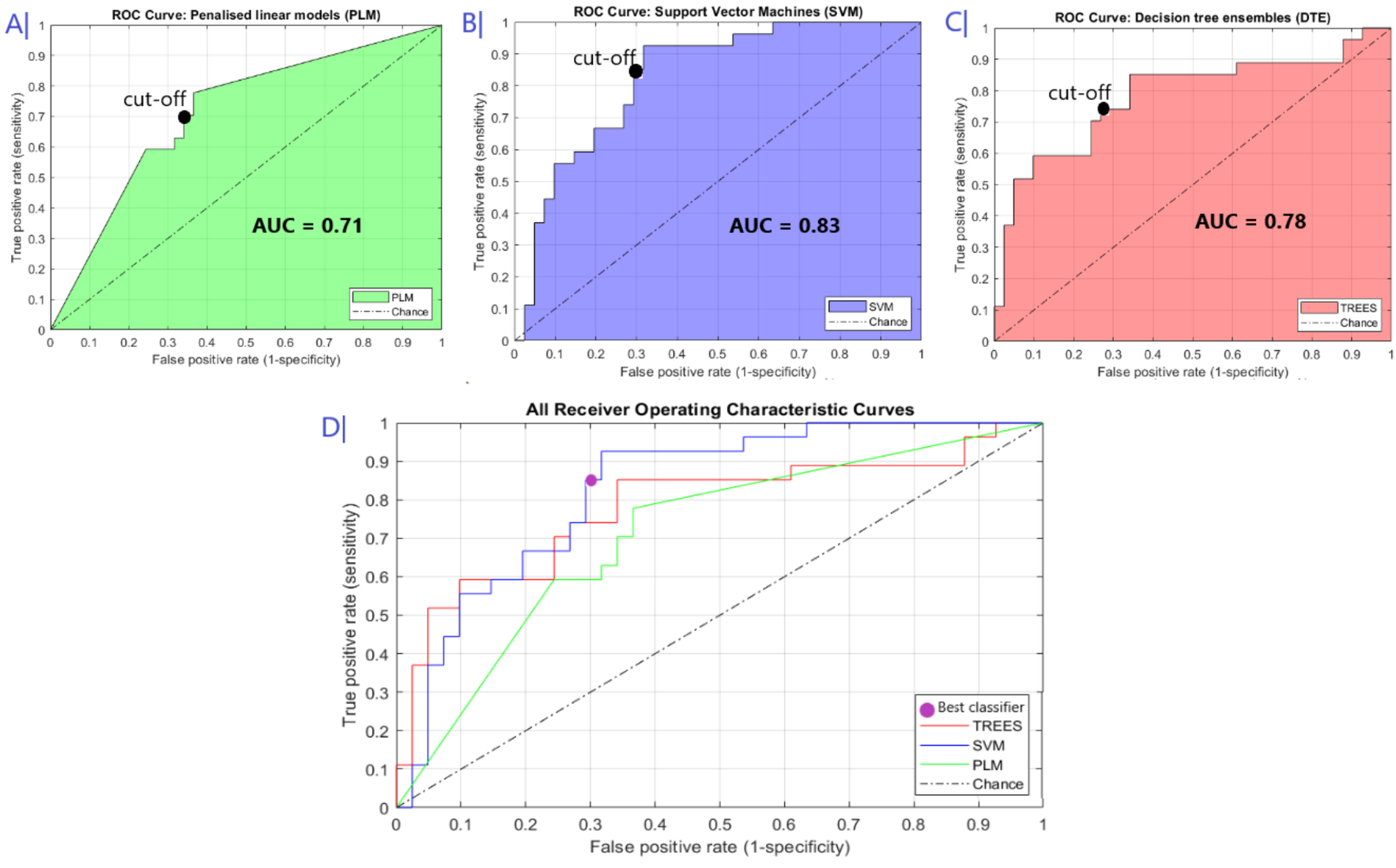


Figure 3.6| An ROC curve showing the diagnostic ability of (A) lasso PLM, (B) linear SVM, and (C) RUS boosted decision trees with varying discrimination threshold, as well as optimum cut-off points for each model. We also show the comparison of the models (D), and the optimum cut-off point of the best classifier, linear SVM.

3.3.5. Relevant features selected for the best prediction model

Since the regression performance of the individual scores was poor in general and the scores could be better predicted with a constant mean response, we did not look at the relevant neuroimaging features. Here, we present the relevant features extracted for the classification models, i.e., prediction of *poorer performers* versus *better performers* comprised of both children living with and without HIV.

There were 40 relevant features out of 489 total neuroimaging predictors obtained via SVM-RFE when we predicted categories of overall cognitive performance. For the other models, there were 48 features from DTE (supplementary Table B.6), and 45 from PLM (supplementary Table B.7). All 40 features from the best prediction model (SVM-RFE) were included in the other models, and there was a total of 49 unique features used by the three prediction models. We present the 40 relevant features in Table 3.4 and Figures 3.7A–F. None of the non-imaging measures were relevant features in the prediction of the cognitive score categories.

Table 3.4| Feature ranking of the 40 relevant features from the best performing modality (linear SVM) obtained via SVM-RFE.

Feature	Weight* $\times 10^{-4}$
Right pallidum volume	293.20
LH precentral volume	163.50
Left lateral ventricle volume	52.00
Right cerebellum white matter volume	36.69
RH middle temporal mean curvature	36.04
LH pars triangularis mean curvature	34.06
RH cerebral white matter volume	21.49
RD in left inferior longitudinal fasciculus	20.04
Right superior frontal volume	19.42
LH cerebral white matter volume	16.26
Left superior parietal area	18.71
Right paracentral volume	14.74
Left accumbens area volume	14.44
RH mean thickness	14.27
Left frontal pole area	13.95

--Table 3.4 continued--

Right postcentral volume	13.44
Left cuneus thickness	13.18
Left supramarginal volume	12.71
Left thalamus proper volume	12.24
Left supramarginal thickness	11.95
Right fusiform volume	11.75
Right transverse-temporal volume	11.48
CC anterior volume	10.96
AD in left uncinate fasciculus	10.81
Right para-hippocampal volume	9.98
CC posterior volume	9.26
Right isthmus cingulate volume	9.18
Left superior parietal volume	8.94
Left lingual volume	8.26
Left inferior lateral ventricle volume	8.14
Left frontal pole volume	8.08
Left hippocampus volume	6.95
Left caudate volume	6.47
Right temporal pole LGI	5.66
CSF volume	4.94
Left pallidum volume	4.72
Left sulcal banks volume	4.72
CC mid-posterior volume	4.43
CC central volume	2.61
Right amygdala volume	2.58

*Weights or coefficients in the final SVM model after recursive feature elimination

Abbreviations

LH - left hemisphere,

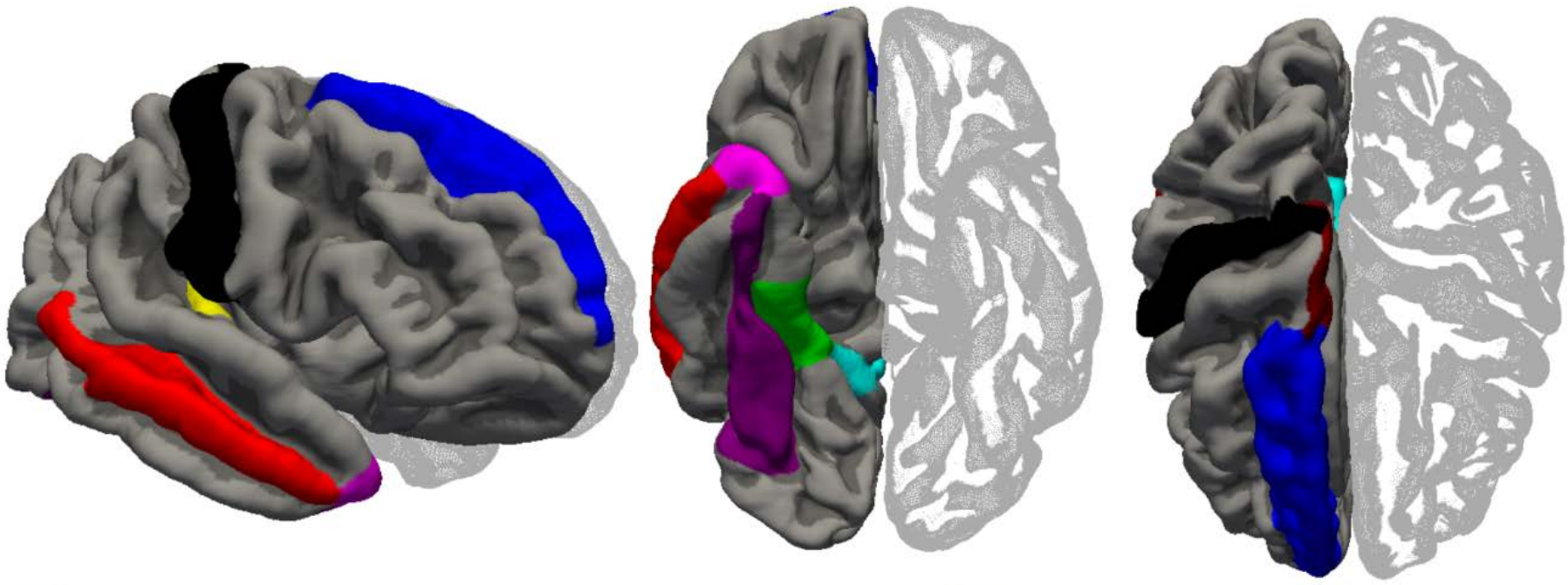
RH - right hemisphere

CSF - cerebrospinal fluid

CC - corpus callosum

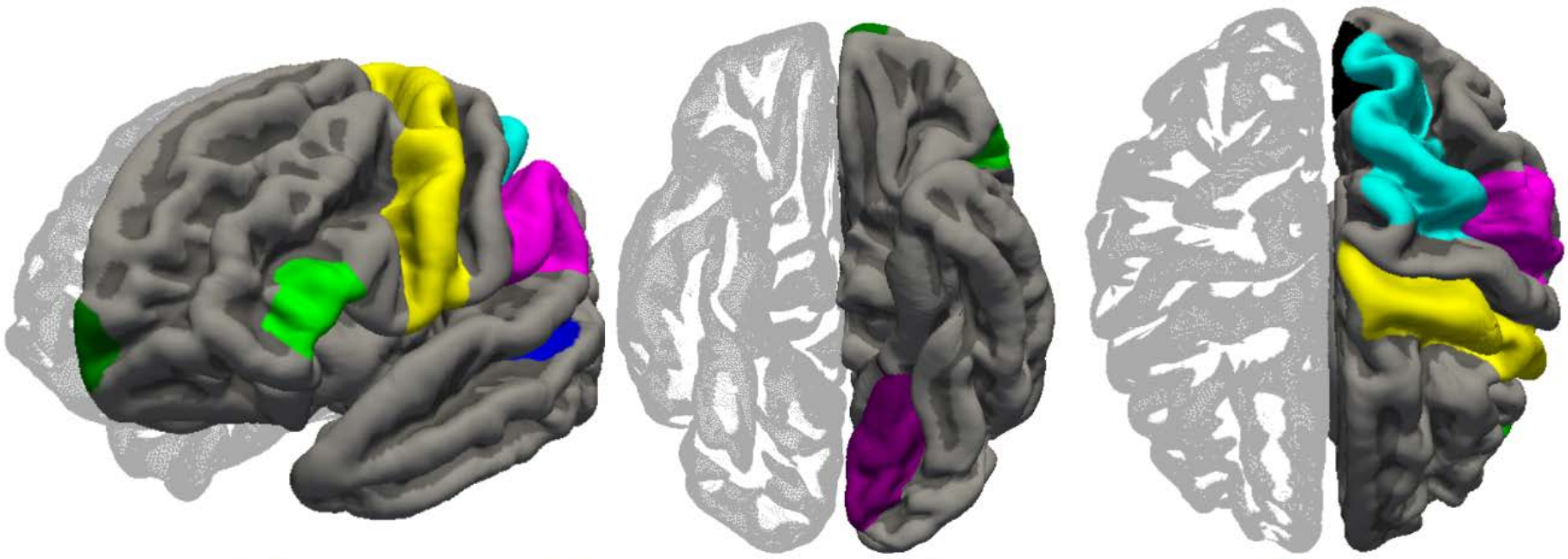
LGI - local gyrification index

A| RIGHT HEMISPHERE



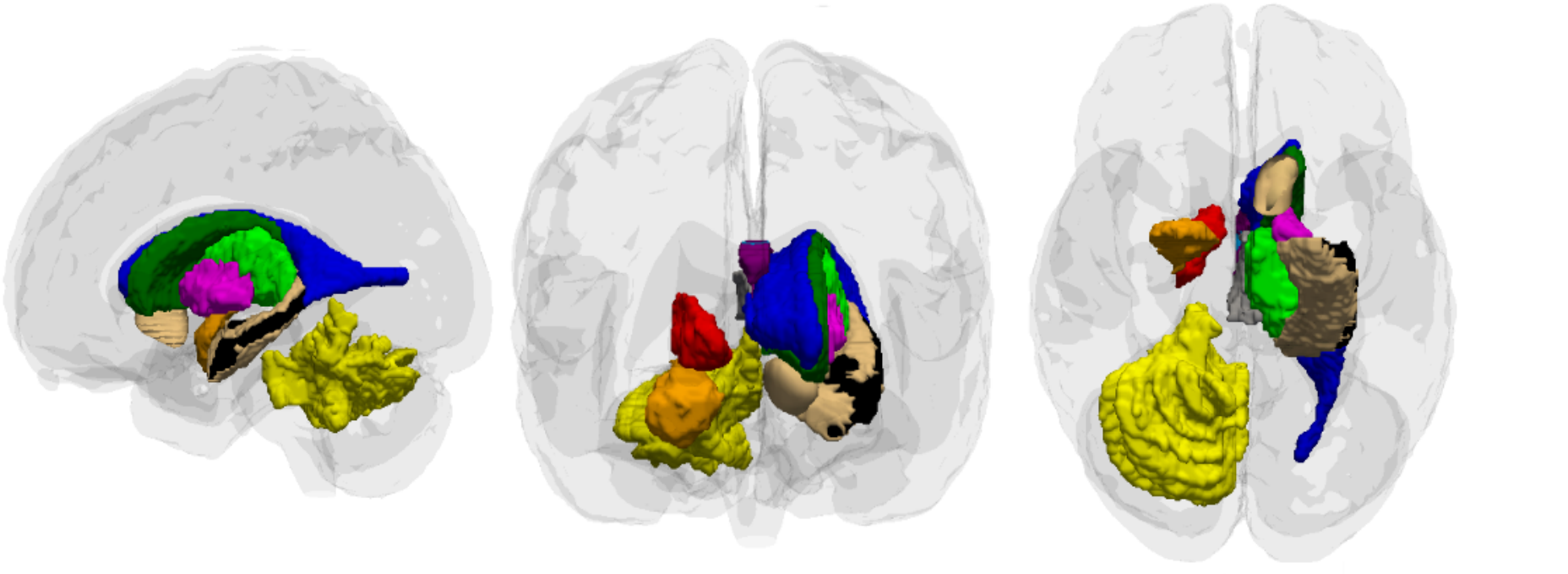
Right transverse temporal volume
Right isthmus cingulate volume
Right para-hippocampal volume
Right temporal pole LGI
Right fusiform volume
Right postcentral volume
Right para-central volume
Right superior-frontal volume
Right middle temporal curvature

B| LEFT HEMISPHERE



Left pars triangularis curvature
Left sulcal banks volume
Left supramarginal volume & thickness
Left lingual volume
Left frontal pole area & volume
Left superior parietal volume & area
Left cuneus thickness
Left precentral volume

C | OTHER REGIONAL & NON-CORTICAL sMRI MEASURES VOLUMES



D | WHITE MATTER TRACTS

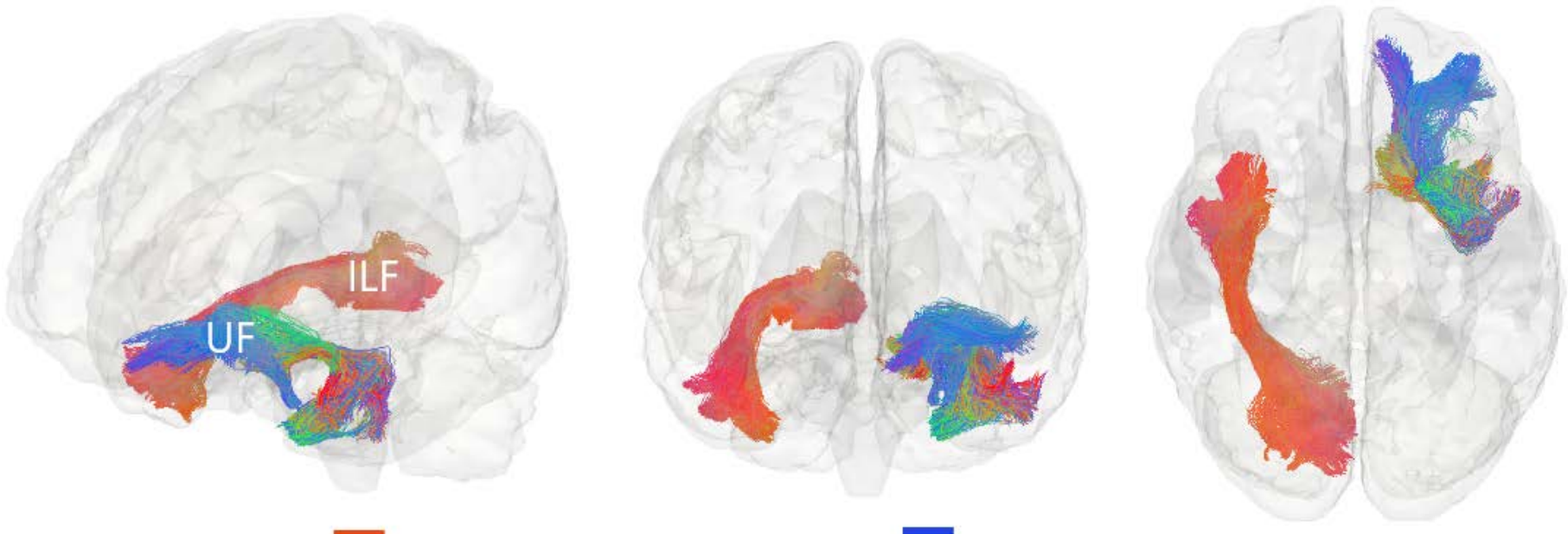


Figure 3.7| Illustration of the relevant regions selected usig SVM-RFE. The subfigures are A—sMRI cortical measures in the right hemisphere (9 features), B—sMRI cortical measures in the left hemisphere (11 features), C—other sMRI regional and non-cortical measures/volumes (15 features), and D—relevant white matter tracts (2 features). The features also include mean area and thickness of the right hemisphere regions as well as left hemisphere cerebral white matter volume (3 features).

3.4 Discussion

In summary, we found that (1) PLMs did not outperform SVMs and DTEs in predicting cognitive abilities from multimodal neuroimaging; (2) neuroimaging could predict categories of cognitive performance more accurately than it did individual cognitive scores; and (3) there was an identifiable subset of multimodal neuroimaging predictors of cognitive abilities in children. Predicting individual cognitive scores led to small training and fitting errors but high generalised errors. Overall, the regression models resulted in low to moderate out-of-sample correlations between actual and predicted scores; they did not perform significantly better than using a constant mean response of the scores to predict unseen data. Further, our hierarchical clustering analysis resulted in two main groups of overall cognitive performance, which comprised better and poorer performing children living with and without HIV. Our classification models using these poorer versus better cognitive performance categories resulted in AUC of 0.71–0.83, accuracies of 65–74%, sensitivities of 70–85%, and specificities of 65–73%, with SVMs being the best performing algorithm in terms of AUC, and PLMs having the lowest predictive performance among the 3 models. Lastly, we identified 40 relevant predictors of overall cognitive performance categories out of 489 candidate imaging measures, which included cortical and subcortical volumes and diffusivity in select white matter tracts.

Regression performance: predicting continuous cognitive scores

Previous studies have found weak to moderate associations (Pearson's r between actual and predicted value) predicting paediatric cognitive abilities from sMRI ($0.25 < r < 0.40$) (Pan et al., 2018; Yao et al., 2023), and resting-state fMRI ($0.09 < r < 0.37$) (Sripada et al., 2021; Sripada et al., 2020). In the comparison of different algorithms, Pat et al. (2022) achieved $-0.01 < r < 0.6$ with task-based fMRI and Yuan et al. (2022) using a combination of sMRI, DTI and resting-state fMRI achieved $-0.84 < r < 0.8$. We combined sMRI, ^{1}H-MRS, and DTI as potential sources of information for cognitive abilities,

since it has been shown that integrating MRI modalities boosts reliability and prediction of cognitive ability in children (Rasero et al., 2021). In line with this literature, our multimodal MRI model predictions had weak to moderate associations with actual cognitive abilities ($0.2 < r < 0.34$). However, there were differences in findings between this study and that of Pat et al. (2022) and Yuan et al. (2022) when comparing the performance of PLMs, DTEs, and SVMs. For example, Yuan and colleagues (2022) found the highest R^2 values on the validation sets among the models being 10.92% explained variance (ridge regression) and 10.74% (elastic net regression). These are both PLMs, whereas our best performing regression models were with linear SVMs, i.e., cross-validated $R^2 = 20\%$ for motor skills via PPT preferred hand, and 14% for categorical fluency score. Pat and colleagues (2022) also found the elastic net regression performing similar to or better than the other algorithms. There were possible differences between this study and theirs which may explain contrasting results, such as the use of fMRI brain measures, executive function (inhibitory response/stop-signal reaction time) prediction, and large sample sizes. Although Yuan and colleagues (2022) used multimodal MRI, the stop-signal task activation data expectedly dominated in estimating stop-signal reaction time compared to DTI and morphometric MRI data. Moreover, compared to our sample size of 132 children of which 89 was usable, both studies used the Adolescent Brain Cognitive Development (ABCD) datasets with more than 3500 subjects.

Earlier paediatric studies (Burgaleta et al., 2014; Karama et al., 2011; Khundrakpam et al., 2017; Lee et al., 2017; Ramsden et al., 2011; Short et al., 2013) have also associated neuroimaging measures and cognitive ability, however, these associations were done within-sample, i.e., not tested on unseen data. The associations were an inference analysis between a single neuroimaging variable and cognitive performance measure, and any strong correlation found may be optimistic or biased upwards. As such, findings within-sample should be interpreted with caution. In this study, we estimated both in-sample and out-of-sample errors using resubstitution validation and 10-fold cross validation

(Kohavi, 1995), respectively. Lower resubstitution errors, which we found here, tend to suggest that the models will predict unseen data poorly, since it may be clear evidence for overfitting, where a model fits perfectly to the training data, capturing "noisy" patterns in the sample, then fails to generalise (Jollans et al., 2019; Paulus et al., 2019). This is a particular problem in small sample datasets, such as ours, although it does occasionally occur in large samples as Yuan and colleagues (2022) found in their analyses. The best way to assess prediction models is how they perform on unseen data, instead of training or fitting errors. However, using cross-validation, small sample sizes tend to lead to large error bars across folds and reported prediction accuracy often decreases as sample size increases (Varoquaux, 2018). We notice this in our models which have high generalised error rates, despite the typical correlation strengths, r's, between actual and predicted scores. We also observe in the plots of predicted versus actual scores that for data points at the more extreme edges of the range, the predicted scores err towards the mean of the sample, suggesting that in the small sample there was no sufficient variability to capture the upper and lower ranges of scores. For this reason, we found that the models do not perform significantly better than prediction using a constant mean response (low validation R^2 values). Models with low and negative validation R^2 are frequently seen in machine learning studies; it indicates that the models do not generalise well (Yuan et al., 2022).

Classification performance: hierarchical clustering and prediction of cognition categories

The prediction or classification of categories of cognition using machine learning algorithms and neuroimaging data is not as widespread in the literature as predicting continuous values (e.g., Pat et al. 2022; Yuan et al., 2022). Classification tends to be disease specific, i.e., a major medical or psychiatric disorder exists and there is an established clinical cut-off criterion. For example, the classification of infant cognitive and motor impairment using logistic regression and DTI data achieving AUC, sensitivity, and

specificity greater than 86% (Schadl et al., 2018); prediction of ADHD diagnostic category in young children using a linear SVM classifier on sMRI cortical measures with 50–61.2% accuracy (Oztekin et al., 2021); and classification of autism spectrum disorder from rs-fMRI functional connectivity, achieving highest accuracy of 71% with kernel SVM (Lanka et al., 2020). Moreover, a recent study predicting various mental disorders in children achieved poor performances, the best model boosted trees could only classify ADHD (0.57 AUC, 56% accuracy) and bipolar disorder (0.55 AUC, 56% accuracy) from sMRI measures with statistically significant prediction performance (Gaus et al., 2023). This may be due in part to the need for classification to be clinically informative or diagnostic, independent of confounding variables, and assessed thoroughly for performance and generalisability (Nielsen et al., 2020). Consequently, cognitive score cut-offs where children are cognitively impaired/delayed or display ADHD tendencies or autism spectrum behaviour are established to a certain extent. But it is more challenging to define categories of cognition in the general population where there is no neurodevelopmental disorder, and the risk is the resulting categories being somewhat arbitrary.

Some cognitive assessment batteries have cut-off scores that indicate a threshold for diagnosis of developmental delay or cognitive impairment, or degrees of cognitive deficit. For example, negative values for the attention performance index on the TOVA are associated with ADHD (Leark et al., 2004), and a value of less than 70 on the full scale intelligence quotient of the Wechsler intelligence scale for children (WISC) is a cut-off for intellectual deficiency (Na & Burns, 2016). Other batteries like KABC (Kaufman & Kaufman, 2014) and Beery-VMI (Beery & Beery, 2004) have descriptive categories, i.e., whether the child's score lies in the low, average, upper average, or high range in the distribution of raw scores among his or her peers of the same age—based on normative data validated extensively in the United States and other high-income settings.

These categories may not translate to the South African context or low-SES households and communities (Harris et al., 2021; Mitchell et al., 2017). Because we wished to identify

categories of generally high and low performers across all tests rather than performance on individual tests, we implemented hierarchical clustering. Although creating artificial categories out of continuous data is not ideal because information is lost, predicting continuous scores is difficult and the cognitive data can be subject to unwanted variations in scoring, e.g., mood and fatigue. Creating categories by clustering across all the available testing dimensions may be more robust to this effect, and to our knowledge this has not been previously done in paediatric HIV. However, Paul and colleagues (2022) implemented a similar approach by using hierarchical clustering to divide their adult HIV cohort into 5 cognitive phenotypes and used ensemble methods to classify the resulting categories of performance with reasonable accuracies. We found 2 main categories in our sample, "poorer performers" and "better performers". For comparison to the external normative data, the poorer performers were characterized by low to below average scores in attention (closer to the ADHD sample), learning, and language, while the better performers had average to above average scores compared to their peers at the same age (Beery & Beery, 2004; Kaufman & Kaufman, 2014; Leark et al., 2004).

By reducing the regression problem treating each domain separately into a classification problem encompassing all domains simultaneously, we could predict these categories using multimodal MRI and machine learning algorithms with performance metrics as high as 0.83 AUC, 74% accuracy, 85% sensitivity, and 73% specificity. This classification performance is similar to that achieved for predicting risk of cognitive impairment (Schadl et al., 2018), ADHD categories (Oztekin et al., 2021), autism spectrum disorder (Lanka et al., 2020), and developmental delay (Ouyang et al., 2020). However, these studies are different to ours in that they used individual modalities, so were not multimodal in nature, and generally had larger sample sizes (102−988 children). Further, the predicted categories in these studies were based on, for example, a clinician determining diagnosis of ADHD and autism from symptoms, behaviour disrupting scales, and parent

interviews, while our approach implemented a data-driven clustering on multiple cognitive domains to help us "diagnose" poorer and better cognitive performance.

Theoretical virtues of PLMs, SVMs, and DTEs

In the classification models using both children living with and without HIV, SVMs (0.83 AUC and 74% accuracy) and DTEs (0.78 AUC and 71% accuracy) perform on par, with SVMs having a slight edge, while PLMs (0.71 AUC and 65% accuracy) tended to have the lowest performances. Unlike PLMs, SVMs and DTEs can model the relationship between response and predictions non-linearly as well as consider the interactions amongst them (Deshmukh et al., 2020; Scholkopf & Smola, 2018). Since linear SVMs led to the best models, non-linearity between categories of cognition and multimodal MRI variables cannot explain the discrepancy. A possible reason may be the ability of SVMs and DTEs to consider interaction effects between features, while PLMs like lasso have serious limitations in this respect (Jain & Xu, 2021). Nevertheless, PLMs have the benefit of being easier to interpret (Molnar, 2020), for example, researchers can directly measure the importance of each predictor by interpreting its weight or coefficient in the model. Moreover, we noted that PLMs, SVMs, and DTEs did not have statistically different prediction residuals in the regression models, i.e., root mean squared and mean absolute errors, but how closely the predicted values followed the trend of the actual values as measured by Pearson's r was dissimilar. Linear SVMs and bagged DTEs tended to have stronger correlation strengths than PLMs (typically lasso regression models). The modelling of interactions effects may likewise explain this difference in performance. It was, however, surprising to find PLMs performing poorly, given other studies in children have reported good regression performances using ridge, lasso, and elastic net regressions on fMRI data to predict cognitive outcomes (Pat et al., 2022; Yuan et al., 2022). We contend using fMRI may have driven either the good performance (Yuan et al., 2022) or the performance on a par with SVMs and DTEs (Pat et al., 2022). This is

further evident in the literature where regression models are improved by using fMRI measures (Vieira et al., 2022).

We implemented Bayesian optimization for finding hyperparameters for each model, which allowed the training of multiple models at the same time and selecting the best PLM, SVM, and DTE algorithms to compare. Among the PLMs, lasso regression models (Tibshirani, 1996) tended to perform better than the other penalisation techniques when it came to classification. We suspect the reason to be that ridge regression is superior if there are many parameters of about the same value influencing the response, elastic net regression is very useful when there are many correlated predictors and leads to a grouping effect for the highly correlated variables, while lasso regression works when there is truly a significant subset of predictors impacting the target (Greenwood et al., 2020). As previously mentioned, the linear SVM kernel performed better than gaussian, quadratic, and cubic kernels. This suggests the data is best separated by a linear decision boundary rather than a non-linear one, therefore the data did not need to be transformed (Scholkopf & Smola, 2018; Schölkopf et al., 2002). For DTEs, random under-sampling boosted (RUSBoost) classifiers performed best, however not significantly better than other bagged or boosted trees. This may be because our categories memberships were not perfectly balanced, and RUSBoost is designed specifically to handle category imbalances.

Additionally, another important metric to consider when choosing between these models is the performance of feature selection (Khaire & Dhanalakshmi, 2022; Mwangi et al., 2014). PLMs and DTEs integrate feature selection in the model building process while SVMs require training with lasso regression to achieve embedded feature selection or repeated training to remove the lowest weighted measures until the optimal subset is obtained. We found the latter technique, SVM-recursive feature elimination, to be very computationally expensive. However, it was successful in identifying important features in the model. More neuroimaging studies have used embedded feature selection than

SVM-recursive feature elimination because it yields the largest improvement in classification (Khaire & Dhanalakshmi, 2022; Tohka et al., 2016).

Feature selection: implications of the brain-cognition biomarkers

Our multimodal predictors of overall cognitive performance via classification consisted mainly of cortical and subcortical volumes (29 out of the 40 relevant predictors). Regional volumes and cortical curvatures have previously been linked with fluid intelligence (Li et al., 2019). Volume of the bilateral globus pallidum is one of the top multimodal predictors in our classifiers. Bilateral globus pallidus volume has been positively correlated with disease severity in children, adolescents, and adults with ADHD (Agoalikum et al., 2023). Notably, our cluster of poorer performers had attention scores closer to the ADHD sample. Globus pallidus controls conscious and proprioceptive movements, its dysfunction has been linked to alcohol and opiate abuse, and studies have demonstrated its involvement in numerous disorders, including ADHD (Frodl & Skokauskas, 2012).

We also found volumes of the left precentral gyrus, a region responsible for controlling voluntary motor movement on the body's contralateral side (Banker & Tadi, 2023), and has been linked to neurodevelopmental differences in child and adult numerical cognition (Skagenholt et al., 2021). Regional cortical morphometric measures like curvature, thickness, and surface area were also predictors in our models. This is in line with previous studies which have successfully used cortical morphology to predict cognitive performance (e.g., Yao et al., 2023). An earlier study showed that more intelligent children develop particularly plastic cortices, and their expression of intelligence is dynamic (Shaw et al., 2006). A more recent study (Wu et al., 2022) using the ABCD dataset has shown that cortical and subcortical morphology, e.g., temporal, parietal, and cingulate cortex surface area predicts fluid intelligence with reasonable accuracy ($r = 0.31$). Diffusivity measures in white matter tracts and metabolite concentrations are rarely associated with cognition in both paediatric and adult literature (Vieira et al., 2022). There was diffusivity in 2 white matter tracts—left inferior longitudinal fasciculus and the left

uncinate fasciculus (UF)—present in our prediction models. Previous studies have associated cognitive performance of children, adolescents, and adults with diffusivity measures in the inferior longitudinal fasciculus, bilateral superior longitudinal fasciculus, left cingulum, and corticospinal tract (Hutton et al., 2020; Raja et al., 2022). White matter integrity of the UF has also been associated with children's verbal memory proficiency, memory retrieval, reward and valuation computation (Olson et al., 2015). Both tracts were in the left hemisphere which is typically responsible for language and learning. This suggests that white matter integrity and microstructural changes play a role in cognitive performance. Finally, we did not find any MR spectroscopic predictors as relevant predictors of cognition, although a study in adults has provided evidence that concentrations of glutamate, creatine, and N-acetyl-aspartate in the cingulate cortex are associated with verbal fluency, episodic memory, and other aspects of cognitive performance in adults (Oeltzschner et al., 2019), also demonstrating the importance of inclusion in predictive models.

Limitations and recommendations for future research

The number of children who completed all cognitive assessments and had good quality multimodal MRI datasets was 89, which although larger than the median sample size in neuroimaging studies, is much less than the recommended amount for good prediction and replicability (Marek et al., 2022). Sample size considerations limited the inclusion of other imaging predictors which were available for some of the sample, namely resting-state fMRI and ^{1}H-MRS in basal ganglia and peritrigonal white matter, which may have been able to capture other neural representations of cognitive performance. PLMs also failed to converge to a solution when the children living with and without HIV were analyses separately due to a low sample. It is a challenge to acquire good quality MRI scans in children, because of their inability to lie still for extended periods of time.

Non-imaging predictors included in this study are limited to age at scan, sex, birthweight, and HIV status (in the analyses across all subjects). Linn and colleagues (2016) argued

that non-imaging confounders ought to be incorporated as predictors in the model instead of being estimated jointly with the imaging variables. These non-imaging variables were included as candidate predictors despite not being confounders. Age being a non-confounding variable was not surprising as large cohort studies have also found limited evidence for interaction effects between HIV and age on cognitive performance in the adult HIV literature (Valcour et al., 2011). However, sex distribution has been previously reported to contribute to the frailty observed in adult HIV (RH. Paul et al., 2018). In our paediatric cohort, neither age nor HIV status were associated with the individual scores nor significantly different between poorer and better performing groups of children. The regression and classification models for children living with and without HIV resulted in poor performances, as such we could not compare how brain-cognition relationships differed between these groups.

Socioeconomic status variables were not, however, included as confounders. A recent cross-sectional study of 11,875 children aged between 9 and 10 found an association between neighbourhood poverty, prefrontal and hippocampal volume, and cognitive performance (Taylor et al., 2020). As such, we recognise that objective measures of SES such as household income may predict both brain development and cognitive performance and are therefore worth including in future regression or classification models.

Similarly, behavioural problems or personal-social cognition, which have been shown to contribute substantially to academic achievement and development as well as carry more implications for interventions in children especially in the context of HIV and low-SES communities (Cockcroft & Cassimjee, 2020; Gabbidon et al., 2020) were not assessed in this study. Future studies may benefit in including them in the calculation of overall cognitive performance. Although information is lost in the creation of categories from continuous data, the use of clustering to represent overall cognitive performance is worth further exploration, especially when clustering algorithms like agglomerative hierarchical clustering often yields local minima based on the sample characteristics and not

necessarily the best solution overall. It may be of interest to compare clusters identified using the hierarchical clustering method with those resulting from other algorithms such k-Means, k-Medoids, k-Nearest neighbour search and radius search clustering (Saxena et al., 2017), or spectral clustering (Tremblay & Loukas, 2020).

Only a single, non-nested cross validation (CV) was used for Bayesian optimization of the model hyperparameters and measuring generalisation performance, while other studies have implemented nested CV to avoid overfitting during training. In nested CV, hyperparameter selection is performed in an inner cross-validation loop which prevents data leak into the models. However, it is much more computationally expensive than non-nested CV. Although the performance measures may therefore be optimistic, a recent study has demonstrated that using in practice nested CV is probably not greatly advantageous over the less computationally costly non-nested CV when selecting amongst models that have fewer hyperparameters to tune (Wainer & Cawley, 2021).

Lastly, our study is limited to neuroimaging measures at age 7, while changes in brain measures with maturation may be more indicative of cognitive performance than static measures at a single timepoint. Interactions between brain regions should also be considered, as measures of connectivity have been found to be indicative of performance in fMRI studies (Pat et al., 2022; Yuan et al., 2022). Another consideration is predicting future cognition instead of prediction at the same age, which has been achieved with high levels of accuracy over a 2-year period in some studies (Adeli et al., 2019; Girault et al., 2019; Khundrakpam et al., 2022; Ouyang et al., 2020; Ullman et al., 2014). In the next chapter, we investigate the prediction of cognitive performance 2 years into the future using earlier multimodal MRI neuroimaging and cognitive scores. This is very important, since the ability to predict future cognitive performance can be used for implementation of early targeted interventions in children with cognitive difficulties.

4

Multimodal magnetic resonance neuroimaging at age 7 predicts categories of overall cognitive performance at age 9 better than cognitive testing

Chapter 4:

- Is it possible to predict cognitive performance at age 9 using neuroimaging and/or cognitive data from age 7?
- Is multimodal MRI better than cognitive scores at predicting subsequent cognitive performance?
- Does using current and historic multimodal MRI (at 2 timepoints) improve prediction of cognitive performance?

Abstract

It is important to understand the relationship between neuroimaging measures and future cognitive outcomes of children from low-socioeconomic status (SES) backgrounds to implement appropriate interventions. However, in order to be of value, neuroimaging needs to provide information that is not readily obtainable by less expensive means. We therefore aimed to (1) predict future cognitive performance in children by using Bayesian optimised, cross-validated (CV) support vector machines and support vector regression (SVMs and SVR) on a combination of structural magnetic resonance imaging (sMRI), diffusion tensor imaging (DTI), and proton magnetic resonance spectroscopy (^{1}H-MRS), and (2) determine whether neuroimaging offers any predictive advantage compared to predicting future performance using earlier cognitive scores.

Our study population comprised 132 children from low-SES communities, who were assessed at ages 7 and 9 years. Cognitive performance in the domains of learning, language, visual-motor integration, attention, and executive function was assessed using the

Kaufman assessment battery for children (KABC), Beery-Buktenica developmental test of visual-motor integrations (Beery-VMI), test of variables of attention (TOVA), Purdue pegboard test (PPT), and the Peabody picture vocabulary test (PPVT) at ages 7 and 9 years. Hierarchical clustering was implemented to group the children into poorer and better performance categories at age 9. Accuracy, specificity, sensitivity, and area under the receiver operating characteristic curve (AUC) were obtained for the 10-fold CV SVM classification models. For prediction of continuous scores, regression performance was assessed via 10-fold CV SVR errors, coefficient of determination (R^2), and Pearson's r between actual and predicted values.

We could classify poorer and better performers at age 9 with 70–77% accuracy, 70–81% specificity, 71–79% sensitivity, and 0.75–0.81 AUC using multimodal MRI and cognitive scores at age 7. Further, we could predict scores for auditory working memory (R^2 = 0.75, r = 0.45), short-term memory (R^2 = 0.43, r = 0.62), visual-motor integration (R^2 = 0.26, r = 0.39), and executive reasoning (R^2 = 0.33, r = 0.27) with moderate to strong Pearson's r. Cortical morphology, subcortical volumes, neurometabolite concentrations, and diffusivity in select white matter tracts were relevant in the prediction model, suggesting that multiple MRI modalities and features should be considered simultaneously to establish correlates of overall cognitive performance.

4.1 Introduction

Discovering neural correlates of cognitive processes, such as problem-solving, attentional control, decision-making, working memory, language, and general intellectual functioning can help improve neurodevelopmental outcomes of children with cognitive difficulties by (i) identifying indicators, risk factors, or biomarkers of cognitive impairment, (ii) understanding the developmental course of cognitive plasticity, and (iii) enabling the development of targeted interventions that can modulate brain networks or regions associated with cognitive functions (Buzzell et al., 2020; DeCicco et al., 2012; Meyer et al., 2014). Therefore, it is very important to understand relationships between cognitive abilities and neuroimaging measures as early as possible during childhood development. Previous studies in children have found associations ranging from weak to strong between cognitive skills and regional cortical thickness (Bathelt et al., 2018; Brito et al., 2017; Habeck et al., 2020), cortical gyrification and sulcal morphology (Chung et al., 2017; Gregory et al., 2016; Tissier et al., 2018), regional and subcortical volumes (Du et al., 2022; Taylor et al., 2020; Ziegler et al., 2013), fractional anisotropy and mean diffusivity (Farah et al., 2020; Hutton et al., 2020; Raja et al., 2022), and neurometabolite concentrations (Oeltzschner et al., 2019; van Dalen et al., 2016). Many of these efforts have relied on univariate statistical inference techniques, such as correlational analysis and generalised linear models, to find relationships between individual brain measures and cognition measured cross-sectionally at the same timepoint. However, predictive modelling, where models are trained using a machine learning algorithm on combinations of current and/or historic neuroimaging measures to estimate unseen data/future cognitive outcomes, may reveal further insights into the neurobiological basis of cognition, as well as enable earlier identification of children who may benefit from intervention. There is therefore potential to use advances in interpretable machine learning algorithms and multimodal neuroimaging in our endeavour to improve neurodevelopmental outcomes of children worldwide.

In recent years, several studies have performed multivariate predictive modelling with various cross-validated learning algorithms and longitudinal neuroimaging in children. For example, predicting over a 2-year period: language scales using multi-task multi-linear regression on cortical thickness, mean curvature, and sulcal depth (Adeli et al., 2019); executive function using random forest regression on cortical thickness measures (Khundrakpam et al., 2022); as well as general intellectual functioning and deep learning on white matter connectomes at birth (Girault et al., 2019). Support vector machine/regression (SVM/R) is one type of supervised learning algorithm (Cortes & Vapnik, 1995), known for its utility in solving classification and regression problems. The goal of SVM/R is to find hyperplanes in n-dimensional space (n is number of predictors/features) with the largest margin, i.e., the distance between the hyperplane and closest data points from each category. The hyperplane maximally separates the categories in the training dataset and can be used to determine how new data points can be categorised (Scholkopf & Smola, 2018; Schölkopf et al., 2002). In the literature, SVM/R models have been used to predict paediatric cognitive performance at least 2 years into the future with moderate accuracies (Ouyang et al., 2020; Ullman et al., 2014). Ouyang and colleagues (2020) could predict cognitive behaviour of neonates at age 2 from whole-brain white matter cortical microstructure at birth. Although the study used a single neuroimaging modality, it was possible to predict developmental delay with reasonable accuracies between 60 and 77%. The earlier study by Ullman et al. (2014) employed a combination of structural and functional brain measures of 6 to 20-year-olds to predict scores of reasoning and working memory, achieving correlations of $0.29 < r < 0.64$ between actual and predicted scores.

In our study, we aimed to use SVM/R and multimodal MRI neuroimaging collected at two timepoints to predict future cognitive scores as well as categories of cognition ("poorer performers" versus "better performers") of children from low-socioeconomic status (SES) backgrounds. We acquired structural MRI (sMRI) volumes, thickness,

gyrification, area, and curvature; diffusion tensor imaging (DTI) fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity; as well as magnetic resonance spectroscopy (^{1}H-MRS) concentrations of choline, N-acetyl aspartate, creatine, and other neurometabolites at ages 7 and 9 years. Cognitive abilities in the domains of learning, planning, executive function, attention, motor coordination, language, and visual-motor integration were assessed at both ages using a battery of standardised tests; these covered all the domains that share a bidirectional link with academic skills (Lovden et al., 2020; Peng & Kievit, 2020).

To directly address our aim, we asked the following questions: (1) Is it possible to predict cognitive performance at age 9 using only neuroimaging and/or cognitive data from age 7? (2) Is multimodal MRI neuroimaging better than historic cognitive scores at predicting subsequent cognitive performance? (3) Does using current and historic multimodal MRI neuroimaging, i.e., at 2 timepoints, improve prediction of cognitive performance?

4.2 Material and methods

4.2.1. Study population

The study included 132 children (demographics: supplementary Table B.1) from low-socioeconomic status communities in Cape Town, South Africa, who were part of a follow-up longitudinal neuroimaging and cognitive study after the conclusion of the children with early antiretroviral therapy (CHER)[2] trial in 2011. The longitudinal follow-up study enrolled children with perinatally-acquired HIV from the CHER trial, as well as HIV negative, age, and community-matched children; some from a connected pneumococcal vaccine trial (Madhi et al., 2010) and others that were involved in a CHER neurodevelopmental sub-study (Laughton et al., 2018). Eligibility criteria included: documented HIV status of child, mother, or caregiver; child's home language either isiXhosa, English, or Afrikaans; no history of head injuries, birth complications, learning disability or major psychiatric disorder that may affect cognitive performance; and child's birthweight $>$ 2000g.

A subset of 69 participants in the current study were therefore virally suppressed, asymptomatic children living with perinatal HIV, and the other 63 participants were children without HIV. This is reflective of the fact that low-SES communities in South Africa are disproportionately affected by HIV (Cleary et al., 2011). We have previously reported on potential multimodal neuroimaging biomarkers that can distinguish children living with HIV from those without HIV (Khobo et al., 2022), and van Wyhe and colleagues (2021) have presented group-wise comparisons of the children's cognitive data. Based on the analyses performed in the previous chapter on the children living with and without HIV separately, for the purpose of predicting cognitive outcomes from multimodal neuroimaging, the cohort was treated as a single homogenous group without considering

[2] The CHER randomised trial found that early treatment reduced mortality and HIV disease progression (Cotton et al., 2013; Laughton et al., 2012; Violari et al., 2008)

HIV status (except whether it is a confounding variable or a predictor of cognitive performance).

4.2.2. Data acquisition

The study was approved by the Human Research Ethics Committee of the University of Cape Town (HREC ref: 453/2021). The children were familiarised with the imaging process with a mock scanner and were given practice tests before the cognitive assessments. Written consent was given by parents or guardians, while oral assent was granted by the children.

Magnetic resonance neuroimaging

The children were scanned at the Cape Universities Brain Imaging Centre (CUBIC) at age 7 $(7.01 - 7.70\ years, mean\ age \pm\ sd = 7.22\ \pm 0.13\ years)$ on a 3 Tesla Magnetom Allegra with a single channel head coil and at age 9 $(8.99 - 10.31\ years, mean\ age \pm\ sd = 9.42\ \pm 0.41\ years)$ on a 3T Skyra MRI scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. We performed neuroimaging without sedation with single sessions lasting under 60 minutes.

The protocol on both scanners included a high-resolution T_1-weighted multi-echo magnetization prepared rapid gradient echo (MEMPRAGE) sequence (van der Kouwe *et al.,* 2008), a twice refocused spin-echo (TRSE) sequence to acquire 2 diffusion-weighted (DW) datasets, and point resolved spectroscopy (PRESS) sequence for single voxel spectroscopy (SVS) data acquisition, as well as chemical shift selective (CHESS) (Haase *et al.*, 1985) water suppression. On the Allegra, the SVS (Hess et al., 2011) and DTI (Alhamud et al., 2012) sequences included a 3-dimensional echo planar imaging (EPI) navigator for prospective motion and shim correction throughout the acquisitions. The protocol parameters at each age are provided in Table 4.1.

Table 4.1| Protocol parameters for MEMPRAGE, TRSE, and PRESS on 3T Allegra scanner at age 7 years and 3T Skyra scanner at age 9 years

Sequence	7 years (3T Allegra)	9 years (3T Skyra)
T1 MEMPRAGE	TR = 2530ms, TI = 1160ms, TEs = [1.53, 3.19, 4.86, 6.53] ms, FOV = 224×224×144 mm^3, voxel size = 1.3×1.0×1.0 mm^3, flip angle = 7°, bandwidth = 657 Hz/px, matrix size (resolution) = 224×168 144 slices	TR = 2530ms, TI = 1100ms, TEs = [1.69, 3.54, 5.39, 7.24] ms, FOV = 224×224×176 mm^3, voxel size = 1.0×1.0×1.0 mm^3, flip angle = 7°, bandwidth = 660 Hz/px, matrix size (resolution) = 224×224 176 slices
DTI TRSE	10000ms TR, 86ms TE, 244×244×144 mm^3 FOV, 2×2×2 mm^3 voxel size, 1940 Hz/px bandwidth, 112×112 matrix size, 72 transversal slices, 4 non-DW b_0 = 0 s/mm^2, b_1 = 1000 s/mm^2, 30 diffusion directions, opposite phase encodings (anterior-posterior and vice versa, AP-PA)	9800ms TR, 92ms TE, 244×244×148 mm^3 FOV, 2×2×2 mm^3 voxel size, 1518 Hz/px bandwidth, 122×122 matrix size, 74 transversal slices, 4 non-DW b_0 = 0 s/mm^2, b_1 = 1000 s/mm^2, 30 diffusion directions, Phase encodings: AP-PA
MFGM SVS PRESS	2000ms TR 30ms TE 15×15×15 mm^3 voxel size 1000 Hz bandwidth 64 averages Vector size 512	2000ms TR 30ms TE 15×15×15 mm^3 voxel size 1300 Hz bandwidth 64 averages Vector size 1024

Sequences
MEMPRAGE - multi-echo magnetization prepared rapid gradient echo, TRSE - twice refocused spin-echo, PRESS - point resolved spectroscopy
Acquisition parameters
TR - time of repetition, TI - time of inversion, TE - time of echo, FOV - field of view, b_0/b_1 (b-values) - reflects strengths or degree of diffusion weighting and used to generate diffusion-weighted images.

Cognitive testing

Cognitive assessment for this cohort at both ages has been previously described by van Wyhe and colleagues (2021). Trained research assistants supervised by a licensed psychologist obtained 11 cognitive scores from 6 cognitive batteries for each child around his or her 7th and 9th birthdays, namely:

Kaufmann assessment battery for children–second edition (KABC-II),

1) learning – short-term memory score
2) Planning – executive reasoning score
3) simultaneous processing – visuospatial processing and problem-solving score
4) sequential processing – auditory working memory score (Kaufman & Kaufman, 2014)

Test of variables of attention (TOVA),

5) attention performance index (API) – a score for how similar the attention performance is to the ADHD profile – values closer and above zero indicate normal attention and negative values indicate poor attention or attention comparable to that of the ADHD sample (Leark et al., 2004)

Semantic fluency test (SFT),

6) category fluency (component of executive functioning) – a score of verbal generativity which is assessed via animal naming (Spreen & Strauss, 2006)

Beery-Buktenica developmental test of visual-motor integrations (Berry-VMI),

7) Visual-motor integration score – a measure for how individuals can integrate their visual and motor abilities (Beery & Beery, 2004)

Peabody picture vocabulary test–fourth edition (PPVT-IV),

8) receptive vocabulary score (Dunn & Dunn, 2007)

Purdue pegboard test (PPT) – scores assessing hand dexterity,

9) preferred hand score

10) non-preferred hand score

11) both hands score (Spreen & Strauss, 2006).

We used scaled scores for KABC, TOVA, Beery-VMI, which represent where a child's raw score is in the distribution of raw scores among his or her peers at the same age where the mean is 10 and the standard deviation is 3. For PPT, SFT and PPVT scores we used raw scores as substantive normative data was not available.

4.2.3. Creation of categories reflecting overall cognitive performance

For classification purposes, we sought to find natural clusters or groups of cognitive performance at age 9 for this cohort, not only to identify children that score similarly, but also to identify the lowest and highest scorers in overall cognitive performance. To achieve this objective, we implemented an agglomerative (bottom-up) hierarchical clustering algorithm (HCA) on the 11-dimensional cognitive score vectors, consisting of the 11 scores at age 9 for each child.

Agglomerative HCA starts with each child in his or her own cluster. Pairs of clusters with the greatest similarity are then grouped iteratively until one large cluster remains (Murtagh & Contreras, 2012, 2017). We used Euclidean distance as a similarity metric to calculate the distances between the children's cognitive score vectors, and group pairs with the smallest distances. Clusters were then linked sequentially based on their distance, to form larger clusters using complete linkage (farthest neighbour method). Complete linkage led to a monotonic cluster tree and achieved a high cophenetic correlation coefficient ($c = 0.72$) compared to the other linkage methods (centroid, median, ward, average, and single linkage), which either had lower c or non-monotonic cluster trees.

To find the most dissimilar clusters, we analysed the resulting cluster tree (dendrogram in Figure 4.1) of HCA by locating the links in the tree with the largest inconsistency coefficient (φ): the ratio of the height (δ) of the link to the average height of links that

lie below it. There were 2 optimal clusters ($\varphi = 1.15$), examination of the scores within each of the 2 clusters revealed that they consisted of better and poorer performing children respectively. The 2 clusters of overall cognitive performance were given the category names *better performers* and *poorer performers*.

In the analysis of cognitive performance in this cohort, van Wyhe and colleagues (2021) detected a main effect of time, with higher scores at age 9 than at age 7 for most outcome variables (SFT, PPT, and PPVT) for both children living with and without HIV. However, Beery-VMI and KABC's sequential processing scores had significant decreases from age 7 to 9, and TOVA had no significant change among all subjects. The increase of cognitive scores over time shows typical development, while a decrease of scores absent any limiting factors in scoring is a sign of cognitive delay or impairment. For these two cognitive scores (Beery-VMI and KABC's sequential processing), we additionally created 2 categories for all subjects: *decreased* and *did not decrease* (the scores for each child may have stayed the same or increased between age 7 and 9). We achieved this by taking the difference in scores (e.g., Beery-VMI score at age 9 minus Beery-VMI score at age 7). The children with a negative result were assigned a *decreased* label and the rest the *did not decrease* label.

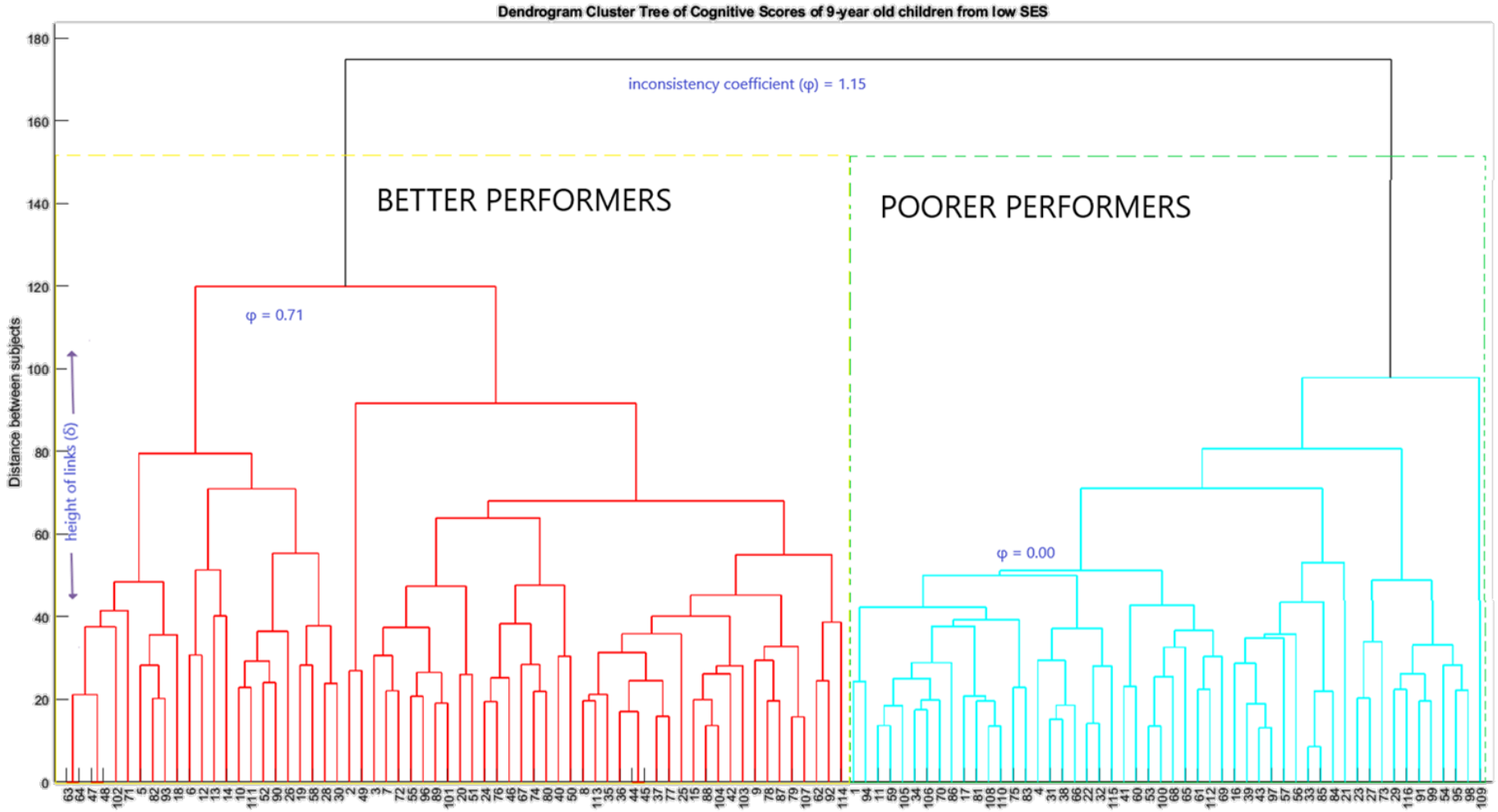


Figure 4.1| A dendrogram (cluster tree) showing the result of hierarchical clustering algorithm applied on 11-dimensional cognitive score vectors of children aged 9. Color-coded are two main clusters of poorer and better performers with a high level of consistency links.

4.2.4. Neuroimaging feature creation and image pre-processing

FreeSurfer version 6.0 software (Dale et al., 1999; Fischl et al., 2002) was used to perform cortical reconstruction and volumetric segmentation on the structural T1-weighted dataset. We calculated cortical thickness, area, mean curvature, volume, and local gyrification index (LGI) for the 34 regions of the Desikan-Killiany (DK) atlas (Desikan et al., 2006) in each cerebral hemisphere, resulting in a total of 5×(2×34) cortical structural features. Furthermore, 47 subcortical and other non-cortical regions from automated segmentation were included, including total volumes e.g., total white and gray matter volumes.

Tortoise v.2.5.2 (Pierpaoli & Walker, 2010) was used for voxel-wise calculation of the diffusion tensor as well as motion and field distortion corrections on the DW dataset. Voxel-wise diffusivity measures on the dataset were computed using AFNI software (Cox, 1996). For each of the 20 white matter tracts in the Johns Hopkins University (JHU) atlas (Mori, 2007), we calculated averaged values of fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD). This resulted in 4×20 diffusion features included as candidate predictors.

LCModel version 6.1 (Provencher, 2001) was used to calculate the concentration of 11 metabolite concentrations and their ratios to total creatine. The metabolites were glutamate (Glu), N-acetylaspartate (NAA), myo-inositol (Ins), choline (Cho), phosphocholine (PCh), total choline (GPCPCh), creatine (Cr), phosphocreatine (Cr), total creatine (CrPCr), glutamate plus glutamine (GluGln), N-acetylaspartylglutamate plus N-acetylaspartate (NAAGNAA). In summary, we generated a total of 489 (340 cortical sMRI + 47 other sMRI + 80 DTI + 22 ^{1}H-MRS) candidate imaging features at each age to use as predictors in the regression and classification models.

4.2.5. Quality control and exclusion criteria

The cognitive assessments were administered in the children's preferred languages and video recorded for quality control. We excluded children who did not complete all components of neuroimaging and cognitive assessments, and within the children living with HIV, those who had detectable viral loads (>400 copies/ml) at the age of scan to maintain a homogenous sample with no acute illness. Images with dropout slices, large motion artefacts, failed alignment to the brain templates were all excluded. Additionally, we eliminated poor-quality ^{1}H-MRS spectra (low signal-to-noise ratio, SNR < 6, and/or broad linewidth, FWHM > 0.07 ppm). After quality control, any children with missing data on any of the measures were excluded so that only children with complete neuroimaging and cognitive data at both age 7 and age 9 were included.

4.2.6. Statistical analyses

We performed the following analyses using MATLAB Statistics and Machine Learning Toolbox, Software Release R2021a (MathWorks, Inc., 2021). We used 5 candidate datasets for predictive modelling of cognitive scores at age 9. They were (i) cognitive scores at age 7—*CS7*, (ii) multimodal MRI at age 7—*MM7*, (iii) cognitive scores at age 7 combined with multimodal MRI at age 7—*CS7 + MM7*, (iv) multimodal MRI at age 9—*MM9*, (v) cognitive scores at age 7 plus multimodal MRI at ages 7 and 9—*CS7 + MM79*.

To determine whether previous cognitive data or multimodal MRI offers better prediction of future cognitive performance, we used SVM classification to predict cognitive performance category and SVR to predict each of the 11 individual scores at age 9 using each of the *CS7* and *MM7* datasets. We then used the combined dataset: *CS7 + MM7* to investigate whether historic data can offer insights on future cognitive performance. Next, we used MM9 models to explore whether current neuroimaging is better than historic data at predicting cognitive performance. Finally, *CS7 + MM79* was used to

evaluate overall whether predictive performance of cognition is improved by multimodal MRI at 2 timepoints.

Classification and regression performance

We employed SVM/R algorithm for the learning process. The goal of SVM algorithm is to find a decision boundary or hyperplane in n-dimensional space (n, number of features) where two or more categories can be distinguished, or in the case of SVR, the function between input features and target variable can be approximated (Cortes & Vapnik, 1995). During training, Bayesian optimisation (Snoek et al., 2012) as part of the machine learning toolbox was used to tune the best-point SVM/R hyperparameters: kernel function (transformation function, e.g., linear, polynomial, and gaussian, applied to input features), kernel scale (controls scaling of the predictors being passed to the kernel function), box constraint (controls trade-off between allowing training errors or misclassified points and forcing rigid parameters), and epsilon (defines margin of tolerance where the errors/misclassified points are not penalised).

For validation, we implemented k-fold cross-validation (CV) (Kohavi, 1995) to estimate how accurately the regression or classification models would perform on an independent data set. We chose k = 10 folds for our relatively small sample size to avoid overfitting and for stability (Tohka et al., 2016). Classification performance for the *better performers* versus *poorer performers* using the 5 datasets was assessed via cross-validation accuracy (percentage of correct predictions) and the area under the receiver operating characteristic curve (AUC), an overall performance measure summarising the diagnostic ability of a classifier as its discrimination threshold is varied. We also calculated sensitivity (true positive rate, TPR) and specificity (true negative rate, TNR) of the classifiers at the optimum cut-off point (point closest to true positive rate of 1 and false positive rate of 0) since there were 2 categories of overall cognitive performance, and both TPR and TPR were weighted equally, i.e., posterior probability of 0.5.

Regression performance for each of the 11 individual scores was assessed with (i) generalised error (validation root mean square error, RMSE) for each observation in the response when it was in a held-out validation fold; (ii) Pearson's correlation coefficient, r, between true and predicted responses; and (iii) coefficient of determination, comparison of residuals when prediction is achieved using the regression model (SSE) vs mean response of the scores (SST), $R^2 = 1 - \frac{SSE}{SST}$.

We performed two extra analyses: (1) Used *MM7* to predict the decrease in Beery-VMI and KABC's sequential processing, i.e., the classification of *decrease* and *did not decrease* categories, and (2) Used changes in multimodal MRI from age 7 to age 9, i.e., |*MM9*–*MM7*|, to classify *decrease* versus *did not decrease* categories. Performance was assessed as described for the classification of *poorer performers* versus *better performers*.

Feature selection

SVM-recursive feature elimination (SVM-RFE) was used to identify important predictors in the prediction models. During SVM-RFE, features are initially assigned weights from the output SVR (regression) or SVM (classification) models. The weakest coefficient is then removed, model is fitted again to determine the new weakest coefficient. The procedure is repeated until irrelevant features are removed, and the remaining feature sets achieves the same validation error or classification accuracy as the initial SVM/R model. We included age at scan, sex, birth weight, and health status (possible confounding variables) as additional candidate predictors in the regression or classification models. We illustrate how classification and regression performance are obtained with a flowchart in Figure 4.2.

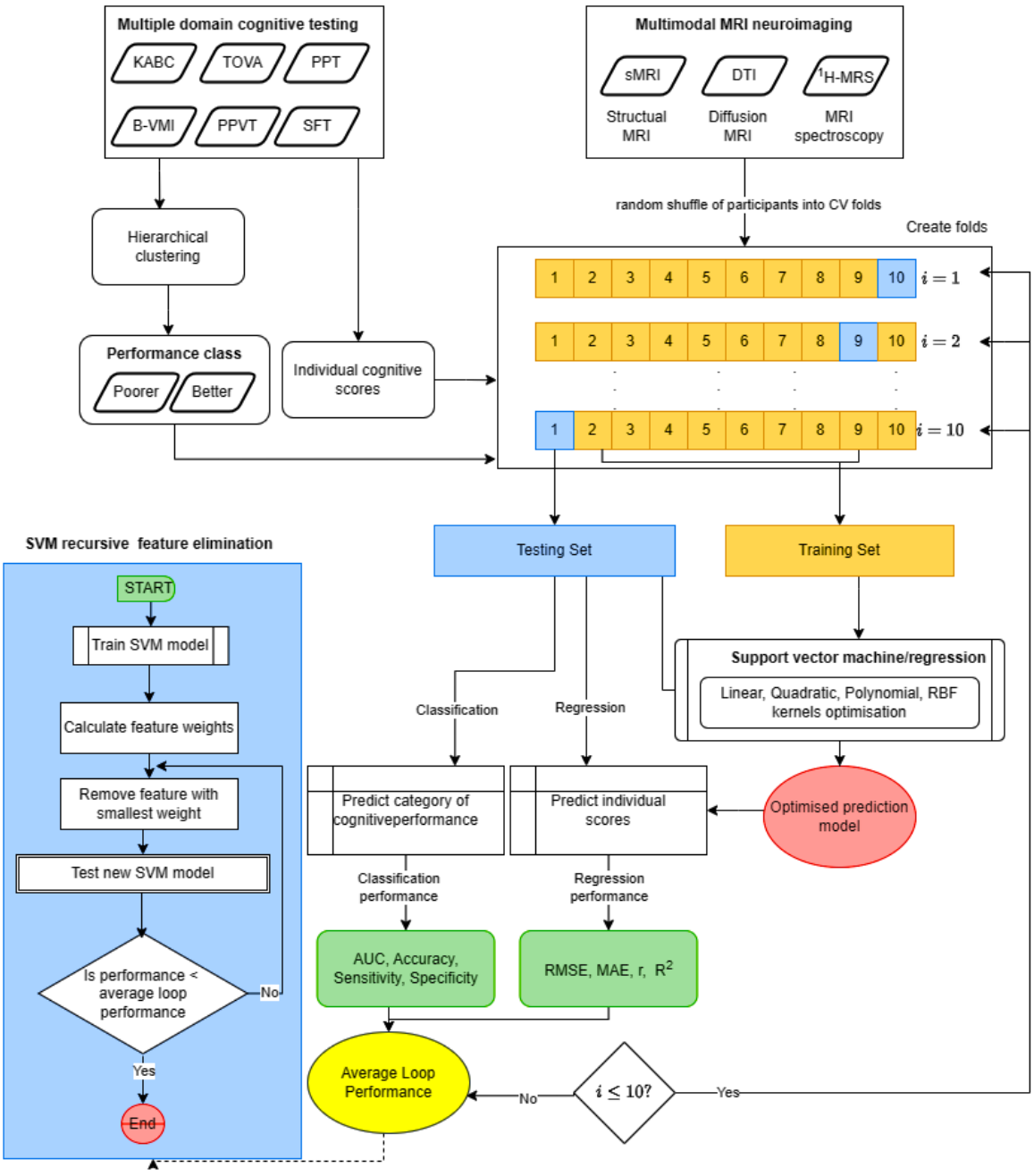


Figure 4.2| Illustration of classification and regression with 10-fold cross validation as well as feature selection used in this study.

4.3 Results

4.3.1. Final sample after exclusions

The exclusion criteria were applied to each child at both ages. Twenty children were eliminated for failure to complete all components of cognitive assessments. Seven subjects had detectable viral loads at ages of scan (5 children around age 7 years and 2 at age 9). Furthermore, eighteen children were excluded for missing one or more images due to motion artefacts (10 children) and poor spectra (8 subjects) leaving a sample of 88 children with complete neuroimaging and cognitive test data at age 7 and 9.

4.3.2. Hierarchical clustering algorithm categories

After performing HCA on the 112 children who had all the 11 scores, we identified 2 clusters of cognitive performance (dendrogram shown in Figure 4.1). The first cluster (*poorer performers*, n = 48) consisted of significantly lower scorers (p's < 0.05, Wilcoxon rank sum test) than the second cluster (*better performers*, n = 64) on measures of learning, planning, sequential processing, PPVT, SFT, TOVA API, and PPT using the non-preferred hand (Table 4.2). The two clusters had approximately equal median scores on simultaneous processing, Beery-VMI, and PPT for preferred and both hands (p's > 0.14). The distribution of scores in the two clusters is shown in Figure 4.3 for the scores that are significantly different between clusters. As can be seen there are no strong correlations between pairs of individual scores (r $< +0.5$).

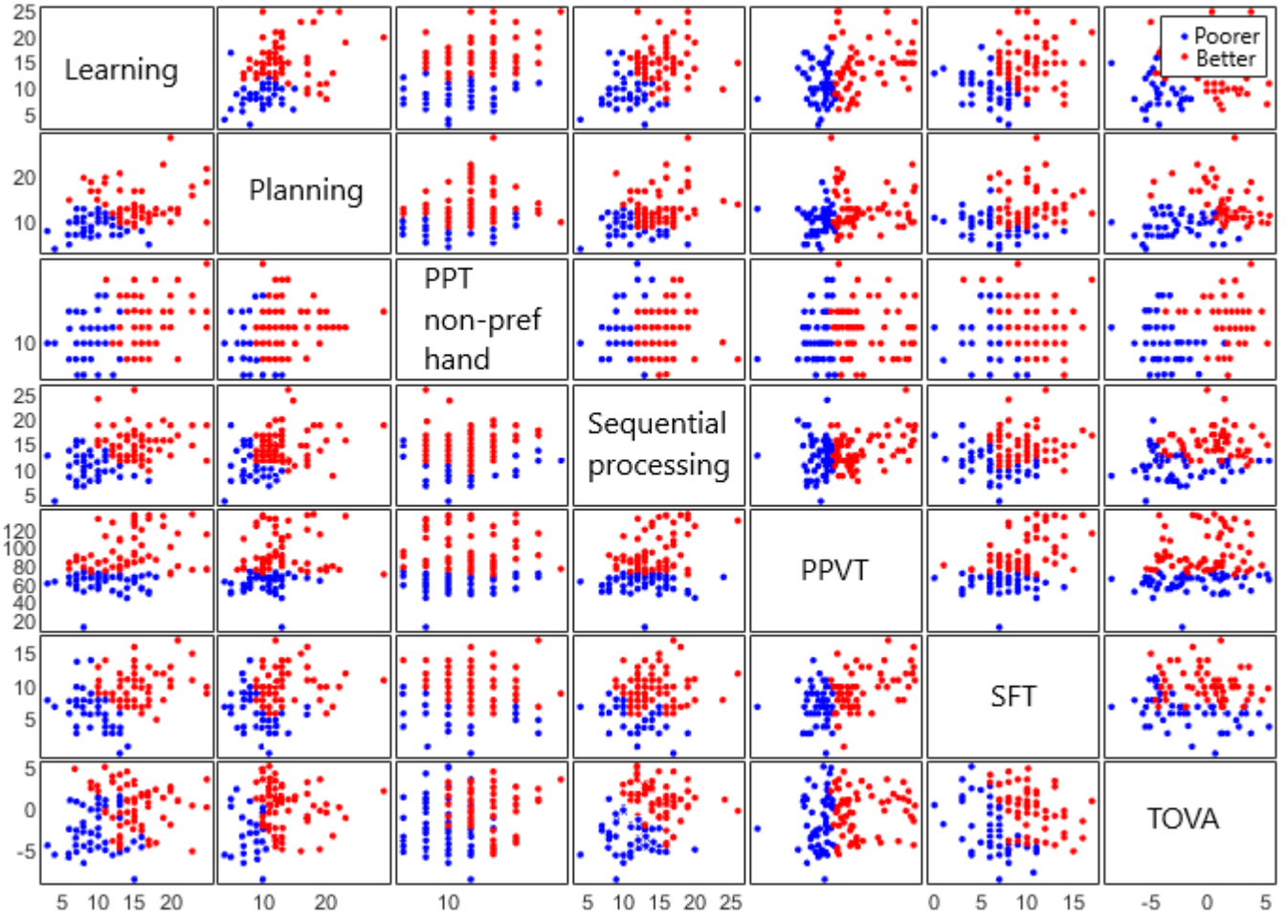


Figure 4.3| The two clusters of poorer and better performers resulting from hierarchical clustering algorithm. Illustrated are scores that are significantly different between the two clusters.

Table 4.2| Sample characteristics of the two clusters obtained via hierarchical clustering, showing *poorer performers* having lower scores than *better performers*.

	Poorer performers (n = 48)	*Better performers* (n = 64)	*p*-values from Wilcoxon rank sum test
Cognitive scores			
KABC-II (scaled)			
Learning	10 (4.50)	14 (6.50)	8×10^{-4}
Planning	10 (4.25)	12 (7.00)	6×10^{-3}
Sequential processing	12 (3.25)	14 (4.00)	7×10^{-3}
Simultaneous processing	14 (8.00)	15 (4.75)	0.14
Beery-VMI (scaled)	7 (3.00)	7 (2.00)	0.46
PPVT-IV (raw)	66 (9.75)	89 (35.50)	1×10^{-11}
SFT (raw)	8 (3.25)	10(3.75)	5×10^{-3}
TOVA API (raw)	-2.44 (5.81)	-0.05 (5.06)	7×10^{-3}
PPT (raw)			
Preferred hand	12 (3.00)	12 (2.00)	0.21
Non-preferred hand	10 (2.25)	11 (2.00)	0.02
Both hands	9 (2.00)	9 (1.75)	0.65
Demographics			
Age (years)	9.09 (0.20)	9.14 (0.13)	0.79
Birth weight (g)	3010 (488)	3200 (450)	0.15
Gender* Female n (%)	28 (58%)	29 (45%)	0.17
Status* HIV n (%)	32 (67%)	31 (48%)	0.05

*A chi-squared (χ^2) test was used to compare percentage of children between the 2 clusters
Values are median (IQR) or n (%)
We used an alpha significance level of $p = 0.05$

Demographic comparison between clusters

The clusters did not differ with respect to age, birthweight, and gender (p's > 0.17, Wilcoxon rank sum test), however *poorer performers* comprised proportionally more children living with HIV, n = 67%, compared to 48% of the *better performers* ($p = 0.054$, chi-squared test, Table 4.2).

Comparing our sample to the wider paediatric population taking the same tests around age 9, the *poorer performers* scored in the average (7–13) range and *better performers* scored in the upper average (14–16) range in the learning and sequential processing domains (Kaufman & Kaufman, 2014). The children in both clusters scored in the average category for planning, and upper average range for the simultaneous processing domain. Median Beery-VMI scores for both clusters were below average (13–19 percentiles) (Beery & Beery, 2004). The *poorer performers* had poor attention scores—their performance was very close to the ADHD sample (median TOVA API = -2.44), while *better performers* showed normal attention. We could not find normative data for PPVT, PPT and SFT scores at age 9, although the data provided by Gardner and Broman (1979) put the scores for both clusters in the below average (10–20 percentile) range.

4.3.3. Classification performance

After quality control, the *poorer performers* had n = 41 children and the *better performers* had n = 47 children. Ten-fold CV SVM classification results for the prediction of these labels using the 5 prediction datasets are shown Table 4.3. Figures 4.4A–F shows the diagnostic ability of each classifier with varying discrimination thresholds.

Using cognitive scores at age 7 (*CS7*) alone resulted in 0.75 AUC whereas using multimodal MRI at the same age (*MM7*) resulted in a higher AUC of 0.80. Using both had a marginal improvement on overall performance compared to using cognitive scores alone: 0.81 AUC, 80% sensitivity, 78% Specificity, and 77% accuracy. Using multimodal neuroimaging at 9 years (*MM9*) has slightly better AUC of 0.83, and using scores at age 7 and all multimodal MRI measures (*CS7 + MM79*) achieved the highest performance: 0.85 AUC, 80 sensitivity, 72% specificity, and 78% accuracy.

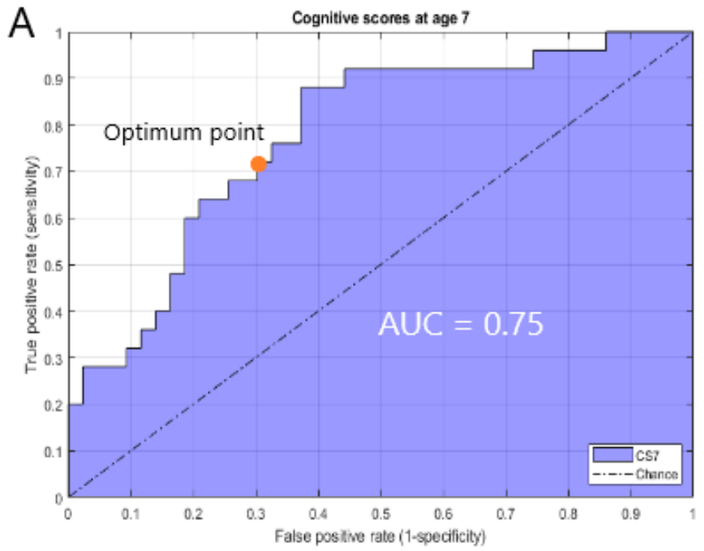
Table 4.3| Classification performance of the 5 datasets used to predict cognition categories *poorer performers* and *better performers* at age 9 years.

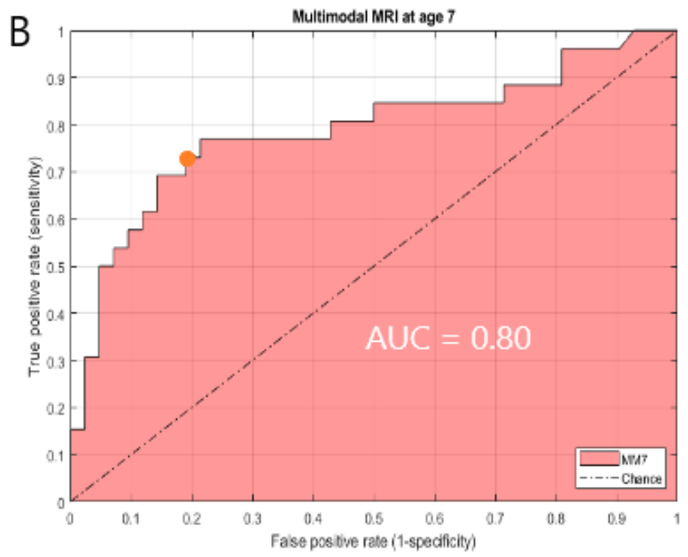
Feature Set*	AUC	Specificity (%)	Sensitivity (%)	Accuracy (%)	Balanced Accuracy (%)#	Relevant Features
CS7	0.75	70	71	70	71	2
MM7	0.80	81	72	76	77	26
CS7 + MM7	0.81	73	79	77	76	29
MM9	0.83	80	73	76	77	24
CS7 + MM79	0.85	71	80	78	76	43

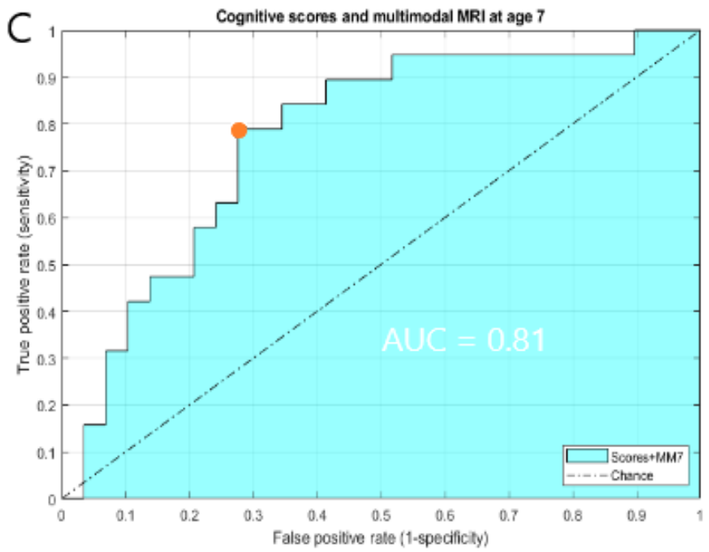
**CS7*—cognitive scores at age 7, *MM7*—multimodal MRI at age 7, *MM9*—multimodal MRI at age 9, *MM79*—multimodal MRI at both ages

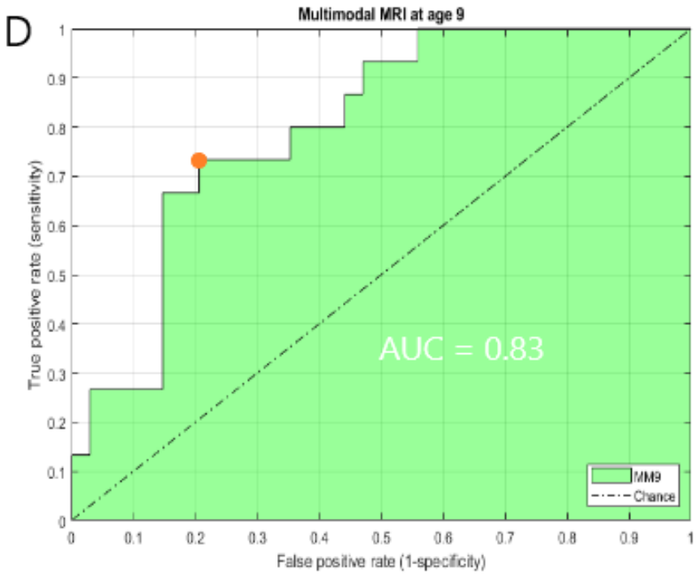
Sensitivity and specificity are evaluated at the optimum cut-off point

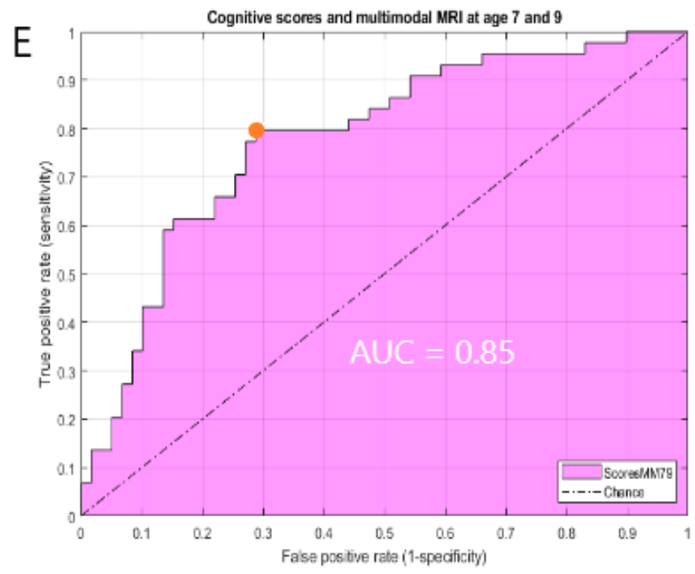
#Balanced accuracy was computed due to imbalances in category memberships. It is the arithmetic mean of sensitivity and specificity.











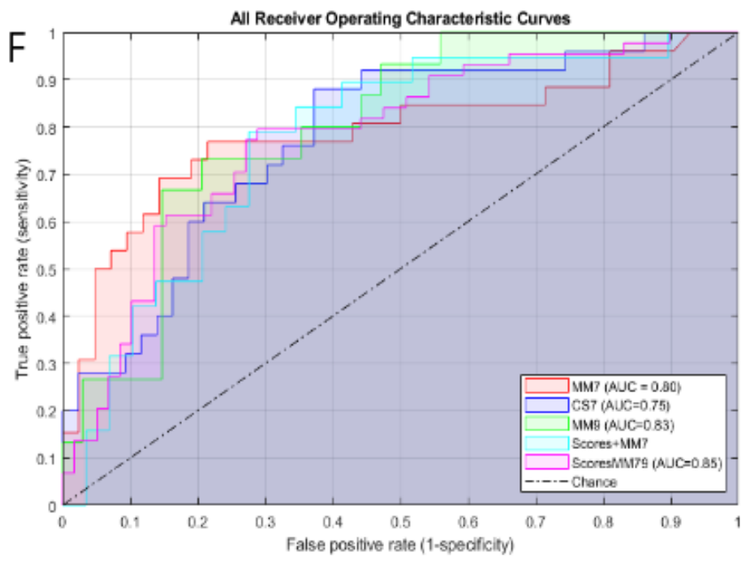


Figure 4.4| ROC curves of the prediction of cognitive scores at age 9 using (A) cognitive scores at age 7, (B) multimodal MRI at age 7, (C) cognitive scores plus multimodal scores at age 7, (D) multimodal MRI at age 9, (E) cognitive scores and multimodal MRI at ages 7 and 9, and (F) all the ROC curves A through E

4.3.4. Regression performance

The performance of SVR in predicting individual cognitive scores (regression) at age 9 with each of the 5 datasets is presented in Table 4.4. Considering for each score only the dataset with the highest regression performance, we see that the following outcomes are best predicted by cognitive scores at age 7 (*CS7*): learning ($R^2 = 0.43$, r = 0.62), sequential processing ($R^2 = 0.45$, r = 0.75), Beery-VMI ($R^2 = 0.26$, r = 0.39), and PPT preferred hand ($R^2 = 0.30$, r = 0.27). However, the regression performance was low to moderate. Only prediction of planning scores could be improved by the addition of imaging data (best predicted by *CS7 + MM7),* although with low $R^2 = 0.33$ and r = 0.27, while semantic fluency was predicted by the *CS7 + MM79* dataset, with $R^2 = 0.18$ and r = 0.44.

With regard to imaging data, *MM7* tended to perform slightly better than *MM9* in predicting scores at age 9, although both models did not perform much better than a constant mean response of the scores (all $R^2 < 0.18$). Similarly, using the *CS7 + MM79* dataset resulted in poor prediction performance ($R^2 <= 0.18$). We show the generalised (validation) errors and associations between true and predicted responses in Figures 4.5A-F for the models with moderate to strong correlation coefficients.

Table 4.4| Performance of SVR with the 5 datasets to predict individual cognitive scores at age 9. The values are coefficient of determination and Pearson's correlation coefficient between actual and predicted values (R^2|r). For each score where there is $R^2 > 10\%$ and correlation coefficient $r > +0.2$, the highest R^2 and r are given in **bold**.

	CS7 (R^2|r)	*MM7* (R^2|r)	*CS7 + MM7* (R^2|r)	*MM9* (R^2|r)	*CS7 + MM79* (R^2|r)
KABC-II (scaled)					
Learning	**0.43|0.62**	0.17|0.28	0.00|-0.32	0.07|-0.03	0.04|-0.25
Planning	0.02|0.31	0.00|-0.36	**0.33|0.27**	0.11|0.13	0.05|0.15
Sequential processing	**0.45|0.75**	0.00|-0.19	0.01|-0.29	0.12|-0.03	0.12|0.16
Simultaneous processing	0.00|0.38	0.01|-0.27	0.02|-0.05	0.15|0.25	0.00|-0.35
Beery-VMI (scaled)	**0.26|0.39**	0.00|-0.29	0.02|-0.01	0.18|0.34	0.20|0.35
PPVT-IV (raw)	0.04|-0.20	0.00|-0.43	0.05|-0.20	0.00|-0.41	0.02|-0.33
SFT (raw)	0.03|0.40	0.13|-0.02	0.00|-0.41	0.00|0.10	**0.18|0.44**
TOVA API (raw)	0.01|0.08	0.07|0.19	0.08|0.22	0.00|-0.43	**0.13|0.23**
PPT (raw)					
Preferred hand	**0.30|0.27**	0.00|-0.48	0.12|0.21	0.11|0.01	0.00|-0.31
Non-preferred hand	0.00|-0.25	0.02|-0.24	0.00|-0.36	0.01|-0.03	0.17|0.17
Both hands	0.01|0.31	0.00|-0.37	0.07|-0.05	0.14|0.06	0.00|-0.30

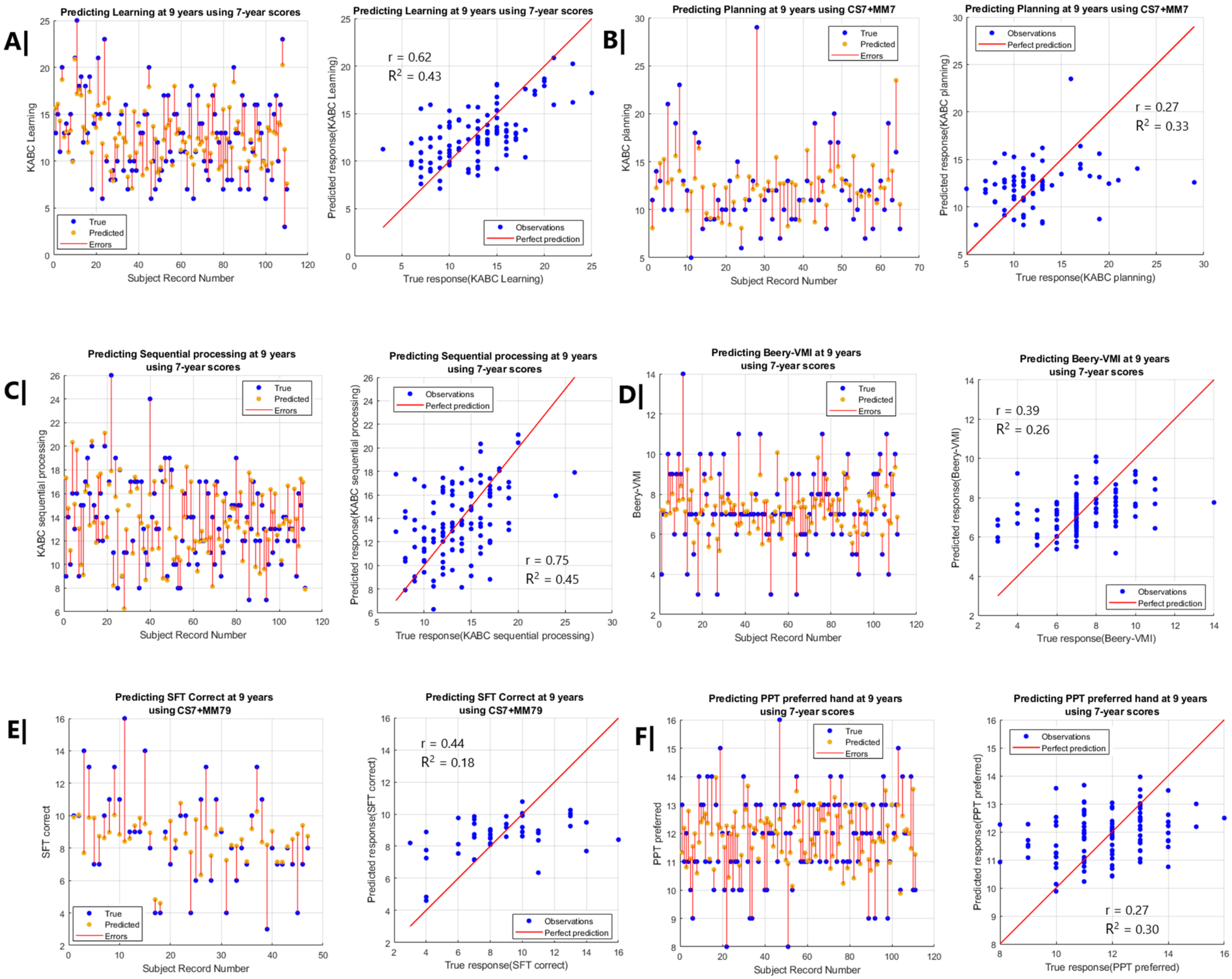


Figure 4.5| The regression performance for the subset of scores (A) KABC learning, (B) KABC planning, (C) KABC sequential processing, (D) Beery-VMI, (E) SFT correct, and (F) PPT preferred hand at age 9 years. For each subfigure, the first graph shows the cross validation errors and the second graph the prediction rate.

4.3.5. Important features in the prediction models

For the regression models, the *CS7* dataset used all 11 scores in the prediction of learning, sequential processing, Beery-VMI, and PPT preferred hand, i.e., there were no redudant predictors. The highest weights in the models predicting learning and sequential processing at age 9 were the same scores at age 7—these were also the strongest correlated pairs of individual scores between the two ages (Figures 4.6A–K). SFT (animal naming) and PPT using the non-preferred hand tended to have the highest weights in the prediction models for Beery-VMI and PPT preferred hand. Similarly, there were many multimodal MRI measures selected in the prediction of planning and semantic fluency, which are not informative given the low regression performance of the models. However, the top features were similar to those extracted during classification. As such, we focused on features of overall cognitive performance, that is, feature selection from the classification of cognition categories rather than of individual scores, since these models had higher predictive accuracies.

Outcomes on the PPVT (receptive vocabulary) and SFT (component of executive function) at age 7 were the only 2 relevant features selected when predicting poorer versus better performers class membership at age 9 using historic cognitive testing (*CS7*) alone. From the *MM7* dataset, 26 out of 489 neuroimaging measures at age 7 (supplementary Table B.8), were selected to predict categories of overall cognitive performance at age 9. Using the *CS7* + *MM7* dataset to predict category membership at age 9 resulted in 29 relevant features (ranked in Table 4.5 and illustrated in Figures 4.7A–D), which were simply the union of the features selected when the classification is performed using the datasets separately. The top 10 predictors are phosphocreatine and glutamate concentrations, area of the caudal anterior cingulate, LGI of right superior frontal, superior parietal and left medial orbitofrontal regions, volumes of the sulcal banks and right thalamus, AD in left cingulum, and the PPVT outcome.

There were 24 multimodal MRI features at age 9 selected from the MM9 dataset for predicting cognitive category membership at 9 years (supplementary Table B.9). Interestingly, the top 15 features are similar to the ones selected using the MM7 dataset, except for the additional inclusion of N-acetylaspartate concentration, and AD in left corticospinal tract. Surprisingly, of the 43 relevant features selected from the CS7+MM79 dataset (supplementary Table B.10), none of the cognitive testing measures are included.

Table 4.5| Feature ranking of the 29 features relevant for predicting cognitive category membership at age 9, obtained via SVM-RFE.

Feature	Weight* $\times 10^{-4}$
PCr concentration	19.53
RH caudal anterior cingulate area	19.17
LH pars opercularis area	17.99
Glu concentration	12.40
RH superior frontal LGI	11.19
LH medial orbitofrontal LGI	11.13
Right thalamus proper volume	11.07
AD in left cingulum	10.52
LH pars triangularis LGI	9.78
LH sulcal banks volume	9.43
RH superior parietal LGI	9.08
PPVT score	9.07
RH para-hippocampal thickness	8.78
Cr concentration	8.57
LH posterior cingulate mean curvature	8.29
RH caudal middle frontal LGI	8.20
RD in left inferior longitudinal fasciculus	7.78
Cho concentration	7.57
Left thalamus proper volume	7.50
SFT (animal naming) score	7.49
RH caudal middle frontal area	7.44
NAA concentration	7.49
LH inferior temporal thickness	7.30
LH supramarginal thickness	6.93
MD in right uncinate fasciculus	6.75

--Table 4.5 continued--

RH supramarginal volume	6.63
MD in right superior longitudinal fasciculus	6.43
Right pallidum volume	6.33
Left pallidum volume	6.22

* Weights or coefficients in the final SVM model after recursive feature elimination

Abbreviations

LH/RH - left/right hemispheres, PCr - phosphocreatine, Glu - glutamate, AD - axial diffusivity, LGI - local gyrification index, Cho - choline, SFT-semantic fluency test, Cr - creatine, MD - mean diffusivity, PPVT - Purdue picture vocabulary test

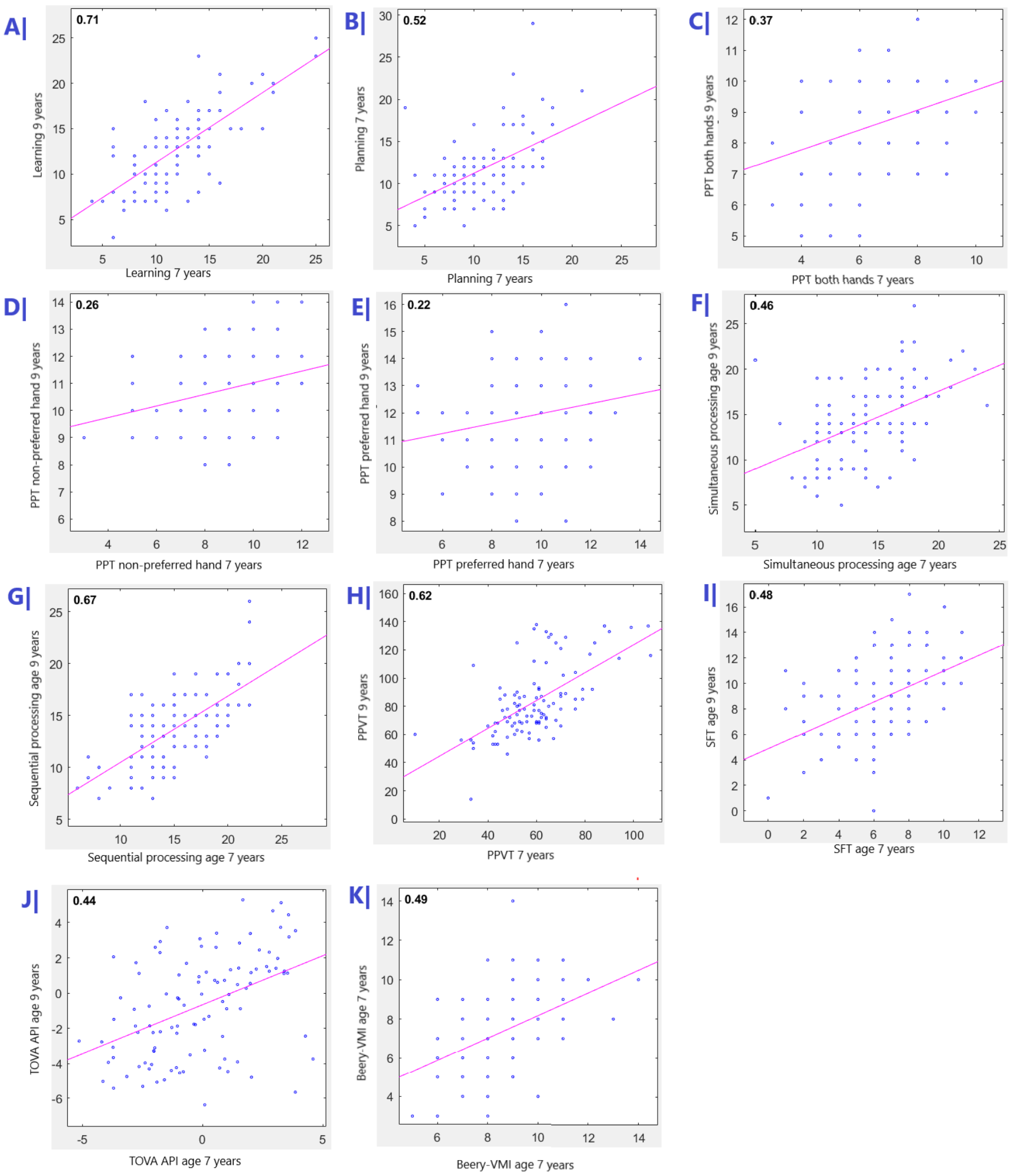
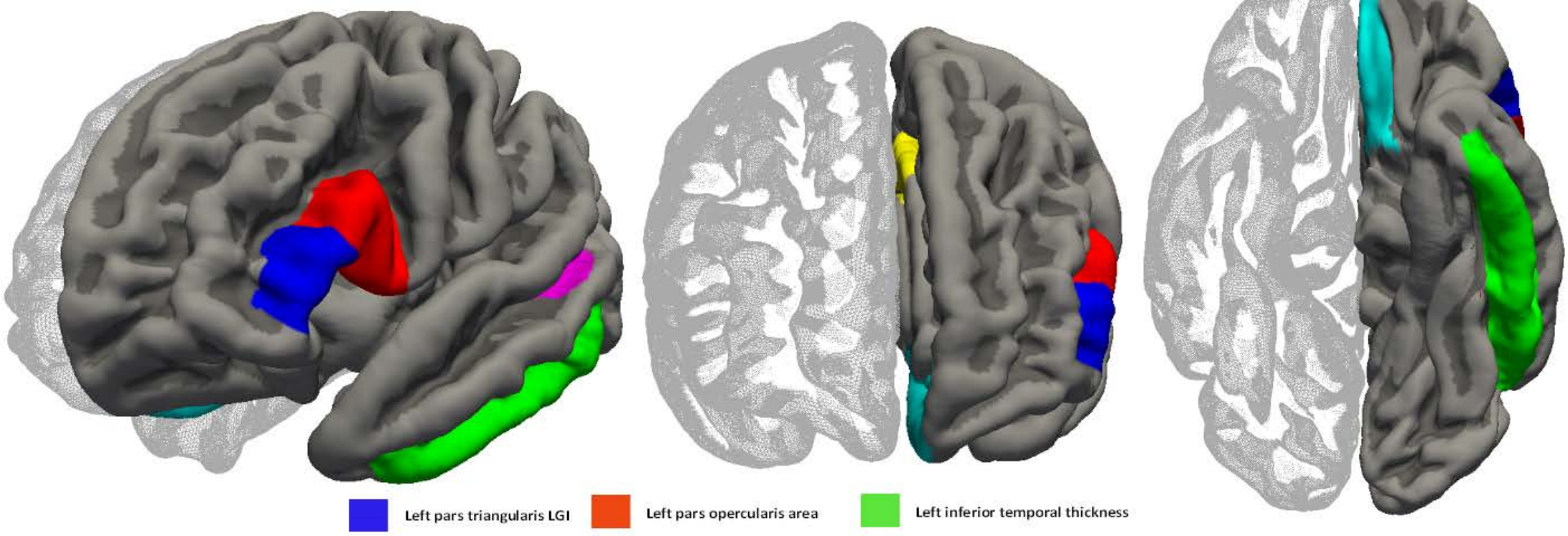


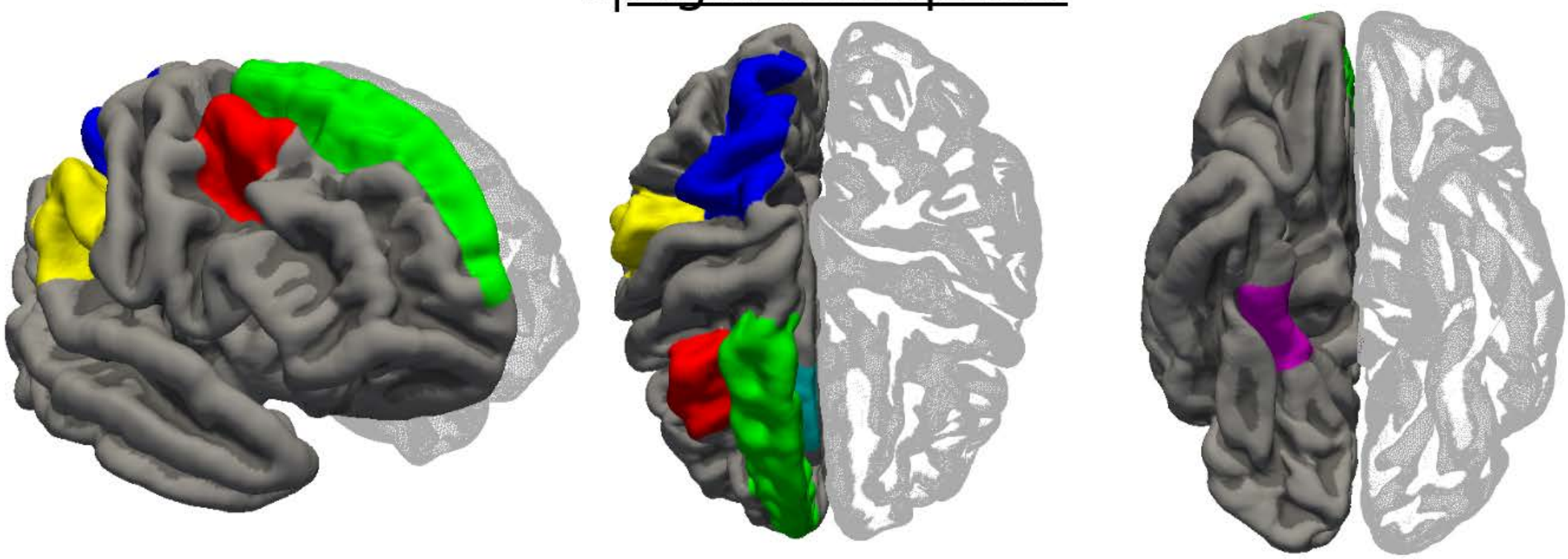
Figure 4.6| Correlation plots between the same scores performed at the different ages, 7 and 9 years. The subplots are for the pairs: (A) KABC learning, (B) KABC planning, (C) PPT with both hands, (D) PPT with the non-preferred hand, (E) PPT with preferred hand, (F) KABC simultaneous processing, (G) KABC sequential processing, (H) PPVT, (I) SFT, (J) TOVA API, and (K) Beery-VMI. The top right corner gives the correlation strengths.

A|Left hemisphere

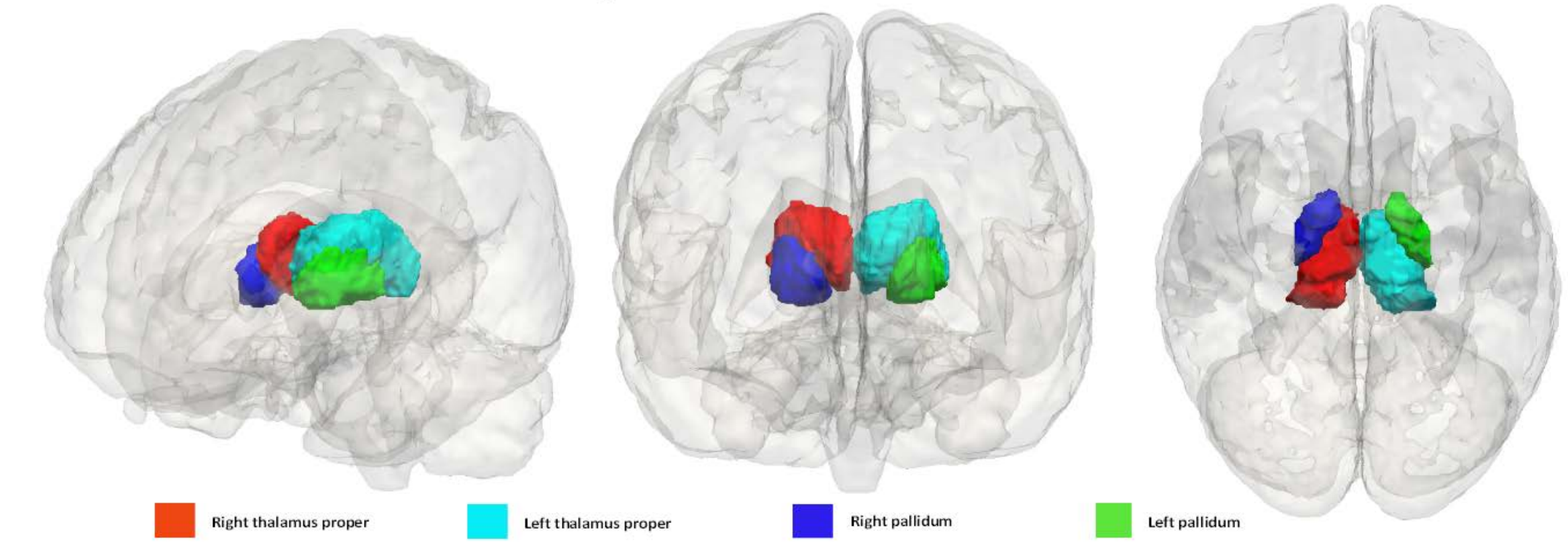


Left pars triangularis LGI
Left pars opercularis area
Left inferior temporal thickness
Left sulcal banks volume
Left medial orbitofrontal LGI
Left posterior cingulate curvature

B|Right hemisphere



Right caudal middle frontal LGI and area
Right supramarginal volume
Right para-hippocampal thickness
Right superior parietal LGI and volume
Right superior frontal LGI
Right caudal anterior cingulate area



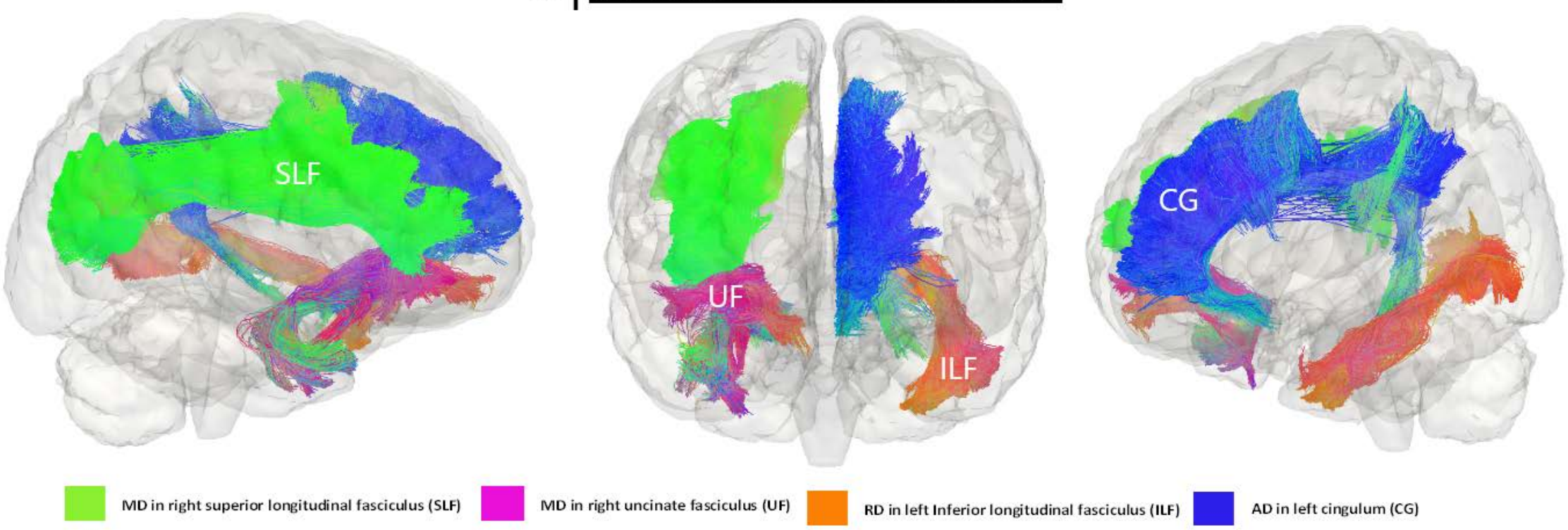


Figure 4.7| Illustration of the relevant regions/features selected usig SVM-RFE. The subfigures are 5A—sMRI cortical measures in the left hemisphere, 5B—sMRI cortical measures in the right hemisphere, 5C—subcortical volumes in both hemispheres, 5D—relevant white matter tracts. Additional to these features are concentrations of creatine (Cr), phosphocreatine (PCr), glutamate (Glu), choline (Cho), and N-aceytl-aspartate (NAA) in midfrontal grey matter region as well as outcomes on the Purdue picture vocabulary test (PPVT) and semantic fluency (animal naming) at age 7.

4.3.6. Predicting the decrease of Beery-VMI and KABC's sequential processing between age 7 and 9

We present the ROC curves for the two extra analyses we performed for predicting the decrease of Beery-VMI and KABC's sequential processing in Figures 4.8A–D. For the classification of *decrease* vs *did not decrease* categories from MM7, Beery-VMI has classification performance: 0.67 AUC, 66% accuracy, 72% specificity, and 70% sensitivity. KABC sequential processing has classification performance of 0.58 AUC, 67% accuracy, 40% specificity, and 76% sensitivity.

For the classification of *decrease* vs *did not decrease* from |MM9–MM7|, Beery-VMI has classification performance: 0.66 AUC, 64% accuracy, 73% specificity, and 56% sensitivity. KABC's sequential processing has classification performance of 0.59 AUC, 58.6% accuracy, 67% specificity, and 56% sensitivity.

We did not extract features leading to this poor classification performance in both cases.

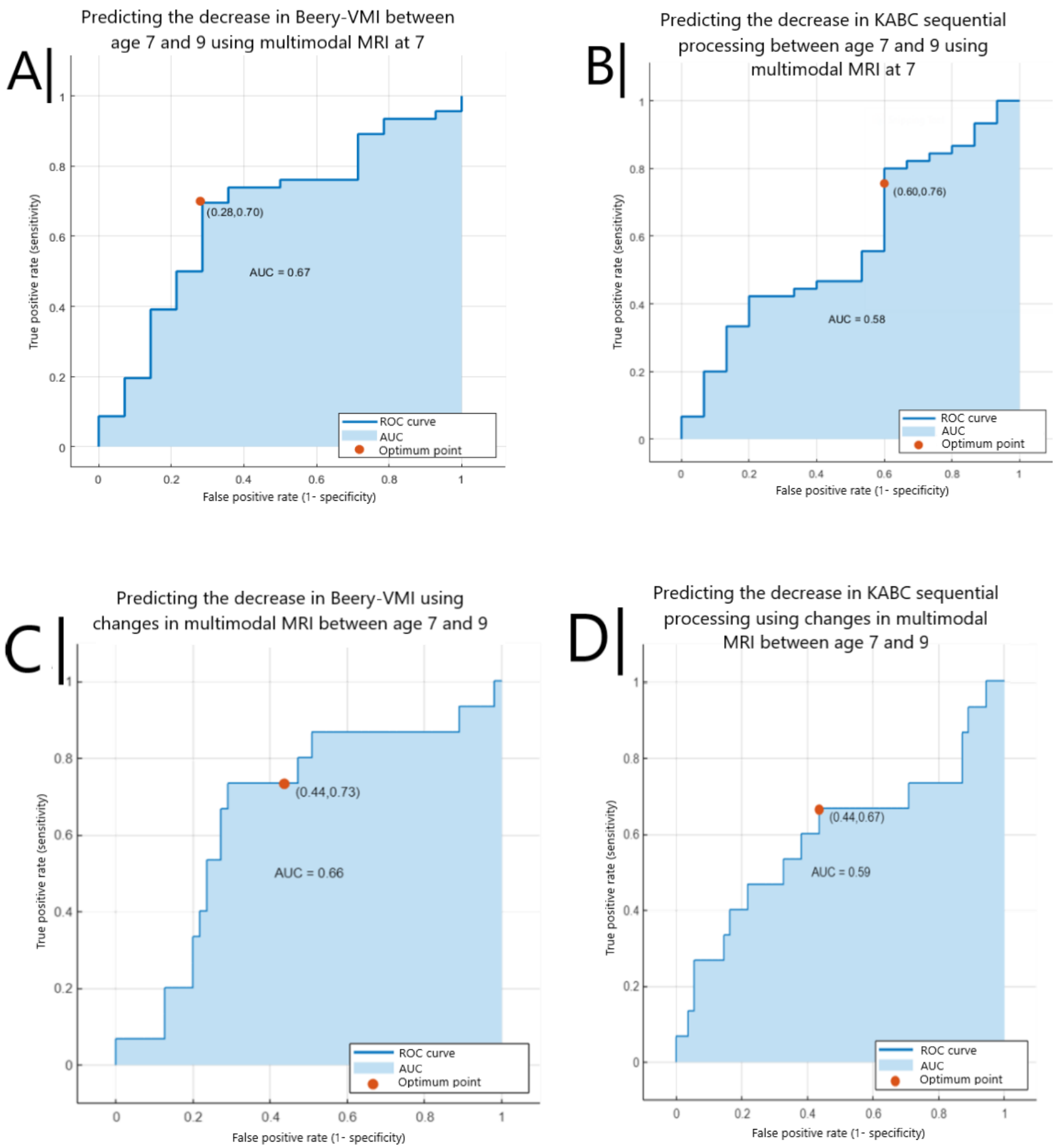


Figure 4.8| ROC curves for the classification of Beery-VMI and KABC's sequential processing *decrease* versus *did not decrease* categories. The subplots are the classification of (A) Beery-VMI and (B) KABC's sequential processing categories using multimodal MRI at age 7, then (C) Beery-VMI and (D) KABC's sequential processing categories using the changes in multimodal MRI between age 7 and 9.

4.4 Discussion

In summary, our investigation provided insights into addressing the research questions at the heart of this study: First, using historic cognitive test score data acquired at age 7, it is possible to make reasonably accurate predictions of continuous cognitive scores at age 9 in the domains of learning/short-term memory, planning/executive reasoning, sequential processing/auditory-working memory, motor coordination, and visual-motor integration ($0.27 < r < 0.62$, $0.18 < R^2 < 0.45$). However, neither multimodal MRI at age 7 nor age 9 can predict continuous cognitive scores much better than a constant mean response of the individual scores ($R^2 < 0.18$). Despite this, when cognitive performance categories are created by clustering across all cognitive scores, it is possible to use historic multimodal MRI and cognitive test data to predict these cognitive performance categories with an AUC between 0.75 and 0.81, accuracies of 70–77%, sensitivities of 71–79%, and specificities of 70–73%. In fact, even though continuous scores are better predicted using prior cognitive test data, multimodal MRI is better than individual cognitive test scores for predicting categories of cognition, i.e., "poorer performers" versus "better performers".

For the prediction of overall cognitive performance categories at age 9, multimodal MRI at the same age performed only marginally better (0.83 AUC, 76% accuracy) than the combination of historic neuroimaging-cognitive score datasets (0.81 AUC, 77% accuracy). Finally, using multimodal MRI at 2 timepoints slightly improved classification performance to 0.85 AUC and 78% accuracy, but this model used substantially more predictors (23 extra features) compared to those used by the simpler multimodal MRI datasets at 7 and 9 years separately. Notably, for prediction of cognitive performance categories, similar features were relevant in the 7-year and 9-year multimodal neuroimaging datasets.

Classification performance: predicting categories of overall cognitive performance

In this study, we could predict categories of poorer and better overall cognitive performance 2 years into the future with moderate predictive performance of 70–77% accuracies, 71–79% sensitivities, and 70–73% specificities using SVMs on historic MRI and cognitive score data. To our knowledge, Ouyang et al. (2020) and Girault et al. (2019) are the only other investigators that have classified future paediatric cognitive performance from neuroimaging. Ouyang and colleagues (2020) achieved comparative performances of 76% and 61% accuracies for predicting categories of general intellectual functioning and language scores, respectively. And Girault and colleagues (2019) achieved higher 83–90% accuracies for classifying composite score categories of cognitive ability. Our studies contrast with previous efforts which have primarily classified cognitive performance with neuroimaging measures at or closer to the age of cognitive assessment. Further, the classification focused on paediatric major psychiatric disorders, for example, ADHD in young children with an accuracy as high as 94% (Oztekin et al., 2021), infant cognitive impairment with accuracy greater than 80% (Schadl et al., 2018), and autism spectrum disorder with a moderate accuracy of 71% (Lanka et al., 2020).

While in this study our prediction was achieved from a multimodal neuroimaging combination of sMRI, DTI, and ^{1}H-MRS at age 7 to predict overall cognitive performance at age 9, Ouyang et al. (2020) and Girault et al. (2019) used DTI measures acquired at birth (30–43 postnatal weeks) to predict cognitive abilities at age 2. Since Girault et al. (2019) and Ouyang et al. (2020) investigated infants, there may be fewer unmeasured environmental factors that contribute to variations in either brain structure or cognitive performance at that point than in school-aged children in our study. Additionally, we studied children from low-socioeconomic status backgrounds, who perform worse than their peers at the same age, as our cognitive data shows and other studies have reported on too (Bush et al., 2020; McMaughan et al., 2020). However, there were two domains of learning and sequential processing where the low-SES children in our study performed

in the average and upper average range, although the expectation was that they would perform worse than their peers at the same age. This may be in part due to how well the cohort was treated, i.e., with three-monthly adherence visits for those living with HIV, six-monthly clinical assessments for children without HIV as well as afterschool learning programmes, opportunities that are not available to all children from low-SES communities.

Furthermore, this study and Ouyang et al. (2020) trained SVMs, while Girault et al. (2019) trained a deep learning classification model, which may explain the better predictive accuracy. However, contrasting results may also be explained by the cognitive performance categories these 3 studies are attempting to predict. Girault and colleagues (2019) combined motor, visual and language scales of the infants into an early learning composite score, then created 2 groups: scores above and below the median, while Ouyang et al. (2020) categorised the infants into normal and low scores based on the cognitive test battery's normative data. Our approach was a data-driven clustering on multiple cognitive domains to categorise poorer and better performing children in the sample.

There are other ways of creating categories from longitudinal cognitive performance data, which we explored in this study. Since van Wyhe and colleagues (2021) found that Beery-VMI (visual-motor integration) and KABC sequential processing (auditory working-memory) scores for our study sample significantly decreased from 7 to 9, while SFT, PPT, and PPVT increased in both children living with and without HIV. We also wondered if multimodal MRI could be used to predict whether a child's score would decrease with age. As such, the children we categorised into those whose cognitive scores decrease over time and those whose scores increased/stayed the same. We found that multimodal MRI at age 7 did not do much better than chance in predicting a decrease in the scores (0.58–0.67 AUC, 60–66% accuracy). We also explored whether changes in multimodal MRI (7 to 9 years) can predict these changes in cognitive measures; but classification

performances were also not convincing (0.58–0.60 AUC, 60–67% accuracy). This suggests that multimodal MRI may not be good at predicting 2-year changes in cognitive performance. Or a decrease in cognitive performance over a 2-year period in this cohort may not involve structural brain changes that a multimodal combination of sMRI, DTI, and ^{1}H-MRS is sensitive to, and the decrease in score could be a result of temporary fluctuations in mood, energy or motivation.

Regression performance: predicting cognitive performance domains

As we noted previously, several studies have predicted scores in different paediatric cognitive domains using neuroimaging and machine learning algorithms; the regression performance of $0.18 < R^2 < 0.45$ and $0.27 < r < 0.62$ between actual and predicted values, is in line with those found by brain-wide association studies (Marek et al., 2022). In the prediction of continuous neurodevelopmental outcomes 2 years into the future as opposed to classification, Ouyang and colleagues (2020) found significant correlations between the predicted and actual scores of general intellectual functioning ($r = 0.54$) and language ($r = 0.47$). Girault et al. (2019), likewise in the regression aspect of their study, found $r = 0.98$ and $r = 0.96$ for full term and preterm infants, respectively, in the prediction of general intellectual functioning at 2 years from white matter connectomes at birth. The disparate regression results could possibly be explained by the methodological differences between these studies and ours that were previously noted while discussing their classification accuracies.

Using all 11 historic cognitive scores rather than multimodal MRI, we could predict auditory working memory ($R^2 = 0.75$, $r = 0.45$) and category fluency ($R^2 = 0.18$, $r = 0.44$), but not receptive verbal language ($R^2 = 0.05$, $r = -0.20$) scores. Interestingly, Adeli and colleagues (2019) could predict infant language scales at age 4 with r between 0.58 and 0.72 using multi-task multi-linear regression and cortical morphology at 24 months of age. However, Adeli et al. (2019) used sMRI data from 9 earlier timepoints, despite having incomplete sulcal depth, LGI, mean curvature, and cortical thickness at

some of them. Similarly, Khundrakpam et al. (2022) used cortical thickness of 5 to 14-year-old children scanned at 2 timepoints 2 years apart to predict executive function composite score with r = 0.61; adding to the evidence that cortical morphology at multiple timepoints can accurately predict future cognition. Comparatively, we could only achieve a low $R^2 = 0.33$ and r = 0.27 when predicting executive reasoning with historic data that included cognitive scores and multimodal MRI. We speculate this is because these authors considered more dimensions of executive function in children, specifically scores of behaviour regulation and meta-cognition, while we focused on reasoning and auditory working memory. Moreover, the earlier study by Ullman et al. (2014) could predict future visuospatial working memory with sMRI (r = 0.29), DTI (r = 0.59), resting-state fMRI data (r = 0.44), and using a combination of all the imaging data (r = 0.64) and the same cognitive score at the earlier timepoint (r = 0.78). In contrast, most of the individual future cognitive scores in this study could be predicted from a combination of earlier scores (category fluency and motor coordination being the most predictive domains) than neuroimaging. The earlier scores of executive reasoning and auditory working memory were some of the individual scores which were moderately correlated with their future counterparts (r = 0.52 and r = 0.67, respectively). While we used a sample size similar to Khundrakpam et al. (2022), n = 82 children, Adeli et al. (2019) used a much smaller sample of n = 24 children, and Ullman et al. (2014) had the largest sample of n = 323. The extent to which these findings are scalable and reproducible remains an open question.

Recently, there have been commentaries on reliability and reproducibility of regression-based multivariable predictive brain-behaviour studies (Callaway, 2022; Gratton et al., 2022; Kong et al., 2022; Rosenberg & Finn, 2022; Spisak et al., 2023), a conversation initiated by Marek et al. (2022) who argued that sample sizes in the several thousands are needed for reliable and reproducible results. The authors used 3 large-scale publicly available MRI datasets: Adolescent Brain Cognitive Development (ABCD, n = 11 875;

9-10 years), Human Connectome Project (HCP, n = 12000; 22-35 years), and UK Biobank (UKB, n = 35 735; 40-69 years). In their analyses, r = 0.39 was the strongest multivariate correlation between an imaging measure and a behavioural phenotype that they could reliably reproduce. They found in-sample r is variable at small sample sizes and stabilises with more sample, out-of-sample r also increases with sample size for cognition, and r was higher for multivariate imaging predictors than in univariate analyses. This study used only 2 neuroimaging modalities: sMRI cortical thickness and fMRI functional connectivity. It is interesting that functional connectivity produced the highest correlation with cognitive scores, and we have not included it in our analyses due to poor resting-state fMRI acquisition data. By comparison, we made use of sMRI (thickness, curvature, area, volumes, and gyrification), DTI, and ^{1}H-MRS in the midfrontal gray matter region. Although sample size does play a role in reliable and reproducible results, the type of data collected and analyses (e.g. learning algorithms) applied to them matters too. We used multivariate analyses, which tend to mitigate the need for large sample sizes (Marek et al., 2022). Furthermore, a large sample is not always required when longitudinal data is used. In our study with had two longitudinal data variables, (1) the combination of multimodal MRI at ages 7 and 9, which was tested in conjunction with the cross-sectional cognitive scores at age 7 dataset, as well as (2) the change in multimodal MRI features from age 7 to 9.

Feature selection and brain-cognition biomarkers

The best predictive regression models utilised only historic cognitive performance to predict future cognitive scores in the domains of short-term memory, executive reasoning, auditory working memory, motor coordination, and visual-motor integration. Although all the domains were used in the prediction models, motor function and receptive vocabulary tended to have the highest weights. This might suggest that good performance in these domains earlier in life is a foundation that is required for development in other domains. By contrast, the best classification models were the ones that used multimodal

MRI features. This suggests that multimodal MRI features may be helpful in predicting categories of overall cognitive performance rather than continuous values in individual domains. A possible explanation is since there may be measurement errors in paediatric intelligence testing (Torres van Grinsven, 2022), multimodal MRI can be more effectively used to predict which range an individual will fall into, rather than their specific outcome on a continuous scale. In addition, the pooling across performance on different tests that is achieved using the clustering method may better index some underlying cognitive capacity that is easier to predict using neuroimaging as suggested by our analysis and other studies (e.g., Paul et al., 2022).

There were 26 relevant features out of 489 for predicting categories of performance at age 9 using prior multimodal MRI, and 24 relevant features when the prediction was done with multimodal MRI at the same age. Notably, the top 15 imaging features were similar at both ages, implying that the most relevant predictors of cognition remain constant at least for a 2-year period. In total, we identified 29 historic predictors of future cognitive performance category. These included midfrontal grey matter metabolite concentrations creatine/phosphocreatine, glutamate, choline, and N-acetyl-aspartate; regional cortical areas of the left caudal anterior cingulate, left pars opercularis, and right caudal middle frontal regions; local gyrification indices of the right superior frontal, medial orbitofrontal, pars triangularis, right caudal middle frontal, and superior parietal regions; regional and other subcortical volumes of thalamus proper, left sulcal banks, supramarginal gyrus, and globus pallidus; cortical curvature of the left posterior cingulate; diffusivity measures AD in left cingulum, RD in left inferior longitudinal fasciculus, MD in right uncinate fasciculus, and MD in right superior longitudinal fasciculus tracts; cortical thickness of the para-hippocampal, inferior temporal, and supramarginal regions; as well as the outcomes of receptive vocabulary and verbal generativity tests.

Cortical morphology measures like thickness, gyrification, curvature, and volumes have been used by previous studies to make good predictions of future cognition (Adeli et al.,

2019; Khundrakpam et al., 2022). Adeli and colleagues (2019) found that curvature was relevant in earlier timepoints (0–9 months), while cortical thickness was quite important in later timepoints (12–48 months); gyrification and sulcal depths were not very predictive of future cognition. In contrast, we find more gyrification measures weighted higher than curvature and cortical thickness measures in the prediction models; however, we have studied older children than Adeli et al. (2019). Khundrakpam and colleagues (2022), could use cortical thickness alone to predict future executive function. Both studies did not perform feature selection as we have done here; this could have informed which cortical morphometric structures contributed to the prediction of future cognition.

We found volumes of the globus pallidus and thalamus; functional activation in these structures, along with the caudate nucleus and putamen, were found by Ullman and colleagues (2014) in the prediction of future working memory and reasoning, suggesting the importance of basal ganglia structures in cognitive abilities of children.

In line with other studies (Girault et al., 2019; Ouyang et al., 2020) that used DTI measures to predict future cognition, we have found diffusivity measures in 3 white matter tracts contributing to the predictive accuracy. However, Girault and colleagues (2019) found white matter connections confined to the frontal lobe, and Ouyang et al. (2020) found FA in right precuneus and bilateral rectus gyri contributing more to the prediction, whereas in this study our DTI predictors were large white matter tracts instead of diffusivity in cortical regions. As the systematic review by Vieira et al. (2022) has noted, more DTI and MR spectroscopy studies are needed to understand the neural underpinnings of cognitive performance in children and adults.

Lastly, receptive vocabulary and verbal generativity were the historic cognitive scores present in the prediction of future overall cognition, it seems very plausible that language skills would predict cognitive performance in other domains.

Limitations and recommendations for future research

We have alluded to the impact that a moderate sample size of n = 112 for hierarchical clustering and n = 88 for regression and classification may have on overall performance and generalisation of these findings. A way to assess whether our models will generalise well is testing on new data, ideally from a different site and independently collected, e.g., ABCD, HCP, and UKB datasets. To test for population-level model generalisability, it would be even better to externally validate our predictive models on a different cohort of children from low-SES backgrounds.

Although the inclusion of 3 modalities of MRI is a strength of this study, we have excluded ^{1}H-MRS in basal ganglia and peritrigonal white matter and resting-state fMRI due to a large amount of missing data at one or both timepoints. Several studies have achieved predictive performance ($-0.4 < r < 0.6$) using fMRI and functional connectivity (Pat et al., 2022; Yuan et al., 2022), and a recent systematic review (Vieira et al., 2022) provides good evidence for including task-based fMRI in multimodal MRI models. We did not collect any task-based fMRI, which may have larger associations with cognitive test performance (Vieira et al., 2022). There are methods of dealing with incomplete data that may be worth exploring in future studies, such as imputation methodology (Oberman & Vink, 2023) or the multi-task multi-linear regression framework proposed by Adeli et al. (2019) in their prediction of infant cognition from incomplete longitudinal neuroimaging data. Additionally, none of the non-imaging and non-cognitive scores predictors, namely: age at scan, sex, birthweight, and HIV status were relevant in the prediction models. Parental education, parental income, and maternal intelligence quotient have also been linked to children's overall intelligence (Flensborg-Madsen et al., 2020) and worth including in future models. Although HIV was not a predictor of overall cognitive performance, it has been reported to cause impairments in cognitive domains of executive function (Rowe et al., 2021) as well as learning and language (Musindo et al., 2022).

The automated cortical reconstruction and volumetric segmentation of the T1-weighted datasets at ages 7 and 9 were done independently as opposed to using the longitudinal stream FreeSurfer. While the longitudinal stream may have been ideal, it assumes a constant head size and tested on adult populations primarily (Reuter et al., 2012), as such was not preferred for our paediatric cohort and analyses which were largely cross-sectional in nature. Lastly, SVM/R is a type of single-task method, where only one type of phenotypic measure is predicted at a time; only a small number of studies perform multitask predictive modelling, which can offer more complex brain-behaviour relationships (Sui et al., 2020). Multitask modelling considers the dependent relationship among measurements derived from a single cognitive test to simultaneously predict various phenotypic variables in a unified framework (ibid.). We have chosen a single-task method, because it offered more interpretability than a multitask model would have done. SVM/R allowed us to map features to the original feature space, extract their weights (β coefficients), and rank them in the prediction models. Since these features were relatively constant for a 2-year period, it would be interesting to explore how they change when more timepoints are added.

5

DISCUSSION: exploring the relationship between magnetic resonance neuroimaging and cognitive outcomes in children

Chapter 5 provides a discussion of findings, limitations, and recommendations for future research.

Highlights & key findings of the dissertation

- In summary, there is evidence in the literature to suggest that neuroimaging measures in children, adolescents, and youth living with and without HIV are indicative of cognitive abilities measured using cognitive testing. Cognitive domains of attention, executive function, and general intellectual functioning can be assessed with multimodal magnetic resonance neuroimaging techniques like structural MRI (sMRI) and diffusion tensor imaging (DTI). Global white matter and grey matter volumes (WM, GM) and microstructural integrity, particular of the corpus callosum subcortical structure, were the neuroimaging measures mostly indicative of cognitive performance in young people (Chapter 2).
- Penalised linear models (PLMs), support vector machines/regression (SVMs/R), and decision tree ensembles (DTEs) do not predict continuous cognitive scores very accurately. However, the models are more accurate in the prediction of poorer and better categories of overall cognitive performance in the sample of children from low-socioeconomic status backgrounds. There are subsets of neuroimaging measures such as the globus pallidus, corpus callosum,

and cerebellar white matter volumes, temporal lobe curvature, and white matter microstructural integrity of the superior longitudinal fasciculus and uncinate fasciculus that are mostly predictive of cognitive performance at age 7 (Chapter 3).

- It is possible to predict overall cognitive performance categories at age 9 using neuroimaging and/or cognitive scores from age 7. Multimodal MRI neuroimaging is better than cognitive scores at predicting subsequent cognitive performance categories. Using current and historic multimodal MRI (neuroimaging at 2 timepoints) improves the prediction of overall cognitive performance. Concentration of N-acetylaspartate, phosphocreatine and glutamate, areas of the caudal anterior cingulate and pars opercularis, gyrification of the frontal lobe, volumes of the thalamus and globus pallidus, and white matter microstructural integrity of the cingulum and superior longitudinal fasciculus, as well as the cognitive domain of language are among the historical predictors of future overall cognitive performance. (Chapter 4).

5.1 Discussion

We aimed to add to the growing evidence on the utility of using machine learning (ML) algorithms to predict cognitive abilities of children from multimodal neuroimaging. It is important to understand relationships between neuroimaging and cognitive performance because it could make it easier to identify children at risk of poor neurodevelopmental outcomes or cognitive impairment. This would add great benefit to performing neuroimaging in children even when there is no clinical need. At-risk children identified in this way could then benefit from targeted interventions to prevent struggles at school and in other areas of life.

Risk factors for cognitive deficits in children and adolescents

Living with HIV, perinatal exposure to HIV, and living in low-socioeconomic status (SES) communities and households are all risk factors for cognitive deficits in children and adolescents (Musindo et al., 2022). Studies show these risk factors are linked; HIV disproportionately affects low-SES communities (Cleary et al., 2011; Wabiri & Taffa, 2013), children with perinatal HIV are more likely to be orphaned because of parents who die from HIV, and orphanhood can lead to under-nutrition in children (Musindo et al., 2022). In some circumstances, the lack of appropriate caregiver stimulation necessary for cognitive development can lead to cognitive delays and deficits in orphaned children. This raises the important question of whether reported cognitive deficits in children living with HIV are due to HIV or socioeconomic factors.

On one hand, the evidence from the systematic review (§2.3.3 synthesis) we conducted, consisting mostly of case-control studies, suggests that HIV may lead to brain changes linked to poorer outcomes in attention (short-term memory and processing speed), executive function, and general intellectual functioning. The sample population in our observational study comprised children from low-SES backgrounds (52% HIV and 48% controls of whom 40% were exposed to HIV). The clustering of overall cognitive

performance scores in our sample at age 7 (§3.3.1) and 9 §4.2.3), revealed that the lowest scoring children were all HIV positive, the highest performers were unexposed controls, and the children straddling the two clusters were a combination of HIV positive and HIV-exposed controls. This would support the idea that HIV is an important factor that negatively influences cognitive development in this cohort, as van Wyhe and colleagues (2021) have previously reported poorer working memory and executive function performance of children living with HIV in this cohort. Furthermore, the prediction models did not misclassify the lowest and highest scoring children, suggesting their brain measures are more different to each other than to the HIV-exposed children.

On the other hand, HIV status was not a relevant feature in any of the classification of poorer and better performers in the sample, i.e., the two broader parent clusters. It might be interesting to investigate health-related characteristics that may affect whether children living with HIV will fall into the poorer or better performing range. However, at both ages both groups of children generally scored worse than their peers at the same age in the United States. (The comparison to normative data from the USA is elaborated on in section §5.2 limitations on cognitive testing.) Absent any cognitive testing related factors, this may support cognitive delays often seen in children from low-SES communities. The children with poorer performance in this South African low-SES cohort corresponded to the lowest scores in their international peers and this performance was most evident in the domains of attention, learning, language, and auditory working memory. These domains are impaired in children with HIV according to several systematic reviews (Ensink & Kuper, 2017; Musindo et al., 2022; Phillips et al., 2016; Rowe et al., 2021). Overall, it is important to note that both children living with HIV and children from low-SES, regardless of HIV status, can benefit from targeted interventions.

Our systematic review showed how diverse neuroimaging findings can be when looking for signatures of cognitive abilities that are affected by a particular disease. Therefore, by combining neuroimaging modalities and cognitive domains, perhaps in a more

interpretable way than the clustering we have implemented, it may be possible to find neuroimaging patterns that index general cognitive abilities. The next step would be to investigate how these patterns are affected by diseases including perinatal HIV; this might be a challenging task because of the heterogenous disease history and characteristics. Although disease history data including viral loads and CD4 counts is available for the children living with HIV in this cohort, we did not use it here, partly because this data did not exist for the control children. However incorporating disease history variables into models is likely to improve predictive ability for poor cognitive outcomes.

Performance of the regression and classification models

In this study, we investigated the prediction of continuous cognitive scores (regression) and categories of overall cognitive performance (classification) created from the continuous data using various supervised learning algorithms, namely: penalised linear models (PLMs), decision tree ensembles (DTEs), support vector machines/regression (SVMs/R) at age 7 (Chapter 3) and elected to use SVMs/R for predicting future cognitive performance at age 9 (Chapter 4). The regression performance was generally poor, shown by the tendency to have low fitting errors and high generalised error. The Pearson's correlation r between predicted and actual value was comparable to values obtained by previous studies. We achieved correlations of $0.20 < r < 0.34$ for same-age predictions (§3.3.2) and $0.27 < r < 0.62$ for predicting future cognitive scores (§4.3.3). The stronger correlations with future than same-age cognitive performance suggests brain development is more of a foundation for acquisition of performance-related skills than an indication that these skills are present or have already been acquired. Previous studies have also found weak to moderate associations between predicted and actual scores ($0.003 < r < 0.98$) using structural MRI and diffusion tensor imaging data separately (Dubois et al., 2018; Khundrakpam et al., 2022; Kline et al., 2020; Mihalik et al., 2019; Ouyang et al., 2020; Sripada et al., 2021) and, less-commonly, multimodal MRI (Marek et al., 2022; Pat et al., 2022). However, the common conclusion which can be drawn from most of

these predictive models is that they are unlikely to be generalisable due to their high cross-validation errors.

For this reason, in this study we also identified categories of generally poorer and better performers across all tests rather than reliance on performance on individual tests. We used hierarchical clustering at age 7 and 9 years, and found the technique to be more robust to some of the variability in test performance data due to random or test-administration-related factors. Interestingly, the lowest and highest scoring children at age 7 who were clustered to the poorer and better performing categories were also clustered into the same categories at age 9. For these children, we could predict from the clustering output alone whether they are HIV positive or HIV-unexposed controls as well as whether they will be poorer or better performers in the future. Although there were few children who either went from poorer to better performance categories and vice versa, multimodal MRI neuroimaging was particularly helpful in classifying these children.

Indeed, the classification models fared much better than their regression counterparts. Looking at the overall performance, the future classification of poorer and better performers with multimodal MRI neuroimaging primarily (§4.3.2: 0.75–0.81 AUC, 70–77% accuracies, 71–79% sensitivities, and 70–81% specificities) was similar to the same-age classifications of these categories (§3.3.3: 0.71–0.83 AUC, 65–74% accuracies, 70–85% sensitivities, and 65–73% specificities). Like in the regression models, we expected future classification of cognitive performance to be higher than same-age predictions from the background knowledge about brain development. This is true when we compare the classification performance metrics of the same supervised learning algorithm, i.e., SVMs. There was a slight imbalance in the members in each poorer and better performance categories, so learning algorithms such as random under-sampling boosted trees would perform better than SVMs in terms of balanced sensitivities and specificities. In cases where SVMs are used, the balanced accuracy or balanced F score measure would be a useful metric in assessing the overall performance of the prediction models. For this

study, we have reported the balanced accuracy measure alongside the other metrics, however it did not significantly differ from the accuracy metric. In general, however, the choice of ML algorithm did not seem to matter if the parameters of the models were optimisable (Chapter 3).

Biomarkers of overall cognitive performance

There were 2 notable results regarding feature selection. First, despite the regression models generalising poorly, similar brain measures were selected for the regression of individual cognitive scores as were selected in the classification of overall cognitive performance. This may suggest that the same subset of features may be useful for predicting performance in individual cognitive domains as for predicting overall cognitive performance. It further reflects the interdependency of the cognitive scores used, that the same underlying cognitive skills and brain regions are most associated with cognitive performance measures. While some regions are specialised for certain tasks, this shows a great deal of overlap and interconnectivity between different regions. Second, the neuroimaging predictors of same-age (§3.3.5) and future (§4.3.5) overall cognitive performance were different. For example, the prediction of future cognitive performance included magnetic resonance spectroscopy measures. In our view, the measures differed likely because not the same children belonged to the poorer and better performance clusters at ages 7 and 9. Furthermore, when multimodal MRI at ages 7 and 9 were used to predict cognitive performance at the same ages, respectively, different features were relevant in the prediction models. This suggests the relationship between multimodal MRI neuroimaging and cognitive performance may change with age as some of the studies have demonstrated (e.g., Adeli et al. 2019). For predicting cognitive performance at age 7, global volumetric features were most prominent, including volumes of the corpus callosum, cerebral and cerebellar white matter, pre- and post-central gyri and lateral ventricles. However, for predicting 9-year cognitive performance, white matter microstructural

properties, gyrification indices and metabolite levels were more important. This is consistent with cognitive abilities being supported by growth of brain gray and white matter in early years, particularly regions involved in motor and sensory functions. However synaptic pruning and white matter maturation may play more of a role in cognitive abilities at a slightly later stage of childhood.

5.2 Limitations and recommendations

Several important limitations need to be considered. Below, we elaborate on the limitations on sample size, comment on the scope for improving the quality of magnetic resonance imaging and cognitive test data, and discuss the generalisability of the results, and conclusions that can be drawn from the regression/classification models. Furthermore, we discuss the challenges in paediatric MRI neuroimaging and what could have been improved in this study as well as providing a potential framework for future research.

Sample size and its relationship with ML algorithms

Our analysis was limited to a sample population of 132 children, however, only a subset of 88 children had complete multimodal MRI neuroimaging and cognitive test data at both ages 7 and 9. In this study, factors that affected the sample size were (i) attrition in the control group in the course of the longitudinal study meaning that some of the children were not assessed or scanned at both timepoints; (ii) exclusion of poor quality magnetic resonance (MR) images due to signal dropout or large motion artefacts, which meant not all neuroimaging modalities had usable data; and (iii) failure to complete all cognitive test batteries because of failing the practice tests beforehand, fatigue, or other mood-related considerations. It is often difficult to recruit and maintain a large cohort in longitudinal multimodal neuroimaging studies, and large-scale studies require more financial resources and time. But our sample size is not small relative to the sample sizes

commonly-used for structural and functional MRI studies: 12–50 participants according to a review by Szucs and Ioannidis (2020) of studies conducted between 1990 and 2018. The relationship between sample size and machine learning (ML) algorithms is complex; in general, the greater the number of observations and predictors the better. Predictive modelling studies with inadequate samples are prone to overfitting of the data and have lower probability of generalising well (Jollans et al., 2019; Rajput et al., 2023; Scheinost et al., 2019). Jollans and colleagues (2019) recommend at least a sample size of 400, while others argue that sample sizes in the several thousands are required for reliability and reproducibility (Marek et al., 2022). In this respect the sample size we used was rather small and it would be desirable to replicate the results on a larger sample, since large datasets increase the likelihood of identifying robust and generalisable brain biomarkers of cognitive performance (Sui et al., 2020).

Promotion of data sharing will allow larger datasets, although there will be heterogeneity in imaging and population characteristics that will need to be considered. For example, we acquired data from two scanners; however, since we did not look at a quantitative change in neuroimaging measures between timepoints, but instead entered them as separate variables in the prediction models, we did not expect there to be extra variability introduced due to brain measures acquired from the two different scanners. Some scans are infrequently acquired, e.g., single-voxel spectroscopy (SVS), while some like T_1 Magnetization Prepared-RApid Gradient Echo (MPRAGE) are almost universally acquired. Therefore, structural MRI is a very good candidate modality for large-scale data pooling. So large-scale publicly available paediatric multimodal MRI datasets like the Adolescent Brain Cognitive Development and UK Biobank will make it possible to perform this kind of study in the future. However, the strengths of our study were as follows, (1) using multivariable/multimodal data in ML analyses, (2) measuring out-of-sample correlation coefficients or classification/regression performance metrics, and (3) predicting cognitive performance evaluated by validated test batteries rather than using self-

report/mental health measures. This study also demonstrates the usefulness of DTI and MR spectroscopy in predicting cognitive performance, and how these other MRI modalities need further attention in understanding cognitive performance of children.

Cognitive testing: normative data and clustering of cognitive scores

Our cognitive assessment is limited to the cognitive domains of attention, executive function, visual perception, language, motor coordination, and visual-motor integration as well as general intellectual functioning (a combination of these domains). Using multiple cognitive batteries to assess multi-domains of cognition is another strength of this study. We could cover all the domains that share a bidirectional link with academic skills (Lovden et al., 2020; Peng & Kievit, 2020). However, we did not test for socio-personal performance (e.g., social cognition, disruptive behaviour, self-concept, or apathy), which is an important domain especially in the context of psychoeducational implications, practice and policy making related to paediatric HIV (Cockcroft & Cassimjee, 2020; Gabbidon et al., 2020; Tuke et al., 2020) in sub-Saharan Africa and children from low socio-economic status backgrounds in general (Dalmaijer et al., 2023).

The quality of the cognitive test scores have secondary influences such as anxiety, fatigue, pain, or response expectancies that may impact test validity (Johnson-Greene, 2013). For example, some of the children were able to complete short-term memory and working memory tests but failed the practice tests for attention performance index (API or ADHD score) multiple times. We understand a combination of these factors may have played a role. Trained neuropsychologists in this study did not administer cognitive tests when practice tests were failed 3 times. When calculating overall cognitive performance, these children were excluded from the analyses since they had missing data for certain cognitive domains (e.g., API). This aspect impacted the overall sample size when classifying the overall cognitive performance categories, whereas the prediction of individual scores included all the children whose data was available.

Normative comparison is a critical concept and perhaps a limitation of cognitive testing (Harvey, 2012). Cognitive test batteries are not diagnostic in and of themselves, instead, they show cognitive performance in comparison to a reference group. There is evidence to suggest the batteries have limited cross-cultural validity in South Africa since the normative data is based on children in high-income settings like the United States (Harris et al., 2021; Mitchell et al., 2017). When creating a score vector representing overall cognitive performance, we used raw scores for some tests and scaled scores for others, where the scaled score is a child's raw score in the distribution of raw scores among his or her peers at the same age. The raw scores were used for Purdue pegboard test (PPT), semantic fluency test (SFT), and Purdue picture vocabulary test (PPVT), since substantive normative data does not exist for these batteries. More research is needed in validating these batteries in African populations as well as the validation of Kaufmann battery for children (KABC-II), test of variables of attention (TOVA), and Beery-VMI, which currently have normative data based on Western populations. However, combining scores in this way would not have affected the categories of overall cognitive performance, because the hierarchical clustering considers the similarity of scores within our sample population.

Multimodal magnetic resonance imaging predictors

Our multimodal neuroimaging investigation was limited to brain measures acquired via MRI, which is an imaging technique favoured both clinically and in the neurosciences because of its excellent soft tissue contrast resolution, high signal-to-noise ratio, absence of ionizing radiation, and the ability to examine different properties of tissue such as morphometry (sMRI), WM microstructure and connectivity (DTI), and local brain metabolism (^{1}H-MRS) using the same scanner. Our multimodal MRI analyses included only sMRI, DTI, and ^{1}H-MRS in the midfrontal grey matter region.

We excluded resting-state fMRI data from the multimodal MRI analysis, because of poor quality data for more than half of our sample population at both timepoints. Since

functional connectivity of resting-state brain networks has been associated with scholastic performance (Chaddock-Heyman et al., 2018), attention behaviour (Ding & Pang, 2021; Yadav et al., 2018), and executive function (Sarma et al., 2021) in children, it is worth investigating whether the inclusion of resting-state fMRI connectivity or measures such as regional homogeneity and mean and fractional amplitude of low frequency fluctuations in multimodal analyses would improve classification and regression models for predicting cognitive performance. Although, task-based fMRI is a closely-related modality that has been shown to predict cognitive outcomes in children (Pat et al., 2022; Yuan et al., 2022), resting-state fMRI may be preferable to task-based fMRI in the context of paediatric neuroimaging, due to its simplicity and it eliminates barriers the children may face in interpretation and performance of tasks for task-based fMRI.

Similarly, we acquired metabolite concentrations in 3 voxels, located in the basal ganglia, midfrontal grey matter (MFGM), and peritrigonal white matter in 132 children. However, in order to maintain a reasonable overall sample size (n = 88), only the MFGM data was used. Furthermore, metabolites that are not measured accurately by our SVS acquisition were not included in the predictive analyses, for example, gamma-aminobutyric acid (GABA). GABA, the brain's principal inhibitory neurotransmitter, has been associated with attention and perception in adults (Porges et al., 2017). Since we identified glutamate (Glu), the primary excitatory neurotransmitter, as one of the predictors of future cognitive performance, it is worth investigating whether GABA is associated with cognitive abilities in children. It is also possible that including metabolite concentrations from several other regions of the brain, for example, using chemical shift imaging, could be informative in the prediction of cognitive scores or categories of cognition in children. Measurement of cerebral blood perfusion (or flow, CBF) using techniques such as MR angiography (MRA) and arterial spin labelling (ASL) may also improve the prediction of cognitive functioning. MRA and ASL quantify perfusion without the need for gadolinium-based intravenous contrast agents (Thomas & Wells, 2011), and are alternatives

to the minimally invasive (i.e., use ionizing radiation) positron emission tomography (PET) and single-photon computed tomography (SPECT) measured cerebral blood flow (CBF) (Crisan et al., 2022). For example, Blokhuis and colleagues (2017) (a study we critically appraised in the systematic review, Chapter 2), found that ASL-measured CBF in grey matter is associated with executive function in control children but not in children living with HIV.

As previously mentioned, MRI is expensive. Therefore, it will require a global effort to improve MRI neuroimaging accessibility, even if focused on cheaper low-field MRI scanners or the search for biomarkers using cheaper measures of brain function like electroencephalogram and near infrared spectroscopy.

Regression and classification models: possible improvements

The main issue with our regression models was overfitting, which likely occurred because of the small sample size. Well-established techniques to prevent overfitting include implementing cross validation (CV), removing irrelevant predictors before training, regularisation to reduce variance, reducing model complexity, and adding more training data (Ying, 2019). We were limited to implementing 10-fold CV, testing ridge/lasso regularisation models, and using less complex and interpretable ML models. We have also implemented 5-fold CV, which led to comparable AUC but lower balanced sensitivity and specificity, suggesting that the higher 10-fold CV was more likely to minimise the effect of class imbalance. However, the models still learned from noise and random fluctuations in the data and could not generalise well to unseen data. Therefore, this study could have benefitted from an increased training data specifically, as a holdout data could have been used for testing and hyperparameter optimisation. Additionally, we could have removed irrelevant predictors before training in a 2-step process: (1) predicting the cognitive scores using the MRI modalities separately and obtained a subset of relevant predictors, (2) retaining only this subset in a further multimodal analysis. This can potentially lead to improved accuracy of the prediction.

The classification performance into better and poorer scorers in this study is limited to our sample, although the framework can be extended to other populations of children from low-SES households and communities. There are ways in which the classification problem could have been made simpler, for example, a median split of the data to categorise low and high performers (e.g., Girault et al. 2019). Another recent study in older patients has used this binning approach to classify cognitive performance differences based on brain network patterns (Kramer et al., 2023).A measure of underlying cognitive capacity can be obtained by performing factor analysis on cognitive scores, although a large sample would be required and decisions about class boundaries would not necessarily be any more straightforward. If the cognitive test batteries translated perfectly in the South African context, it may have been better to perform classification using the standard category splits of the test batteries (*low*, *average*, *upper average*, and *high*) instead of using clustering. This would make the classification performance more likely to be generalisable to the broader population of children and would provide more clinically useful and interpretable predictions

The ML algorithms considered here were limited to decision tree ensembles, support vector machines/regression, and penalised linear models. In general, these involved high training time and had poor prediction accuracies for responses with low sample sizes. Because of the small sample size, the tuning of parameters for the classifiers and regression models were done with the same data used for cross validation, and we chose parameter values that gave the lowest cross-validation average error. This can be another potential source of overfitting since the parameters are tweaked until the predictive models perform optimally. Furthermore, we are limited to simple ML algorithms, while deep learning structures such feed forward neural networks (FFNN) and convolutional neural networks (CNN) could improve the classification and regression problem (Liu et al., 2021). However, FFNN and CNN can be difficult to interpret due to their over-parameterised black-box nature (Li et al., 2022) and require much larger datasets.

Connectome-based predictive modelling (CPM) is another approach that has gained popularity in recent years (Sui et al., 2020). A CPM is constructed by using the interregional functional connections from whole-brain data as predictors most pertinent to the predicted target, and it has been used to predict behaviour from connectivity data according to the review by Sui and colleagues (2020).

In our classification and regression modelling, we did not directly handle multiple comparisons and false discovery rates (FDRs) as these are usually understood in the context of statistical testing rather than machine learning. We used cross-validation resampling methods to help assess the stability and reliability of the models' performances. Beyond classification and regression, there are other approaches of investigating the relationship between multimodal brain measures and cognitive performance scores. For example, canonical correlation analysis can be used to investigate which dimensions of multimodal neuroimaging are most related to which dimensions of cognitive performance as a continuous multivariate measure (Mohammadi-Nejad et al., 2017).

The dissertation has also focused on cognitive-to-cognitive and neuroimaging-to-cognitive prediction models. It can also be useful to investigate cognitive-to-neuroimaging and neuroimaging-to-neuroimaging predictive models. Although it can be useful to predict brain structure from cognitive performance since it is cheaper to administer cognitive tests, there is often not enough information available in these tests scores. Furthermore, there are also many more brain measures obtained from neuroimaging than there are cognitive domains, as such it becomes a challenge to determine which measures should be predicted. However, the neuroimaging-to-cognition prediction models can also be used to answer some of the cognitive-to-neuroimaging problems. If it is established that a subset of neuroimaging measures is related to overall cognitive performance or specific cognitive domain outcomes, and these domains are impacted i.e., cognitive deficits exist, then the parts of the brain that are affected can be predicted. Neuroimaging-to-neuroimaging prediction can be useful in determining whether aging plays a crucial role in

cognitive development. Similarly, we do not currently know whether the relationship between neuroimaging and cognitive performance at earlier than 7 years and after 9 years of age will yield the same results. Future work will investigate the cross-sectional effects in neuroimaging-to-cognitive prediction modelling.

5.3 Summary and conclusion

We have uncovered evidence in the literature suggesting perinatally-acquired HIV may lead to changes in structural, functional, and biochemical nature of the brain, and these changes are linked to cognitive domains of attention, executive function, and overall cognitive functioning. Furthermore, this evidence also showed relationships between cognitive performance and multimodal MRI of children without HIV.

We subsequently conducted a study using longitudinal MRI neuroimaging and cognitive data from children with and without HIV from low-SES backgrounds. The goal of the study was 2-fold: testing various machine learning algorithms for predicting cognitive performance from neuroimaging, and using the best classifiers and regression models to predict future cognitive outcomes of the children with and without HIV. The algorithms, namely, penalised linear regression, decision tree ensembles, and support vector machines, predicted continuous cognitive scores with low fitting errors but high generalised cross-validation errors. The classification of poorer and better performers in the sample had good prediction accuracies, and multimodal MRI was better than cognitive testing for classifying future cognitive performance. Notwithstanding limitations such as small sample size, this study suggests multimodal MRI can be a useful compliment to cognitive testing in predicting future cognitive scores. The next steps to improve this research would be to use more training data, add more timepoints, include functional MRI and chemical shift imaging in addition to the current multimodal MRI modalities, explore different ways of feature construction and implement deep learning models for regression and classification.

CITATIONS

Abo Hamza, E., Tindle, R., Pawlak, S., Bedewy, D., & Moustafa, A. A. (2024). The impact of poverty and socioeconomic status on brain, behaviour, and development: a unified framework. *Rev Neurosci*, *35*(6), 597-617. https://doi.org/10.1515/revneuro-2023-0163

Ackermann, C., Andronikou, S., Laughton, B., Kidd, M., Dobbels, E., Innes, S., van Toorn, R., & Cotton, M. (2014). White matter signal abnormalities in children with suspected HIV-related neurologic disease on early combination antiretroviral therapy. *Pediatr Infect Dis J*, *33*(8), e207-212. https://doi.org/10.1097/INF.0000000000000288

Ackermann, C., Andronikou, S., Saleh, M. G., Kidd, M., Cotton, M. F., Meintjes, E. M., & Laughton, B. (2020). Diffusion tensor imaging point to ongoing functional impairment in HIV-infected children at age 5, undetectable using standard neurodevelopmental assessments. *AIDS Res Ther*, *17*(1), 20. https://doi.org/10.1186/s12981-020-00278-z

Ackermann, C., Andronikou, S., Saleh, M. G., Laughton, B., Alhamud, A. A., van der Kouwe, A., Kidd, M., Cotton, M. F., & Meintjes, E. M. (2016). Early Antiretroviral Therapy in HIV-Infected Children Is Associated with Diffuse White Matter Structural Abnormality and Corpus Callosum Sparing. *AJNR Am J Neuroradiol*, *37*(12), 2363-2369. https://doi.org/10.3174/ajnr.A4921

Ackermann, C., van Toorn, R., & Andronikou, S. (2019). Human immunodeficiency virus-related cerebral white matter disease in children. *Pediatr Radiol*, *49*(5), 652-662. https://doi.org/10.1007/s00247-018-4310-x

Adeli, E., Kwon, D., Zhao, Q., Pfefferbaum, A., Zahr, N. M., Sullivan, E. V., & Pohl, K. M. (2018). Chained regularization for identifying brain patterns specific to HIV infection. *Neuroimage*, *183*, 425-437. https://doi.org/10.1016/j.neuroimage.2018.08.022

Adeli, E., Meng, Y., Li, G., Lin, W., & Shen, D. (2019). Multi-task prediction of infant cognitive scores from longitudinal incomplete neuroimaging data. *Neuroimage*, *185*, 783-792. https://doi.org/10.1016/j.neuroimage.2018.04.052

Agoalikum, E., Klugah-Brown, B., Wu, H., Hu, P., Jing, J., & Biswal, B. (2023). Structural differences among children, adolescents, and adults with attention-deficit/hyperactivity disorder and abnormal Granger causality of the right pallidum and whole-brain. *Front Hum Neurosci*, *17*, 1076873. https://doi.org/10.3389/fnhum.2023.1076873

Alhamud, A., Tisdall, M. D., Hess, A. T., Hasan, K. M., Meintjes, E. M., & van der Kouwe, A. J. (2012). Volumetric navigators for real-time motion correction in diffusion tensor imaging. *Magn Reson Med*, *68*(4), 1097-1108. https://doi.org/10.1002/mrm.23314

Andronikou, S., Ackermann, C., Laughton, B., Cotton, M., Tomazos, N., Spottiswoode, B., Mauff, K., & Pettifor, J. M. (2014). Correlating brain volume and callosal thickness with clinical and laboratory indicators of disease severity in children with HIV-related brain disease. *Childs Nerv Syst*, *30*(9), 1549-1557. https://doi.org/10.1007/s00381-014-2434-3

Ashby, J., Foster, C., Garvey, L., Wan, T., Allsop, J., Paramesparan, Y., Taylor-Robinson, S. D., Fidler, S., & Winston, A. (2015). Cerebral function in perinatally HIV-infected young adults and their HIV-uninfected sibling controls. *HIV Clin Trials*, *16*(2), 81-87. https://doi.org/10.1179/1528433614Z.0000000003

Awad, M., & Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer nature. https://library.oapen.org/bitstream/handle/20.500.12657/28170/1/1001824.pdf

Banker, L., & Tadi, P. (2023). Neuroanatomy, Precentral Gyrus. In *StatPearls*. https://www.ncbi.nlm.nih.gov/pubmed/31334938

Bathelt, J., Gathercole, S. E., Johnson, A., & Astle, D. E. (2018). Differences in brain morphology and working memory capacity across childhood. *Dev Sci*, *21*(3), e12579. https://doi.org/10.1111/desc.12579

Beery, K. E., & Beery, N. A. (2004). *The Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery-VMI) with supplemental Developmental tests of Visual Perception and Motor Coordination and Stepping Stones Age Norms from Birth to Age Six* (5th ed.). Pearson Assessments.

Benavente-Fernandez, I., Siddiqi, A., & Miller, S. P. (2020). Socioeconomic status and brain injury in children born preterm: modifying neurodevelopmental outcome. *Pediatr Res*, *87*(2), 391-398. https://doi.org/10.1038/s41390-019-0646-7

Bertholdo, D., Watcharakorn, A., & Castillo, M. (2013). Brain proton magnetic resonance spectroscopy: introduction and overview. *Neuroimaging Clin N Am*, *23*(3), 359-380. https://doi.org/10.1016/j.nic.2012.10.002

Blokhuis, C., Mutsaerts, H., Cohen, S., Scherpbier, H. J., Caan, M. W. A., Majoie, C., Kuijpers, T. W., Reiss, P., Wit, F., & Pajkrt, D. (2017). Higher subcortical and white matter cerebral blood flow in perinatally HIV-infected children. *Medicine (Baltimore)*, *96*(7), e5891. https://doi.org/10.1097/MD.0000000000005891

Botvinik-Nezer, R., & Wager, T. D. (2023). Reproducibility in Neuroimaging Analysis: Challenges and Solutions. *Biol Psychiatry Cogn Neurosci Neuroimaging*, *8*(8), 780-788. https://doi.org/10.1016/j.bpsc.2022.12.006

Brito, N. H., Piccolo, L. R., Noble, K. G., Pediatric Imaging, N., & Genetics, S. (2017). Associations between cortical thickness and neurocognitive skills during childhood vary by family socioeconomic factors. *Brain Cogn*, *116*, 54-62. https://doi.org/10.1016/j.bandc.2017.03.007

Burgaleta, M., Johnson, W., Waber, D. P., Colom, R., & Karama, S. (2014). Cognitive ability changes and dynamics of cortical thickness development in healthy

children and adolescents. *Neuroimage*, *84*, 810-819. https://doi.org/10.1016/j.neuroimage.2013.09.038

Bush, N. R., Wakschlag, L. S., LeWinn, K. Z., Hertz-Picciotto, I., Nozadi, S. S., Pieper, S., Lewis, J., Biezonski, D., Blair, C., Deardorff, J., Neiderhiser, J. M., Leve, L. D., Elliott, A. J., Duarte, C. S., Lugo-Candelas, C., O'Shea, T. M., Avalos, L. A., Page, G. P., & Posner, J. (2020). Family Environment, Neurodevelopmental Risk, and the Environmental Influences on Child Health Outcomes (ECHO) Initiative: Looking Back and Moving Forward. *Front Psychiatry*, *11*, 547. https://doi.org/10.3389/fpsyt.2020.00547

Buzzell, G. A., Lahat, A., & Fox, N. A. (2020). Chapter 24 - The neural correlates of cognitive control and the development of social behavior. In J. Rubenstein, P. Rakic, B. Chen, & K. Y. Kwan (Eds.), *Neural Circuit and Cognitive Development (Second Edition)* (pp. 523-538). Academic Press. https://doi.org/10.1016/B978-0-12-814411-4.00024-X

Caceres, G. A., Scambray, K. A., Malee, K., Smith, R., Williams, P. L., Wang, L., Jenkins, L. M., & Pediatric, H. I. V. A. C. S. (2024). Relationship between brain structural network integrity and emotional symptoms in youth with perinatally-acquired HIV. *Brain Behav Immun*, *116*, 101-113. https://doi.org/10.1016/j.bbi.2023.11.026

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1-27. https://doi.org/10.1080/03610927408827101

Callaway, E. (2022). Can brain scans reveal behaviour? Bombshell study says not yet. *Nature*, *603*(7903), 777-778. https://doi.org/10.1038/d41586-022-00767-3

Chaddock-Heyman, L., Weng, T. B., Kienzler, C., Erickson, K. I., Voss, M. W., Drollette, E. S., Raine, L. B., Kao, S. C., Hillman, C. H., & Kramer, A. F. (2018). Scholastic performance and functional connectivity of brain networks in children. *PLoS One*, *13*(1), e0190073. https://doi.org/10.1371/journal.pone.0190073

Chung, Y. S., Hyatt, C. J., & Stevens, M. C. (2017). Adolescent maturation of the relationship between cortical gyrification and cognitive ability. *Neuroimage*, *158*, 319-331. https://doi.org/10.1016/j.neuroimage.2017.06.082

Cleary, S., Silal, S., Birch, S., Carrara, H., Pillay-van Wyk, V., Rehle, T., & Schneider, H. (2011). Equity in the use of antiretroviral treatment in the public health care system in urban South Africa. *Health Policy*, *99*(3), 261-266. https://doi.org/10.1016/j.healthpol.2010.10.016

Cockcroft, K., & Cassimjee, N. (2020). HIV/AIDS in South African schools: Neuropsychological and psychoeducational implications. *International Journal of School and Educational Psychology*, *8*(4), 269-278. https://doi.org/10.1080/21683603.2020.1796124

Cohen, S., Caan, M. W., Mutsaerts, H. J., Scherpbier, H. J., Kuijpers, T. W., Reiss, P., Majoie, C. B., & Pajkrt, D. (2016). Cerebral injury in perinatally HIV-infected children compared to matched healthy controls. *Neurology*, *86*(1), 19-27. https://doi.org/10.1212/WNL.0000000000002209

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273-297.

Costa, V. G., & Pedreira, C. E. (2022). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, *56*, 4765–4800. https://doi.org/10.1007/s10462-022-10275-5

Cotton, M. F., Violari, A., Otwombe, K., Panchia, R., Dobbels, E., Rabie, H., Josipovic, D., Liberty, A., Lazarus, E., Innes, S., Van Rensburg, A. J., Pelser, W., Truter, H., Madhi, S. A., Handelsman, E., Jean-Philippe, P., McIntyre, J. A., Gibb, D. M., & Babiker, A. G. (2013). Early time-limited antiretroviral therapy versus deferred therapy in South African infants infected with HIV: Results from the children with HIV early antiretroviral (CHER) randomised trial. *The Lancet*, *382*(9904), 1555-1563. https://doi.org/10.1016/S0140-6736(13)61409-9

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162-173. https://doi.org/10.1006/cbmr.1996.0014

Crisan, G., Moldovean-Cioroianu, N. S., Timaru, D. G., Andries, G., Cainap, C., & Chis, V. (2022). Radiopharmaceuticals for PET and SPECT Imaging: A Literature Review over the Last Decade. *Int J Mol Sci*, *23*(9). https://doi.org/10.3390/ijms23095023

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179-1p. https://doi.org/10.1006/nimg.1998.0395

Dalmaijer, E. S., Gibbons, S. G., Bignardi, G., Anwyl-Irvine, A. L., Siugzdaite, R., Smith, T. A., Uh, S., Johnson, A., & Astle, D. E. (2023). Direct and indirect links between children's socio-economic status and education: pathways via mental health, attitude, and cognition. *Curr Psychol*, *42*(12), 9637-9651. https://doi.org/10.1007/s12144-021-02232-2

Darby, R. R., Joutsa, J., & Fox, M. D. (2019). Network localization of heterogeneous neuroimaging findings. *Brain*, *142*(1), 70-79. https://doi.org/10.1093/brain/awy292

David, W., Gerry, M., Michael, S., & Gillian, A. (2019). Relationship between childhood socioeconomic position and adverse childhood experiences (ACEs): a systematic review. *Journal of Epidemiology and Community Health*, *73*(12), 1087. https://doi.org/10.1136/jech-2019-212738

Dean, O., Buda, A., Adams, H. R., Mwanza-Kabaghe, S., Potchen, M. J., Mbewe, E. G., Kabundula, P. P., Moghaddam, S. M., Birbeck, G. L., & Bearden, D. R.

(2020). Brain Magnetic Resonance Imaging Findings Associated With Cognitive Impairment in Children and Adolescents With Human Immunodeficiency Virus in Zambia. *Pediatr Neurol*, *102*, 28-35. https://doi.org/10.1016/j.pediatrneurol.2019.08.014

DeCicco, J. M., Solomon, B., & Dennis, T. A. (2012). Neural correlates of cognitive reappraisal in children: An ERP study. *Developmental Cognitive Neuroscience*, *2*(1), 70-80. https://doi.org/10.1016/j.dcn.2011.05.009

Deshmukh, J., Jangid, M., Gupte, S., Ghosh, S., & Ingle, S. (2020). Ensemble Method Combination: Bagging and Boosting. In H. Vasudevan, A. Michalas, N. Shekokar, & M. Narvekar (Eds.), *Advanced Computing Technologies and Applications* (pp. 399-409). Springer Singapore. https://doi.org/10.1007/978-981-15-3242-9_38

Desikan, R. S., Ségonne, 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*(3), 968-980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol*, *3*(2), 185-205. https://doi.org/10.1142/s0219720005001004

Ding, L., & Pang, G. (2021). Identification of Brain Regions with Enhanced Functional Connectivity with the Cerebellum Region in Children with Attention Deficit Hyperactivity Disorder: A Resting-State fMRI Study. *International Journal of General Medicine*, *14*, 2109-2115. https://doi.org/10.2147/IJGM.S303339

Du, J., Rolls, E. T., Gong, W., Cao, M., Vatansever, D., Zhang, J., Kang, J., Cheng, W., & Feng, J. (2022). Association between parental age, brain structure, and behavioral and cognitive problems in children. *Mol Psychiatry*, *27*(2), 967-975. https://doi.org/10.1038/s41380-021-01325-5

Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos Trans R Soc Lond B Biol Sci*, *373*(1756). https://doi.org/10.1098/rstb.2017.0284

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test - 4th edition* (4th ed.). Pearson Assessments.

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L., Dodd, L., Kaplan, R., Lacombe, D., & Verweij, J. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*, *45*(2), 228-247. https://doi.org/10.1016/j.ejca.2008.10.026

Ellingsen, K. M. (2016). Standardized assessment of cognitive development: Instruments and issues. *Early childhood assessment in school and clinical child psychology*, 25-49. https://doi.org/10.1007/978-1-4939-6349-2

Ensink, R. J. H., & Kuper, H. (2017). Is hearing impairment associated with HIV? A systematic review of data from low- and middle-income countries. *Trop Med Int Health*, *22*(12), 1493-1504. https://doi.org/10.1111/tmi.12993

Fan, C., Yao, L., Zhang, J., Zhen, Z., & Wu, X. (2023). Advanced Reinforcement Learning and Its Connections with Brain Neuroscience. *Research (Wash D C)*, *6*, 0064. https://doi.org/10.34133/research.0064

Farah, R., Tzafrir, H., & Horowitz-Kraus, T. (2020). Association between diffusivity measures and language and cognitive-control abilities from early toddler's age to childhood. *Brain Struct Funct*, *225*(3), 1103-1122. https://doi.org/10.1007/s00429-020-02062-1

Firouzabadi, F. D., Ramezanpour, S., Firouzabadi, M. D., Yousem, I. J., Puts, N. A. J., & Yousem, D. M. (2022). Neuroimaging in Attention-Deficit/Hyperactivity Disorder: Recent Advances. *AJR Am J Roentgenol*, *218*(2), 321-332. https://doi.org/10.2214/AJR.21.26316

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*. https://doi.org/10.1016/S0896-6273(02)00569-X

Flensborg-Madsen, T., Falgreen Eriksen, H. L., & Mortensen, E. L. (2020). Early life predictors of intelligence in young adulthood and middle age. *PLoS One*, *15*(1), e0228144. https://doi.org/10.1371/journal.pone.0228144

Fletcher, J. M., & Miciak, J. (2017). Comprehensive Cognitive Assessments are not Necessary for the Identification and Treatment of Learning Disabilities. *Arch Clin Neuropsychol*, *32*(1), 2-7. https://doi.org/10.1093/arclin/acw103

Forsythe, S. S., McGreevey, W., Whiteside, A., Shah, M., Cohen, J., Hecht, R., Bollinger, L. A., & Kinghorn, A. (2019). Twenty Years Of Antiretroviral Therapy For People Living With HIV: Global Costs, Health Achievements, Economic Benefits. *Health Aff (Millwood)*, *38*(7), 1163-1172. https://doi.org/10.1377/hlthaff.2018.05391

Frodl, T., & Skokauskas, N. (2012). Meta-analysis of structural MRI studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatr Scand*, *125*(2), 114-126. https://doi.org/10.1111/j.1600-0447.2011.01786.x

Gabbidon, K., Chenneville, T., & Adeli, S. (2020). HIV/AIDS in Kenyan schools: Neuropsychological and psychoeducational implications. *International Journal of School and Educational Psychology*. https://doi.org/10.1080/21683603.2020.1796122

Gabis, L., Belman, A., Huang, W., Milazzo, M., & Nachman, S. (2006). Clinical and imaging study of human immunodeficiency virus-1-infected youth receiving highly active antiretroviral therapy: Pilot study using magnetic resonance spectroscopy. *JOURNAL OF CHILD NEUROLOGY*. https://doi.org/10.2310/7010.2006.00126

Gardner, R. A., & Broman, M. (1979). The Purdue Pegboard: Normative data on 1334 school children [doi:10.1080/15374417909532912]. *Journal of Clinical Child Psychology*, *8*, 156-162. https://doi.org/10.1080/15374417909532912

Gartlehner, G., Affengruber, L., Titscher, V., Noel-Storr, A., Dooley, G., Ballarini, N., & König, F. (2020). Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2020.01.005

Gaus, R., Polsterl, S., Greimel, E., Schulte-Korne, G., & Wachinger, C. (2023). Can we diagnose mental disorders in children? A large-scale assessment of machine learning on structural neuroimaging of 6916 children in the adolescent brain cognitive development study. *JCPP Adv*, *3*(4), e12184. https://doi.org/10.1002/jcv2.12184

Geethanath, S., & Vaughan, J. T., Jr. (2019). Accessible magnetic resonance imaging: A review. *J Magn Reson Imaging*, *49*(7), e65-e77. https://doi.org/10.1002/jmri.26638

Girault, J. B., Munsell, B. C., Puechmaille, D., Goldman, B. D., Prieto, J. C., Styner, M., & Gilmore, J. H. (2019). White matter connectomes at birth accurately predict cognitive abilities at age 2. *Neuroimage*, *192*, 145-155. https://doi.org/10.1016/j.neuroimage.2019.02.060

Gratton, C., Nelson, S. M., & Gordon, E. M. (2022). Brain-behavior correlations: Two paths toward reliability. *Neuron*, *110*(9), 1446-1449. https://doi.org/10.1016/j.neuron.2022.04.018

Greenwood, C. J., Youssef, G. J., Letcher, P., Macdonald, J. A., Hagg, L. J., Sanson, A., McIntosh, J., Hutchinson, D. M., Toumbourou, J. W., Fuller-Tyszkiewicz, M., & Olsson, C. A. (2020). A comparison of penalised regression methods for informing the selection of predictive markers. *PLoS One*, *15*(11), e0242730. https://doi.org/10.1371/journal.pone.0242730

Gregory, M. D., Kippenhan, J. S., Dickinson, D., Carrasco, J., Mattay, V. S., Weinberger, D. R., & Berman, K. F. (2016). Regional Variations in Brain Gyrification Are Associated with General Cognitive Ability in Humans. *Curr Biol*, *26*(10), 1301-1305. https://doi.org/10.1016/j.cub.2016.03.021

Haase, A., Frahm, J., Hanicke, W., & Matthaei, D. (1985). 1H NMR chemical shift selective (CHESS) imaging. *Physics in Medicine and Biology*, *30*(4), 341. https://doi.org/10.1088/0031-9155/30/4/008

Habeck, C., Gazes, Y., Razlighi, Q., & Stern, Y. (2020). Cortical thickness and its associations with age, total cognition and education across the adult lifespan. *PLoS One*, *15*(3), e0230298. https://doi.org/10.1371/journal.pone.0230298

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysi* (1 ed.). Springer New York, NY. https://doi.org/10.1007/978-1-4757-3462-1

Harris, M., Franzsen, D., & de Witt, P. A. (2021). Relevance of Norms and Psychometric Properties of Three Standardised Visual Perceptual Tests for Children Attending Mainstream Schools in Gauteng. *South African Journal of Occupational Therapy*, *51*, 4-13. https://doi.org/10.17159/2310-3833/2021/vol51n3a2

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues Clin Neurosci*, *14*(1), 91-99. https://www.ncbi.nlm.nih.gov/pubmed/22577308

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (1 ed.). Springer New York, NY. https://doi.org/10.1007/978-0-387-21606-5

Heany, S. J., Phillips, N., Brooks, S., Fouche, J. P., Myer, L., Zar, H., Stein, D. J., & Hoare, J. (2020). Neural correlates of maintenance working memory, as well as relevant structural qualities, are associated with earlier antiretroviral treatment initiation in vertically transmitted HIV. *J Neurovirol*, *26*(1), 60-69. https://doi.org/10.1007/s13365-019-00792-5

Herting, M. M., Uban, K. A., Williams, P. L., Gautam, P., Huo, Y., Malee, K., Yogev, R., Csernansky, J., Wang, L., Nichols, S., Van Dyke, R., & Sowell, E. R. (2015). Default Mode Connectivity in Youth With Perinatally Acquired HIV. *Medicine (Baltimore)*, *94*(37), e1417. https://doi.org/10.1097/MD.0000000000001417

Hess, A. T., Tisdall, M. D., Andronesi, O. C., Meintjes, E. M., & van der Kouwe, A. J. W. (2011). Real-time motion and B0 corrected single voxel spectroscopy using volumetric navigators. *Magnetic Resonance in Medicine*. https://doi.org/10.1002/mrm.22805

Hoare, J., Fouche, J. P., Phillips, N., Heany, S. J., Myer, L., Zar, H. J., & Stein, D. J. (2021). Alcohol use is associated with mental health problems and brain structural alterations in adolescents with perinatally acquired HIV infection on ART. *Alcohol*, *97*, 59-66. https://doi.org/10.1016/j.alcohol.2021.09.006

Hoare, J., Fouche, J. P., Phillips, N., Joska, J. A., Donald, K. A., Thomas, K., & Stein, D. J. (2015). Clinical associations of white matter damage in cART-treated HIV-positive children in South Africa. *J Neurovirol*, *21*(2), 120-128. https://doi.org/10.1007/s13365-014-0311-1

Hoare, J., Fouche, J. P., Spottiswoode, B., Donald, K., Philipps, N., Bezuidenhout, H., Mulligan, C., Webster, V., Oduro, C., Schrieff, L., Paul, R., Zar, H., Thomas, K., & Stein, D. (2012). A diffusion tensor imaging and neurocognitive study of

HIV-positive children who are HAART-naïve "slow progressors". *Journal of NeuroVirology*, *18*, 205-212. https://doi.org/10.1007/s13365-012-0099-9

Hoare, J., Heany, S. J., Fouche, J. P., Phillips, N., Joska, J. A., Myer, L., Zar, H. J., & Stein, D. J. (2019). Initiation of antiretroviral therapy after the critical neuronal developmental period of the second postnatal year affects white matter microstructure in adolescents living with HIV. *Journal of NeuroVirology*, *25*, 254–262. https://doi.org/10.1007/s13365-018-0712-7

Hoare, J., Ransford, G. L., Phillips, N., Amos, T., Donald, K., & Stein, D. J. (2014). Systematic review of neuroimaging studies in vertically transmitted HIV positive children and adolescents. *METABOLIC BRAIN DISEASE*, *29*, 221–229. https://doi.org/10.1007/s11011-013-9456-5

Holmes, M. J., Robertson, F. C., Little, F., Randall, S. R., Cotton, M. F., Van Der Kouwe, A. J. W., Laughton, B., & Meintjes, E. M. (2017). Longitudinal increases of brain metabolite levels in 5-10 year old children. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0180973

Howieson, D. (2019). Current limitations of neuropsychological tests and assessment procedures. *Clin Neuropsychol*, *33*(2), 200-208. https://doi.org/10.1080/13854046.2018.1552762

Hutton, J. S., Dudley, J., Horowitz-Kraus, T., DeWitt, T., & Holland, S. K. (2020). Associations between home literacy environment, brain white matter integrity and cognitive abilities in preschool-age children. *Acta Paediatr*, *109*(7), 1376-1386. https://doi.org/10.1111/apa.15124

Ito, R., Nakae, K., Hata, J., Okano, H., & Ishii, S. (2019). Semi-supervised deep learning of brain tissue segmentation. *Neural Netw*, *116*, 25-34. https://doi.org/10.1016/j.neunet.2019.03.014

Jain, R., & Xu, W. (2021). HDSI: High dimensional selection with interactions algorithm on feature selection and testing. *PLoS One*, *16*(2), e0246159. https://doi.org/10.1371/journal.pone.0246159

Jankiewicz, M., Holmes, M. J., Taylor, P. A., Cotton, M. F., Laughton, B., van der Kouwe, A. J. W., & Meintjes, E. M. (2017). White Matter Abnormalities in Children with HIV Infection and Exposure. *FRONTIERS IN NEUROANATOMY*, *11*. https://doi.org/10.3389/fnana.2017.00088

Johnson-Greene, D. (2013). Secondary Influences on Neuropsychological Test Performance. *Archives of Clinical Neuropsychology*, *28*(8), 866-867. https://doi.org/10.1093/arclin/act063

Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivières, 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*. https://doi.org/10.1016/j.neuroimage.2019.05.082

Karama, S., Colom, R., Johnson, W., Deary, I. J., Haier, R., Waber, D. P., Lepage, C., Ganjavi, H., Jung, R., & Evans, A. C. (2011). Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. *Neuroimage*, *55*(4), 1443-1453. https://doi.org/10.1016/j.neuroimage.2011.01.016

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Assessment Battery for Children, Second Edition*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118660584.ese1324

Khaire, U. M., & Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*, *34*(4), 1060-1073. https://doi.org/10.1016/j.jksuci.2019.06.012

Khobo, I. L., Jankiewicz, M., Holmes, M. J., Little, F., Cotton, M. F., Laughton, B., van der Kouwe, A. J. W., Moreau, A., Nwosu, E., Meintjes, E. M., & Robertson, F. C. (2022). Multimodal magnetic resonance neuroimaging measures characteristic of early cART-treated pediatric HIV: A feature selection approach. *Hum Brain Mapp*, *43*(13), 4128-4144. https://doi.org/10.1002/hbm.25907

Khundrakpam, B., Booij, L., Jeon, S., Karama, S., Tohka, J., & Evans, A. C. (2022). Individualized prediction of future cognition based on developmental changes in cortical anatomy. *Neuroimage: Reports*, *2*(4), 100127. https://doi.org/https://doi.org/10.1016/j.ynirp.2022.100127

Khundrakpam, B. S., Lewis, J. D., Reid, A., Karama, S., Zhao, L., Chouinard-Decorte, F., Evans, A. C., & Brain Development Cooperative, G. (2017). Imaging structural covariance in the development of intelligence. *Neuroimage*, *144*(Pt A), 227-240. https://doi.org/10.1016/j.neuroimage.2016.08.041

Kline, J. E., Sita Priyanka Illapani, V., He, L., & Parikh, N. A. (2020). Automated brain morphometric biomarkers from MRI at term predict motor development in very preterm infants. *NeuroImage: Clinical*, *28*, 102475. https://doi.org/10.1016/j.nicl.2020.102475

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Articial Intelligence (IJCAI)*, *14*(2), 1137-1145. https://doi.org/10.1067/mod.2000.109031

Kong, X.-Z., Zhang, C., Liu, Y., & Pu, Y. (2022). Scanning reproducible brain-wide associations: sample size is all you need? *Psychoradiology*, *2*(3), 67-68. https://doi.org/10.1093/psyrad/kkac010

Kramer, C., Stumme, J., da Costa Campos, L., Rubbert, C., Caspers, J., Caspers, S., & Jockwitz, C. (2023). Classification and prediction of cognitive performance differences in older age based on brain network patterns using a machine learning approach. *Netw Neurosci*, *7*(1), 122-147. https://doi.org/10.1162/netn_a_00275

Lanka, P., Rangaprakash, D., Dretsch, M. N., Katz, J. S., Denney, T. S., Jr., & Deshpande, G. (2020). Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. *Brain Imaging Behav*, *14*(6), 2378-2416. https://doi.org/10.1007/s11682-019-00191-8

Laughton, B., Cornell, M., Boivin, M., & Van Rie, A. (2013). Neurodevelopment in perinatally HIV-infected children: a concern for adolescence. *Journal of the International AIDS Society*, *16*, 18603. https://doi.org/https://doi.org/10.7448/IAS.16.1.18603

Laughton, B., Cornell, M., Grove, D., Kidd, M., Springer, P. E., Dobbels, E., Van Rensburg, A. J., Violari, A., Babiker, A. G., Madhi, S. A., Jean-Philippe, P., Gibb, D. M., & Cotton, M. F. (2012). Early antiretroviral therapy improves neurodevelopmental outcomes in infants. *AIDS*. https://doi.org/10.1097/QAD.0b013e328355d0ce

Laughton, B., Cornell, M., Kidd, M., Springer, P. E., Dobbels, E. F. M. T., Rensburg, A. J. V., Otwombe, K., Babiker, A., Gibb, D. M., Violari, A., Kruger, M., & Cotton, M. F. (2018). Five year neurodevelopment outcomes of perinatally HIV-infected children on early limited or deferred continuous antiretroviral therapy. *Journal of the International AIDS Society*. https://doi.org/10.1002/jia2.25106

Leark, R. A., Wallace, D. R., & Fitzgerald, R. (2004). Test-retest reliability and standard error of measurement for the Test of Variables of Attention (T.O.V.A.) with healthy school-age children. *Assessment*. https://doi.org/10.1177/1073191104269186

Lee, S. J., Steiner, R. J., Yu, Y., Short, S. J., Neale, M. C., Styner, M. A., Zhu, H., & Gilmore, J. H. (2017). Common and heritable components of white matter microstructure predict cognitive function at 1 and 2 y. *Proc Natl Acad Sci U S A*, *114*(1), 148-153. https://doi.org/10.1073/pnas.1604658114

Lewis-de los Angeles, C. P., Williams, P. L., Huo, Y., Wang, S. D., Uban, K. A., Herting, M. M., Malee, K., Yogev, R., Csernansky, J. G., Nichols, S., Van Dyke, R. B., Sowell, E. R., Wang, L., Pediatric, H. I. V. A. C. S., the Pediatric Imaging, N., & Genetics, S. (2017). Lower total and regional grey matter brain volumes in youth with perinatally-acquired HIV infection: Associations with HIV disease severity, substance use, and cognition. *Brain Behav Immun*, *62*, 100-109. https://doi.org/10.1016/j.bbi.2017.01.004

Li, J., Gao, L., & Ye, Z. (2021). Study of Brain Structure in HIV Vertically Infected Adolescents. *AIDS Res Hum Retroviruses*, *37*(9), 647-656. https://doi.org/10.1089/AID.2020.0030

Li, T., Wang, X., Luo, T., Yang, Y., Zhao, B., Yang, L., Zhu, Z., & Zhu, H. (2019). Adolescent fluid intelligence prediction from regional brain volumes and cortical curvatures using BlockPC-XGBoost. In *Adolescent Brain Cognitive Development Neurocognitive Prediction: First Challenge, ABCD-NP 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019,*

Proceedings 1 (pp. 167-175). Springer, Cham. https://doi.org/10.1007/978-3-030-31901-4_20

Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., & Dou, D. (2022). Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, *64*(12), 3197-3234. https://doi.org/10.1007/s10115-022-01756-8

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel)*, *23*(1). https://doi.org/10.3390/e23010018

Linn, K. A., Gaonkar, B., Doshi, J., Davatzikos, C., & Shinohara, R. T. (2016). Addressing Confounding in Predictive Models with an Application to Neuroimaging. *International Journal of Biostatistics*. https://doi.org/10.1515/ijb-2015-0030

Liu, W. K., Gan, Z., Fleming, M., Liu, W. K., Gan, Z., & Fleming, M. (2021). Deep Learning for Regression and Classification. In *Mechanistic Data Science for STEM Education and Applications* (pp. 171-214). Springer, Cham. https://doi.org/10.1007/978-3-030-87832-0_6

Lovden, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and Cognitive Functioning Across the Life Span. *Psychol Sci Public Interest*, *21*(1), 6-41. https://doi.org/10.1177/1529100620920576

Madhi, Shabir A., Adrian, P., Cotton, Mark F., McIntyre, James A., Jean-Philippe, P., Meadows, S., Nachman, S., Käyhty, H., Klugman, Keith P., & Violari, A. (2010). Effect of HIV Infection Status and Anti-Retroviral Treatment on Quantitative and Qualitative Antibody Responses to Pneumococcal Conjugate Vaccine in Infants. *The Journal of Infectious Diseases*. https://doi.org/10.1086/653704

Madzime, J., Holmes, M., Cotton, M. F., Laughton, B., van der Kouwe, A. J. W., Meintjes, E. M., & Jankiewicz, M. (2022). Altered White Matter Tracts in the Somatosensory, Salience, Motor, and Default Mode Networks in 7-Year-Old Children Living with Human Immunodeficiency Virus: A Tractographic Analysis. *Brain Connect*, *12*(4), 302-319. https://doi.org/10.1089/brain.2020.0948

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., . . . Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654-660. https://doi.org/10.1038/s41586-022-04492-9

Martin Bejarano-Garcia, M., Ruiz-Saez, B., Zamora, B., Martinez de Aragon, A., Garcia-Navarro, C., Jimenez-de Ory, S., Velo, C., Ramos, J. T., Sainz, T., Escosa, L., Nunez-Enamorado, N., Stephan-Otto, C., Navarro, M. L., &

Gonzalez-Tome, M. I. (2021). Brain activity in well-controlled perinatally HIV-infected young adults: a fMRI pilot study. *Rev Neurol*, *72*(10), 343-351. https://doi.org/10.33588/rn.7210.2020536 (Actividad cerebral en jovenes infectados por el virus de la inmunodeficiencia humana por transmision vertical: estudio piloto de resonancia magnetica funcional.)

Martinez-Heras, E., Grussu, F., Prados, F., Solana, E., & Llufriu, S. (2021). Diffusion-Weighted Imaging: Recent Advances and Applications. *Semin Ultrasound CT MR*, *42*(5), 490-506. https://doi.org/10.1053/j.sult.2021.07.006

Mbugua, K. K., Holmes, M. J., Cotton, M. F., Ratai, E. M., Little, F., Hess, A. T., Dobbels, E., Van Der Kouwe, A. J. W., Laughton, B., & Meintjes, E. M. (2016). HIV-associated CD4+/CD8+ depletion in infancy is associated with neurometabolic reductions in the basal ganglia at age 5 years despite early antiretroviral therapy. *AIDS*. https://doi.org/10.1097/QAD.0000000000001082

McDermott, C. L., Seidlitz, J., Nadig, A., Liu, S., Clasen, L. S., Blumenthal, J. D., Reardon, P. K., Lalonde, F., Greenstein, D., Patel, R., Chakravarty, M. M., Lerch, J. P., & Raznahan, A. (2019). Longitudinally Mapping Childhood Socioeconomic Status Associations with Cortical and Subcortical Morphology. *J Neurosci*, *39*(8), 1365-1373. https://doi.org/10.1523/JNEUROSCI.1808-18.2018

McMaughan, D. J., Oloruntoba, O., & Smith, M. L. (2020). Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. *Front Public Health*, *8*, 231. https://doi.org/10.3389/fpubh.2020.00231

Meyer, A. L. (2022). Need to Revise Frascati Criteria for HIV-Associated Neurocognitive Disorders to Improve Relevance for Diverse Global Populations. *Neurol Clin Pract*, *12*(4), 328-330. https://doi.org/10.1212/CPJ.0000000000200063

Meyer, M., Bekkering, H., Janssen, D. J. C., de Bruijn, E. R. A., & Hunnius, S. (2014). Neural Correlates of Feedback Processing in Toddlers. *Journal of Cognitive Neuroscience*, *26*(7), 1519-1527. https://doi.org/10.1162/jocn_a_00560

Mihalik, A., Brudfors, M., Robu, M., Ferreira, F. S., Lin, H., Rau, A., Wu, T., Blumberg, S. B., Kanber, B., & Tariq, M. (2019). ABCD neurocognitive prediction challenge 2019: Predicting individual fluid intelligence scores from structural MRI using probabilistic segmentation and kernel ridge regression. In *Lecture NoteLecture Notes in Computer Sciences in Computer Science* (pp. 133-142). Springer, Cham. https://doi.org/10.1007/978-3-030-31901-4_16

Mills, K. L., Siegmund, K. D., Tamnes, C. K., Ferschmann, L., Wierenga, L. M., Bos, M. G. N., Luna, B., Li, C., & Herting, M. M. (2021). Inter-individual variability in structural brain development from late childhood to young adulthood. *NeuroImage*, *242*, 118450. https://doi.org/10.1016/j.neuroimage.2021.118450

Mitchell, J. M., Tomlinson, M., Bland, R. M., Houle, B., Stein, A., & Rochat, T. J. (2017). Confirmatory factor analysis of the Kaufman assessment battery in a

sample of primary school-aged children in rural South Africa. *South African Journal of Psychology*, *48*(4), 434-452. https://doi.org/10.1177/0081246317741822

Mohammadi-Nejad, A. R., Hossein-Zadeh, G. A., & Soltanian-Zadeh, H. (2017). Structured and Sparse Canonical Correlation Analysis as a Brain-Wide Multi-Modal Data Fusion Approach. *IEEE Trans Med Imaging*, *36*(7), 1438-1448. https://doi.org/10.1109/TMI.2017.2681966

Molnar, C. (2020). *Interpretable machine learning*. Leanpub. https://books.google.com/books?hl=en&lr=&id=jBm3DwAAQBAJ&oi=fnd&pg=PP1&dq=Molnar,+C.+(2020).+Interpretable+machine+learning&ots=EhrUUpGCV5&sig=2pU0VaiguS1iDGEBVPKMzETansk

Mori, S. (2007). Human White Matter Atlas. *American Journal of Psychiatry*, *164*(7), 1005-1005. https://doi.org/10.1176/appi.ajp.164.7.1005

Muheriwa Matemba, S. R., Cianelli, R., McCabe, B. E., & Chilemba, E. (2021). Health-related interventions for youth living with perinatally-acquired HIV in sub-Saharan Africa: A systematic review of the evidence. *Journal of HIV/AIDS & Social Services*, *20*(1), 1-32. https://doi.org/10.1080/15381501.2021.1900975

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(1), 86-97. https://doi.org/10.1002/widm.53

Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: an overview, II. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(6), e1219. https://doi.org/10.1002/widm.1219

Musielak, K. A., & Fine, J. G. (2016). An Updated Systematic Review of Neuroimaging Studies of Children and Adolescents with Perinatally Acquired HIV. *Journal of Pediatric Neuropsychology*. https://doi.org/10.1007/s40817-015-0009-1

Musindo, O., Krabbendam, L., Mutahi, J., Garcia, M. P., Bangirana, P., & Kumar, M. (2022). Neurocognitive deficits and socioeconomic risk factors among children and adolescents living with HIV in sub-Saharan Africa: a systematic review. *Child Adolesc Psychiatry Ment Health*, *16*(1), 31. https://doi.org/10.1186/s13034-022-00465-y

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A Review of Feature Reduction Techniques in Neuroimaging. *Neuroinformatics*, *12*, 229-244. https://doi.org/10.1007/s12021-013-9204-3

Na, S. D., & Burns, T. G. (2016). Wechsler intelligence scale for children-V: Test review. *Applied Neuropsychology: Child*, *5*(2), 156-160. https://doi.org/10.1080/21622965.2015.1015337

Nagarajan, R., Sarma, M. K., Thomas, M. A., Chang, L., Natha, U., Wright, M., Hayes, J., Nielsen-Saines, K., Michalik, D. E., Deville, J., Church, J. A., Mason, K., Critton-Mastandrea, T., Nazarian, S., Jing, J., & Keller, M. A. (2012).

Neuropsychological function and cerebral metabolites in HIV-infected youth. *J Neuroimmune Pharmacol*, *7*(4), 981-990. https://doi.org/10.1007/s11481-012-9407-7

Nalwanga, D., & Musiime, V. (2022). Children living with HIV: a narrative review of recent advances in pediatric HIV research and their implications for clinical practice. *Ther Adv Infect Dis*, *9*, 20499361221077544. https://doi.org/10.1177/20499361221077544

Nichols, S. L. (2022). Central Nervous System Impact of Perinatally Acquired HIV in Adolescents and Adults: an Update. *Curr HIV/AIDS Rep*, *19*(1), 121-132. https://doi.org/10.1007/s11904-021-00598-3

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. *Biol Psychiatry Cogn Neurosci Neuroimaging*, *5*(8), 791-798. https://doi.org/10.1016/j.bpsc.2019.11.007

Novotny, E., Ashwal, S., & Shevell, M. (1998). Proton magnetic resonance spectroscopy: an emerging technology in pediatric neurology research. *Pediatr Res*, *44*(1), 1-10. https://doi.org/10.1203/00006450-199807000-00001

Nozyce, M. L., Lee, S. S., Wiznia, A., Nachman, S., Mofenson, L. M., Smith, M. E., Yogev, R., McIntosh, K., Stanley, K., & Pelton, S. (2006). A behavioral and cognitive profile of clinically stable HIV-infected children. *Pediatrics*, *117*(3), 763-770. https://doi.org/10.1542/peds.2005-0451

Nwosu, E. C., Holmes, M. J., Cotton, M. F., Dobbels, E., Little, F., Laughton, B., van der Kouwe, A., Meintjes, E. M., & Robertson, F. (2021). Cortical structural changes related to early antiretroviral therapy (ART) interruption in perinatally HIV-infected children at 5 years of age. *IBRO Neuroscience Reports*. https://doi.org/10.1016/j.ibneur.2021.02.001

Nwosu, E. C., Robertson, F. C., Holmes, M. J., Cotton, M. F., Dobbels, E., Little, F., Laughton, B., van der Kouwe, A., & Meintjes, E. M. (2018). Altered brain morphometry in 7-year old HIV-infected children on early ART. *Metabolic Brain Disease*. https://doi.org/10.1007/s11011-017-0162-6

Oberman, H. I., & Vink, G. (2023). Toward a standardized evaluation of imputation methodology. *Biom J*, e2200107. https://doi.org/10.1002/bimj.202200107

Oeltzschner, G., Wijtenburg, S. A., Mikkelsen, M., Edden, R. A. E., Barker, P. B., Joo, J. H., Leoutsakos, J.-M. S., Rowland, L. M., Workman, C. I., & Smith, G. S. (2019). Neurometabolites and associations with cognitive deficits in mild cognitive impairment: a magnetic resonance spectroscopy study at 7 Tesla. *Neurobiology of Aging*, *73*, 211-218. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2018.09.027

Olson, I. R., Von Der Heide, R. J., Alm, K. H., & Vyas, G. (2015). Development of the uncinate fasciculus: Implications for theory and developmental disorders. *Dev Cogn Neurosci*, *14*, 50-61. https://doi.org/10.1016/j.dcn.2015.06.003

Ouyang, M., Peng, Q., Jeon, T., Heyne, R., Chalak, L., & Huang, H. (2020). Diffusion-MRI-based regional cortical microstructure at birth for predicting neurodevelopmental outcomes of 2-year-olds. *Elife*, *9*. https://doi.org/10.7554/eLife.58116

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*. https://doi.org/10.1186/s13643-016-0384-4

Oztekin, I., Finlayson, M. A., Graziano, P. A., & Dick, A. S. (2021). Is there any incremental benefit to conducting neuroimaging and neurocognitive assessments in the diagnosis of ADHD in young children? A machine learning investigation. *Dev Cogn Neurosci*, *49*, 100966. https://doi.org/10.1016/j.dcn.2021.100966

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*, n71 https://doi.org/10.1136/bmj.n71

Pan, J., Sawyer, K., McDonough, E., Slotpole, L., & Gansler, D. (2018). Cognitive, Neuroanatomical, and Genetic Predictors of Executive Function in Healthy Children and Adolescents. *Dev Neuropsychol*, *43*(7), 535-550. https://doi.org/10.1080/87565641.2018.1516770

Pat, N., Wang, Y., Bartonicek, A., Candia, J., & Stringaris, A. (2022). Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cereb Cortex*. https://doi.org/10.1093/cercor/bhac235

Paul, R., Cooley, S. A., Garcia-Egan, P. M., & Ances, B. M. (2018). Cognitive Performance and Frailty in Older HIV-Positive Adults. *J Acquir Immune Defic Syndr*, *79*(3), 375-380. https://doi.org/10.1097/QAI.0000000000001790

Paul, R., Prasitsuebsai, W., Jahanshad, N., Puthanakit, T., Thompson, P., Aurpibul, L., Hansudewechakul, R., Kosalaraksa, P., Kanjanavanit, S., Ngampiyaskul, C., Luesomboon, W., Lerdlum, S., Pothisri, M., Visrutaratna, P., Valcour, V., Nir, T. M., Saremi, A., Kerr, S., Ananworanich, J., . . . Thailand Study, G. (2018). Structural Neuroimaging and Neuropsychologic Signatures in Children With Vertically Acquired HIV. *Pediatr Infect Dis J*, *37*(7), 662-668. https://doi.org/10.1097/INF.0000000000001852

Paul, R. H., Cho, K., Belden, A., Carrico, A. W., Martin, E., Bolzenius, J., Luckett, P., Cooley, S. A., Mannarino, J., Gilman, J. M., Miano, M., & Ances, B. M. (2022). Cognitive Phenotypes of HIV Defined Using a Novel Data-driven Approach. *J Neuroimmune Pharmacol*, *17*(3-4), 515-525. https://doi.org/10.1007/s11481-021-10045-0

Paulus, M. P., Kuplicki, R., & Yeh, H. W. (2019). Machine Learning and Brain Imaging: Opportunities and Challenges. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2019.07.007

Peng, P., & Kievit, R. A. (2020). The Development of Academic Achievement and Cognitive Abilities: A Bidirectional Perspective. *Child Dev Perspect*, *14*(1), 15-20. https://doi.org/10.1111/cdep.12352

Phillips, N., Amos, T., Kuo, C., Hoare, J., Ipser, J., Thomas, K. G., & Stein, D. J. (2016). HIV-Associated Cognitive Impairment in Perinatally Infected Children: A Meta-analysis. *Pediatrics*, *138*(5). https://doi.org/10.1542/peds.2016-0893

Pierpaoli, C., & Walker, L. (2010). *TORTOISE: an integrated software package for processing of diffusion MRI data* ISMRM 18th annual meeting, Stockholm, Sweden. https://www.nichd.nih.gov/sites/default/files/inline-files/TORTOISE_MRI_software%20_package.pdf

Plaia, A., Buscemi, S., Fürnkranz, J., & Mencía, E. L. (2022). Comparing Boosting and Bagging for Decision Trees of Rankings. *Journal of Classification*, *39*(1), 78-99. https://doi.org/10.1007/s00357-021-09397-2

Porges, E. C., Woods, A. J., Edden, R. A., Puts, N. A., Harris, A. D., Chen, H., Garcia, A. M., Seider, T. R., Lamb, D. G., Williamson, J. B., & Cohen, R. A. (2017). Frontal Gamma-Aminobutyric Acid Concentrations Are Associated With Cognitive Performance in Older Adults. *Biol Psychiatry Cogn Neurosci Neuroimaging*, *2*(1), 38-44. https://doi.org/10.1016/j.bpsc.2016.06.004

Provencher, S. W. (2001). Automatic quantitation of localized in vivo1H spectra with LCModel. *NMR in Biomedicine*. https://doi.org/10.1002/nbm.698

Raja, R., Na, X., Moore, A., Otoo, R., Glasier, C. M., Badger, T. M., & Ou, X. (2022). Associations Between White Matter Microstructures and Cognitive Functioning in 8-Year-Old Children: A Track-Weighted Imaging Study. *J Child Neurol*, *37*(6), 471-490. https://doi.org/10.1177/08830738221083487

Rajput, D., Wang, W. J., & Chen, C. C. (2023). Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics*, *24*(1), 48. https://doi.org/10.1186/s12859-023-05156-9

Rakesh, D., Whittle, S., Sheridan, M. A., & McLaughlin, K. A. (2023). Childhood socioeconomic status and the pace of structural neurodevelopment: accelerated, delayed, or simply different? *Trends Cogn Sci*. https://doi.org/10.1016/j.tics.2023.03.011

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S., Ellis, C., Shakeshaft, C., Seghier, M. L., & Price, C. J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature*, *479*(7371), 113-116. https://doi.org/10.1038/nature10514

Randall, S. R., Warton, C. M. R., Holmes, M. J., Cotton, M. F., Laughton, B., van der Kouwe, A. J. W., & Meintjes, E. M. (2017). Larger subcortical gray matter structures and smaller corpora callosa at age 5 years in HIV infected children on

early ART. *Frontiers in Neuroanatomy*. https://doi.org/10.3389/fnana.2017.00095

Rasero, J., Sentis, A. I., Yeh, F. C., & Verstynen, T. (2021). Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLoS Comput Biol*, *17*(3), e1008347. https://doi.org/10.1371/journal.pcbi.1008347

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, *61*(4), 1402-1418. https://doi.org/10.1016/j.neuroimage.2012.02.084

Robertson, F. C., Holmes, M. J., Cotton, M. F., Dobbels, E., Little, F., Laughton, B., van der Kouwe, A. J. W., & Meintjes, E. M. (2018). Perinatal HIV Infection or Exposure Is Associated With Low N-Acetylaspartate and Glutamate in Basal Ganglia at Age 9 but Not 7 Years. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2018.00145

Rokach, L. (2019). Ensemble learning: pattern classification using ensemble methods. In *Machine Perception and Artificial Intelligence*. World Scientific. https://doi.org/10.1142/9789811201967_0001

Rosenberg, M. D., & Finn, E. S. (2022). How to establish robust brain-behavior relationships without thousands of individuals. *Nat Neurosci*, *25*(7), 835-837. https://doi.org/10.1038/s41593-022-01110-9

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7

Rowe, K., Buivydaite, R., Heinsohn, T., Rahimzadeh, M., Wagner, R. G., Scerif, G., & Stein, A. (2021). Executive function in HIV-affected children and adolescents: a systematic review and meta-analyses. *AIDS Care*, *33*(7), 833-857. https://doi.org/10.1080/09540121.2021.1873232

Santana, C. P., de Carvalho, E. A., Rodrigues, I. D., Bastos, G. S., de Souza, A. D., & de Brito, L. L. (2022). rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis. *Sci Rep*, *12*(1), 6030. https://doi.org/10.1038/s41598-022-09821-6

Sarma, M. K., Pal, A., Keller, M. A., Welikson, T., Ventura, J., Michalik, D. E., Nielsen-Saines, K., Deville, J., Kovacs, A., Operskalski, E., Church, J. A., Macey, P. M., Biswal, B., & Thomas, M. A. (2021). White matter of perinatally HIV infected older youths shows low frequency fluctuations that may reflect glial cycling. *Sci Rep*, *11*(1), 3086. https://doi.org/10.1038/s41598-021-82587-5

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664-681. https://doi.org/https://doi.org/10.1016/j.neucom.2017.06.053

Schadl, K., Vassar, R., Cahill-Rowley, K., Yeom, K. W., Stevenson, D. K., & Rose, J. (2018). Prediction of cognitive and motor development in preterm children using exhaustive feature selection and cross-validation of near-term white matter microstructure. *NeuroImage: Clinical*, *17*, 667-679. https://doi.org/10.1016/j.nicl.2017.11.023

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M., Salehi, M., Gao, S., Shen, X., O'Connor, D., & Barron, D. S. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage*, *193*, 35-45. https://doi.org/10.1016/j.neuroimage.2019.02.057

Scheinost, D., Pollatou, A., Dufford, A. J., Jiang, R., Farruggia, M. C., Rosenblatt, M., Peterson, H., Rodriguez, R. X., Dadashkarimi, J., Liang, Q., Dai, W., Foster, M. L., Camp, C. C., Tejavibulya, L., Adkinson, B. D., Sun, H., Ye, J., Cheng, Q., Spann, M. N., . . . Westwater, M. L. (2022). Machine learning and prediction in fetal, infant, and toddler neuroimaging: a review and primer. *Biological Psychiatry*. https://doi.org/https://doi.org/10.1016/j.biopsych.2022.10.014

Schneider, C. L., Mohajeri-Moghaddam, S., Mbewe, E. G., Kabundula, P. P., Dean, O., Buda, A., Potchen, M. J., Mwanza-Kabaghe, S., Saylor, D., Adams, H. R., Birbeck, G. L., & Bearden, D. R. (2020). Cerebrovascular Disease in Children Perinatally Infected With Human Immunodeficiency Virus in Zambia. *Pediatr Neurol*, *112*, 14-21. https://doi.org/10.1016/j.pediatrneurol.2020.08.003

Scholkopf, B., & Smola, A. J. (2018). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press. https://books.google.com/books?hl=en&lr=&id=7r34DwAAQBAJ&oi=fnd&pg=PR13&dq=Learning+with+kernels:+support+vector+machines,+regularization,+optimization,+and+beyond&ots=pxL1vflQSS&sig=pNXxeRkoVXHZTV5JEmxqbTbDUv8

Schölkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press. https://books.google.com/books?hl=en&lr=&id=y8ORL3DWt4sC&oi=fnd&pg=PR13&dq=Learning+with+kernels:+support+vector+machines,+regularization,+optimization,+and+beyond&ots=bNsV4zRbHB&sig=j_o95WrOFtgKKgnipb7fF840b0k

Schwartz, L. H., Litiere, S., de Vries, E., Ford, R., Gwyther, S., Mandrekar, S., Shankar, L., Bogaerts, J., Chen, A., Dancey, J., Hayes, W., Hodi, F. S., Hoekstra, O. S., Huang, E. P., Lin, N., Liu, Y., Therasse, P., Wolchok, J. D., & Seymour, L. (2016). RECIST 1.1-Update and clarification: From the RECIST committee. *Eur J Cancer*, *62*, 132-137. https://doi.org/10.1016/j.ejca.2016.03.081

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006). Intellectual ability and cortical development

in children and adolescents. *Nature*, *440*(7084), 676-679. https://doi.org/10.1038/nature04513

Short, S. J., Elison, J. T., Goldman, B. D., Styner, M., Gu, H., Connelly, M., Maltbie, E., Woolson, S., Lin, W., Gerig, G., Reznick, J. S., & Gilmore, J. H. (2013). Associations between white matter microstructure and infants' working memory. *Neuroimage*, *64*, 156-166. https://doi.org/10.1016/j.neuroimage.2012.09.021

Skagenholt, M., Skagerlund, K., & Traff, U. (2021). Neurodevelopmental differences in child and adult number processing: An fMRI-based validation of the triple code model. *Dev Cogn Neurosci*, *48*, 100933. https://doi.org/10.1016/j.dcn.2021.100933

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*. https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

Sosu, E. M., Dare, S., Goodfellow, C., & Klein, M. (2021). Socioeconomic status and school absenteeism: A systematic review and narrative synthesis. *Review of Education*, *9*(3), e3291. https://doi.org/https://doi.org/10.1002/rev3.3291

Spisak, T., Bingel, U., & Wager, T. D. (2023). Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, *615*(7951), E4-E7. https://doi.org/10.1038/s41586-023-05745-x

Spreen, O., & Strauss, E. A. (2006). *Compendium of Neuropsychological Tests. Administration, norms, and commentary.* (3rd ed.). Oxford University Press.

Sripada, C., Angstadt, M., Taxali, A., Clark, D. A., Greathouse, T., Rutherford, S., Dickens, J. R., Shedden, K., Gard, A. M., Hyde, L. W., Weigard, A., & Heitzeg, M. (2021). Brain-wide functional connectivity patterns support general cognitive ability and mediate effects of socioeconomic status in youth. *Transl Psychiatry*, *11*(1), 571. https://doi.org/10.1038/s41398-021-01704-0

Sripada, C., Rutherford, S., Angstadt, M., Thompson, W. K., Luciana, M., Weigard, A., Hyde, L. H., & Heitzeg, M. (2020). Prediction of neurocognition in youth from resting state fMRI. *Mol Psychiatry*, *25*(12), 3413-3421. https://doi.org/10.1038/s41380-019-0481-6

Stoll, C. R. T., Izadi, S., Fowler, S., Green, P., Suls, J., & Colditz, G. A. (2019). The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods*. https://doi.org/10.1002/jrsm.1369

Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biol Psychiatry*, *88*(11), 818-828. https://doi.org/10.1016/j.biopsych.2020.02.016

Symms, M., Jager, H. R., Schmierer, K., & Yousry, T. A. (2004). A review of structural magnetic resonance neuroimaging. *J Neurol Neurosurg Psychiatry*, *75*(9), 1235-1244. https://doi.org/10.1136/jnnp.2003.032714

Szucs, D., & Ioannidis, J. P. (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*, 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Tan, C. Y. (2024). Socioeconomic Status and Student Learning: Insights from an Umbrella Review. *Educational Psychology Review*, *36*(3), 100. https://doi.org/10.1007/s10648-024-09929-3

Taylor, R. L., Cooper, S. R., Jackson, J. J., & Barch, D. M. (2020). Assessment of Neighborhood Poverty, Cognitive Function, and Prefrontal and Hippocampal Volumes in Children. *JAMA Netw Open*, *3*(11), e2023774. https://doi.org/10.1001/jamanetworkopen.2020.23774

Thomas, D., & Wells, J. (2011). MR angiography and arterial spin labelling. *Methods Mol Biol*, *711*, 327-345. https://doi.org/10.1007/978-1-61737-992-5_16

Thomas, M. S. C., & Coecke, S. (2023). Associations between Socioeconomic Status, Cognition, and Brain Structure: Evaluating Potential Causal Pathways Through Mechanistic Models of Development. *Cogn Sci*, *47*(1), e13217. https://doi.org/10.1111/cogs.13217

Thompson, P. M., & Jahanshad, N. (2015). Novel Neuroimaging Methods to Understand How HIV Affects the Brain. *CURRENT HIV/AIDS REPORTS*, *12*, 289–298. https://doi.org/10.1007/s11904-015-0268-6

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tissier, C., Linzarini, A., Allaire-Duquette, G., Mevel, K., Poirel, N., Dollfus, S., Etard, O., Orliac, F., Peyrin, C., Charron, S., Raznahan, A., Houde, O., Borst, G., & Cachia, A. (2018). Sulcal Polymorphisms of the IFC and ACC Contribute to Inhibitory Control Variability in Children and Adults. *eNeuro*, *5*(1). https://doi.org/10.1523/ENEURO.0197-17.2018

Tohka, J., Moradi, E., Huttunen, H., & Initiative, A. S. D. N. (2016). Comparison of Feature Selection Techniques in Machine Learning for Anatomical Brain MRI in Dementia. *Neuroinformatics*. https://doi.org/10.1007/s12021-015-9292-3

Toich, J. T. F., Taylor, P. A., Holmes, M. J., Gohel, S., Cotton, M. F., Dobbels, E., Laughton, B., Little, F., Van Der Kouwe, A. J. W., Biswal, B., & Meintjes, E. M. (2018). Functional connectivity alterations between networks and associations with infant immune health within networks in HIV infected children on early treatment: A study at 7 years. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2017.00635

Torres van Grinsven, V. (2022). Sources of measurement error in pediatric intelligence testing. *Methodological Innovations*, *15*(1), 96-104. https://doi.org/10.1177/20597991221077919

Tournier, J. D. (2019). Diffusion MRI in the brain – Theory and concepts. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *112-113*, 1-16. https://doi.org/10.1016/j.pnmrs.2019.03.001

Tran, L. T., Roos, A., Fouche, J. P., Koen, N., Woods, R. P., Zar, H. J., Narr, K. L., Stein, D. J., & Donald, K. A. (2016). White matter microstructural integrity and neurobehavioral outcome of hiv-exposed uninfected neonates. *Medicine (United States)*. https://doi.org/10.1097/MD.0000000000002577

Tremblay, N., & Loukas, A. (2020). Approximating spectral clustering via sampling: a review. In *Sampling Techniques for Supervised or Unsupervised Tasks* (pp. 129-183). Springer, Cham. https://doi.org/10.1007/978-3-030-29349-9_5

Tuke, R., Sikorskii, A., Zalwango, S. K., Webster, K. D., Ismail, A., Pobee, R. A., Barkin, J. L., Boivin, M. J., Giordani, B., & Ezeamama, A. E. (2020). Psychosocial Adjustment in Ugandan Children: Coping With Human Immunodeficiency Virus Exposure, Lifetime Adversity, and Importance of Social Support. *New Directions for Child and Adolescent Development*. https://doi.org/10.1002/cad.20354

Uban, K. A., Herting, M. M., Williams, P. L., Ajmera, T., Gautam, P., Huo, Y., Malee, K. M., Yogev, R., Csernansky, J. G., Wang, L., Nichols, S. L., Sowell, E. R., Pediatric, H. C., the Pediatric Imaging, N., & Genetics, S. (2015). White matter microstructure among youth with perinatally acquired HIV is associated with disease severity. *AIDS*, *29*(9), 1035-1044. https://doi.org/10.1097/QAD.0000000000000648

Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *J Neurosci*, *34*(5), 1592-1598. https://doi.org/10.1523/JNEUROSCI.0842-13.2014

Valcour, V., Paul, R., Neuhaus, J., & Shikuma, C. (2011). The Effects of Age and HIV on Neuropsychological Performance. *J Int Neuropsychol Soc*, *17*(1), 190-195. https://doi.org/10.1017/S1355617710001438

van Dalen, Y. W., Blokhuis, C., Cohen, S., Ter Stege, J. A., Teunissen, C. E., Kuhle, J., Kootstra, N. A., Scherpbier, H. J., Kuijpers, T. W., Reiss, P., Majoie, C., Caan, M. W. A., & Pajkrt, D. (2016). Neurometabolite Alterations Associated With Cognitive Performance in Perinatally HIV-Infected Children. *Medicine (Baltimore)*, *95*(12), e3093. https://doi.org/10.1097/MD.0000000000003093

van den Hof, M., Jellema, P. E. J., Ter Haar, A. M., Scherpbier, H. J., Schrantee, A., Kaiser, A., Caan, M. W. A., Majoie, C., Reiss, P., Wit, F., Mutsaerts, H. M. M., & Pajkrt, D. (2021). Normal structural brain development in adolescents

treated for perinatally acquired HIV: a longitudinal imaging study. *AIDS*, *35*(8), 1221-1228. https://doi.org/10.1097/QAD.0000000000002873

van den Hof, M., Ter Haar, A. M., Caan, M. W. A., Spijker, R., van der Lee, J. H., & Pajkrt, D. (2019). Brain structure of perinatally HIV-infected patients on long-term treatment: A systematic review. *Neurol Clin Pract*, *9*(5), 433-442. https://doi.org/10.1212/CPJ.0000000000000637

van der Kouwe, A. J. W., Benner, T., Salat, D. H., & Fischl, B. (2008). Brain morphometry with multiecho MPRAGE. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.12.025

van Genderen, J. G., Chia, C., Van den Hof, M., Mutsaerts, H., Reneman, L., Pajkrt, D., & Schrantee, A. (2022). Brain Differences in Adolescents Living With Perinatally Acquired HIV Compared With Adoption Status Matched Controls: A Cross-sectional Study. *Neurology*, *99*(15), e1676-e1684. https://doi.org/10.1212/WNL.0000000000200946

van Genderen, J. G., Van den Hof, M., Ter Haar, A. M., Blokhuis, C., Keil, V. C., Pajkrt, D., & Mutsaerts, H. (2021). A Longitudinal Analysis of Cerebral Blood Flow in Perinatally HIV Infected Adolescents as Compared to Matched Healthy Controls. *Viruses*, *13*(11). https://doi.org/10.3390/v13112179

van Wyhe, K. S., Laughton, B., Cotton, M. F., Meintjes, E. M., van der Kouwe, A. J. W., Boivin, M. J., Kidd, M., & Thomas, K. G. F. (2021). Cognitive outcomes at ages seven and nine years in South African children from the children with HIV early antiretroviral (CHER) trial: a longitudinal investigation. *Journal of the International AIDS Society*. https://doi.org/10.1002/jia2.25734

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*, *80, Part A*, 68-77. https://doi.org/10.1016/j.neuroimage.2017.06.061

Vastag, Z., Fira-Mladinescu, O., & Rosca, E. C. (2022). HIV-Associated Neurocognitive Disorder (HAND): Obstacles to Early Neuropsychological Diagnosis. *Int J Gen Med*, *15*, 4079-4090. https://doi.org/10.2147/IJGM.S295859

Vieira, B. H., Pamplona, G. S. P., Fachinello, K., Silva, A. K., Foss, M. P., & Salmon, C. E. G. (2022). On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. *Intelligence*, *93*, 101654. https://doi.org/https://doi.org/10.1016/j.intell.2022.101654

Violari, A., Cotton, M. F., Gibb, D. M., Babiker, A. G., Steyn, J., Madhi, S. A., Jean-Philippe, P., & McIntyre, J. A. (2008). Early antiretroviral therapy and mortality among HIV-infected infants. *The New England journal of medicine*. https://doi.org/10.1056/NEJMoa0800971

Vishnuvarthanan, G., Rajasekaran, M. P., Subbaraj, P., & Vishnuvarthanan, A. (2016). An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images.

Applied Soft Computing, *38*, 190-212. https://doi.org/https://doi.org/10.1016/j.asoc.2015.09.016

Vu, M. T., Adali, T., Ba, D., Buzsaki, G., Carlson, D., Heller, K., Liston, C., Rudin, C., Sohal, V. S., Widge, A. S., Mayberg, H. S., Sapiro, G., & Dzirasa, K. (2018). A Shared Vision for Machine Learning in Neuroscience. *J Neurosci*, *38*(7), 1601-1607. https://doi.org/10.1523/JNEUROSCI.0508-17.2018

Wabiri, N., & Taffa, N. (2013). Socio-economic inequality and HIV in South Africa. *BMC Public Health*. https://doi.org/10.1186/1471-2458-13-1037

Wade, B. S. C., Valcour, V. G., Puthanakit, T., Saremi, A., Gutman, B. A., Nir, T. M., Watson, C., Aurpibul, L., Kosalaraksa, P., Ounchanum, P., Kerr, S., Dumrongpisutikul, N., Visrutaratna, P., Srinakarin, J., Pothisri, M., Narr, K. L., Thompson, P. M., Ananworanich, J., Paul, R. H., & Jahanshad, N. (2019). Mapping abnormal subcortical neurodevelopment in a cohort of Thai children with HIV. *NeuroImage: Clinical*. https://doi.org/10.1016/j.nicl.2019.101810

Waffenschmidt, S., Knelangen, M., Sieben, W., Bühn, S., & Pieper, D. (2019). Single screening versus conventional double screening for study selection in systematic reviews: A methodological systematic review. *BMC Medical Research Methodology*. https://doi.org/10.1186/s12874-019-0782-0

Wainer, J., & Cawley, G. (2021). Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, *182*, 115222. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115222

Wang, P., Li, J., Wang, X., Thapa, D., & Wu, G. Y. (2018). Asymptomatic Human Immunodeficiency Virus Vertical Transmitted Adolescents' Brain Functional Changes: Based on Resting-State Functional Magnetic Resonance Imaging. *AIDS Res Hum Retroviruses*, *34*(8), 699-704. https://doi.org/10.1089/AID.2017.0267

Wedderburn, C. J., Subramoney, S., Yeung, S., Fouche, J. P., Joshi, S. H., Narr, K. L., Rehman, A. M., Roos, A., Ipser, J., Robertson, F. C., Groenewold, N. A., Gibb, D. M., Zar, H. J., Stein, D. J., & Donald, K. A. (2020). Neuroimaging young children and associations with neurocognitive development in a South African birth cohort study. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2020.116846

Wu, Y., Besson, P., Azcona, E. A., Bandt, S. K., Parrish, T. B., Breiter, H. C., & Katsaggelos, A. K. (2022). A multicohort geometric deep learning study of age dependent cortical and subcortical morphologic interactions for fluid intelligence prediction. *Sci Rep*, *12*(1), 17760. https://doi.org/10.1038/s41598-022-22313-x

Yadav, S. K., Gupta, R. K., Garg, R. K., Venkatesh, V., Gupta, P. K., Singh, A. K., Hashem, S., Al-Sulaiti, A., Kaura, D., Wang, E., Marincola, F. M., & Haris, M. (2017). Altered structural brain changes and neurocognitive performance in pediatric HIV. *Neuroimage Clin*, *14*, 316-322. https://doi.org/10.1016/j.nicl.2017.01.032

Yadav, S. K., Gupta, R. K., Hashem, S., Bhat, A. A., Garg, R. K., Venkatesh, V., Gupta, P. K., Singh, A. K., Chaturvedi, S., Ahmed, S. N., Azeem, M. W., & Haris, M. (2018). Changes in resting-state functional brain activity are associated with waning cognitive functions in HIV-infected children. *Neuroimage Clin*, *20*, 1204-1210. https://doi.org/10.1016/j.nicl.2018.10.028

Yadav, S. K., Gupta, R. K., Hashem, S., Nisar, S., Azeem, T., Bhat, A. A., Syed, N., Garg, R. K., Venkatesh, V., Kamal, M., Fakhro, K., Frenneaux, M. P., & Haris, M. (2020). Brain microstructural changes support cognitive deficits in HIV uninfected children born to HIV infected mothers. *Brain Behav Immun Health*, *2*, 100039. https://doi.org/10.1016/j.bbih.2020.100039

Yao, J. K., Voorhies, W. I., Miller, J. A., Bunge, S. A., & Weiner, K. S. (2023). Sulcal depth in prefrontal cortex: a novel predictor of working memory performance. *Cereb Cortex*, *33*(5), 1799-1813. https://doi.org/10.1093/cercor/bhac173

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Yu, X., Gao, L., Wang, H., Yin, Z., Fang, J., Chen, J., Li, Q., Xu, H., & Gui, X. (2019). Neuroanatomical Changes Underlying Vertical HIV Infection in Adolescents. *Front Immunol*, *10*, 814. https://doi.org/10.3389/fimmu.2019.00814

Yuan, D., Hahn, S., Allgaier, N., Owens, M. M., Chaarani, B., Potter, A., & Garavan, H. (2022). Machine learning approaches linking brain function to behavior in the ABCD STOP task. *Hum Brain Mapp*. https://doi.org/10.1002/hbm.26172

Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC press. https://books.google.com/books?hl=en&lr=&id=BDB50Ev2ur4C&oi=fnd&pg=PP1&dq=Ensemble+methods:+foundations+and+algorithms&ots=OzBFDilNQK&sig=5QaKpJyoFCDvIUdbGc8jh_GAK3Y

Ziegler, G., Dahnke, R., Winkler, A. D., & Gaser, C. (2013). Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents. *Neuroimage*, *82*, 284-294. https://doi.org/10.1016/j.neuroimage.2013.05.088

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. https://doi.org/10.1111/j.1467-9868.2005.00503.x

APPENDICES

Appendix A

Glossary of definitions, principles, and abbreviations

Hierarchical Clustering Algorithm (HCA)

Linkages (mentioned in §3.2.6, §4.2.3)

A linkage is the distance between two clusters.

- Single linkage (nearest neighbour) uses the smallest distance between objects in the two clusters.
- Complete linkage (farthest neighbour) uses the largest distance between all pairs of objects in any two clusters.
- Average linkage uses the average distance between all pairs of objects in any two clusters.
- Centroid linkage uses the Euclidean distance between the centroids of the two clusters.
- Median linkage uses the Euclidean distance between weighted centroids of the two clusters.
- Ward's linkage uses the incremental sum of squares, that is, the increase in the total within-cluster sum of squares because of joining two clusters. The within-cluster sum of squares is defined as the sum of squares of the distance between all objects in the cluster and the centroid of the cluster.

Cophenetic distance (between two observations) is represented in a dendrogram by the height of the link at which two observations are first joined.

Cophenetic correlation coefficient (c) is the linear correlation coefficient between the cophenetic distances obtained from the tree and the original distances used to

construct the tree. Coefficient closer to 1 indicates higher validity of the output from hierarchical clustering algorithm.

----------**mentioned in** §3.2.6, §4.2.3----------

Machine learning algorithms

Penalties (§3.2.7 penalised regression)

Regularisation or penalisation is a way to avoid overfitting by reducing/shrinking high-valued regression coefficients. It achieves this by biasing data towards particular values via adding tuning parameters to encourage those values: (mentioned §3.1)

- L^1-norm regularisation adds a penalty equal to the absolute value of the magnitude of coefficients. Some coefficients can become zero and eliminated in the model. This method is used by lasso regression. (also mentioned §3.3.3)
- L^2-norm regularisation adds a penalty equal to the square of the magnitude of the coefficients, and all the parameters are reduced by the same factor, although some approach zero and can be eliminated. Ridge regression and SVMs use this method.

Kernels (§3.2.7 support vector machines)

Kernel SVM is implemented when the data is not linearly separable. It is achieved by using kernels discussed below (mentioned §3.1 kernel SVM).
Kernels are mathematical functions used by the kernel SVM algorithm to transform the data from a lower dimension where the points are often not linearly separable to a high dimensional plane where they can be separated easily. Kernel functions include:

- Polynomial kernel takes the dot product of the data points in the original space and the polynomial function in the new space.

- Gaussian radial basis function (RBF) kernel finds the dot products and squares of all the features in the dataset. The RBF kernel is formed by taking an infinite sum over polynomial kernels.
- Sigmoid kernel function is equivalent to the 2-layer perceptron model of the neural network, used as an activation function for artificial neurons.

----------**Mentioned in** §3.2.7 support vector machines----------

Ensemble methods (discussed in §3.2.7 decision tree ensembles)

In a decision tree, the dataset is split based on different conditions like mutual information. Mutual information is a measure between random variables that quantifies the amount of information obtained about one random variable through the other random variable. Multiple decision trees are combined via boosting or bagging (mentioned §3.1).

Bagging (bootstrap aggregation)

Bagging aggregates the prediction of several weak learners to select the best prediction.

- Random forest generates many bootstrap replicas of the data and grow decision trees on the replicas. For dataset size N, each bootstrap replica is obtained by randomly selecting N out of N observations with replacement. Every tree in the ensemble can randomly select predictors for each decision split. (mentioned §3.1, §3.2.7).

Boosting

Boosting trains the trees sequentially and corrects errors of the previous weak learner. (mentioned §3.1boosted trees)

- Adaptive boosting (AdaBoost) trains learners sequentially. For learner q, AdaBoost increases weights for misclassified observations and reduces weights for correctly classified observations. The next learner q+1 is then trained on the data with updated weights.

- Random under-sampling boosting (RUSBoost) is designed specifically for class imbalances and can be used with AdaBoost. RUSBoost takes N, basic unit for sampling, as the number of members in the class with the fewest training members. Categories with more members are under sampled by taking only N observations. (mentioned §3.3.3)
- Least-Squares boosting (LSBoost) fits regression ensembles to minimise mean-squared error. At every step, LSBoost fits a new learner to the difference between the observed response and the collective prediction of previous learners.
- Totally corrective boosting (TotalBoost) designed for multiclass classification. TotalBoost maximises the minimal margin in the training set, the margin between the difference between the predicted soft classification score for the true class and the largest score for the false categories.

----------**discussed in** §3.2.7 decision tree ensembles----------

Multimodal magnetic resonance imaging (§3.2.4, §4.2.4)

Structural magnetic resonance imaging (sMRI) (§3.2.4 & §4.2.4 sMRI features)

sMRI quantifies morphometric characteristics of brain tissue such as cortical and sub-cortical volumes, thickness as well as local gyrification, curvature, and area.

- Thickness is the depth of the outer layer of neural tissues in the cerebrum and is closely associated with motor control. In FreeSurfer, cortical thickness corresponds to the distance between the white surface, i.e., gray-white matter boundary and the pial surface, i.e., gray matter-cerebrospinal fluid boundary.
- Local gyrification index (LGI) quantifies the amount of cortex buried within the sulcal folds compared to the amount on the outer visible cortex.
- Curvature in general is measured as 1/r by FreeSurfer, where r is the radius of an inscribed circle. Mean curvature is the average of the two principal curvatures with units (1/mm).

- Area corresponds to the white surface, i.e., gray-white matter boundary.
- Volume corresponds to the product of the surface area and cortical thickness.

----------**Mentioned in** §3.2.4 & §4.2.4 sMRI features----------

Diffusion tensor imaging (DTI) (§3.2.4 & §4.2.4 DTI features)

DTI measures the directional dependence of water molecules in tissue and can provide information about white matter microstructural characteristics.

- Fractional anisotropy (FA) is the extent to which water molecules travel in organised or random direction. It is a scalar value between 0 and 1. Higher FA typically shows increased axonal count, neural density, and myelination.
- Mean diffusivity (MD) is the overall diffusion, calculated as the mean of the three eigenvalues of the diffusion tensor. MD is a measure of organisation: Low MD values show a well-organised structure and dense axonal packing.
- Radial diffusivity (RD) is the diffusion in the direction perpendicular to the axonal fibres. Higher RD decreases shows decreased myelination and axonal count.
- Axial diffusivity (AD) is the diffusion in the direction parallel to the fibre tracts. Higher AD shows well-organised structure and fibre coherence.

----------**Mentioned in** §3.2.4 & §4.2.4 DTI features----------

Proton magnetic resonance spectroscopy (^{1}H-MRS) (§3.2.4 & §4.2.4 MRS features)

Spectroscopy allows for the assessment of regional brain biochemistry. ^{1}H-MRS can quantify the concentration of various metabolites:

- Creatine (Cr) is a natural source of energy and often used to calculate ratios of other metabolites because of its stability. Phosphocreatine (PCr) is phosphorylated form of creatine to mobilise phosphates to recycle adenosine triphosphate.

- Choline (Cho) is needed to form membranes that surround tissue. Higher Cho typically shows inflammation or infarction. Phosphocholine (PCh), like PCr, is a phosphorylated form of Choline
- Myo-inositol (Ins) indicates gliosis (proliferation of glial cells), inflammation, and myelin degradation.
- N-acetylaspartate (NAA) indicates axonal and neuronal density, so loss of neurons is signalled by a decreased concentration.
- Glutamate/Glutamine (Glu/Gln) is an excitatory neurotransmitter and is important to memory, cognition, and mood regulation.
- N-acetylaspartylglutamate (NAAG) is a precursor to NAA and it modulates glutamate release.

---------- **Mentioned in** §3.2.4 & §4.2.4 MRS features----------

Appendix B: Supplementary Tables for Chapters 3 & 4

Demographics data of the study (Chapter 3 & Chapter 4)

We had 132 children part of our study. The follow is a table of demographics variables (sample size, health status, sex, birth weight, ages at scan) for the children at both age 7 and 9. How these variables differ between clusters of "poorer" and "better" performing children are given in the main text, Tables 3.1 & 4.2.

Table B.1| Demographics characteristics of the sample used in the study

Participants	132 children
Health status	63 controls, 69 HV
Sex	62 Female, 70 Male
Birth weight	Mean = 3150g, standard error = 55g
Age at 7 years scan	Mean = 7.22 years, standard error = 0.013 years
Age at 9 years scan	Mean = 9.41 years, standard error = 0.05 years

Regression and classification performance of HIV and controls separately

In Chapters 3 and 4, the children living with and without HIV were treated as a single homogenous group representing the low-socioeconomic status background cohort. Below are the results for the analysis performed in the groups separately.

Table B.2 show the regression performance resulting from the individual prediction of the 11 cognitive scores at age 7 of children living with HIV using PLMs, SVMs, and DTEs and multimodal MRI at age 7, and Table B.3 is likewise the performance for the children living without HIV (controls). Since the sample sizes are very small, we have implemented the models with 5-fold cross validation.

Table B.2| PLMs, SVMs, and DTEs regression performance for children living with HIV using the multimodal MRI dataset at age 7.

Cognitive test	Model category	RMSE	MAE	r	R^2	Algorithm with lowest CV errors**
Learning	PLM	5.24	4.14	0.04	0.00	
	SVM	4.16	3.02	0.13	0.05	Quadratic SVM
	DTE	4.67	4.22	0.09	0.03	
Planning	PLM	7.08	6.77	0.04	-0.78	
	SVM	4.68	4.28	0.06	0.01	Linear SVM
	DTE	4.83	4.44	0.08	0.00	
Simultaneous processing	PLM	4.77	3.95	-0.38	-0.11	
	SVM	4.81	3.99	-0.03	-0.09	Boosted DTE
	DTE	4.79	3.94	0.07	-0.08	
Sequential processing	PLM	5.96	5.07	-0.23	-1.01	
	SVM	4.35	3.80	-0.10	0.02	Lasso PLM
	DTE	4.56	3.96	-0.10	-0.09	
Beery-VMI	PLM	2.37	2.04	-0.67	-0.16	
	SVM	2.31	1.99	0.14	-0.05	Linear SVM
	DTE	2.34	2.01	0.07	-0.09	
PPVT	PLM	16.29	11.96	-0.11	-1.33	
	SVM	17.07	12.36	0.18	-0.09	Linear SVM
	DTE	17.11	12.28	0.15	-0.09	
SFT	PLM	3.40	2.94	-0.06	-0.39	
	SVM	3.11	2.66	-0.03	-0.09	Boosted DTE
	DTE	2.95	2.46	0.17	0.05	
TOVA API	PLM	3.19	2.78	-0.37	-0.10	
	SVM	3.16	2.70	-0.01	-0.06	Bagged DTE
	DTE	3.11	2.70	-0.02	-0.02	
PPT Preferred	PLM	2.56	2.24	-0.41	-0.11	
	SVM	2.37	2.07	0.06	0.11	Bagged DTE
	DTE	2.49	2.16	0.08	0.00	
PPT non-preferred	PLM	2.43	2.08	-0.54	-0.12	
	SVM	2.43	2.09	-0.08	-0.09	Bagged DTE
	DTE	2.43	2.09	0.05	-0.10	
PPT Both hands	PLM	2.48	2.17	-0.35	-0.58	
	SVM	2.18	1.88	0.16	-0.06	Linear SVM
	DTE	2.20	1.90	0.13	-0.09	

Table B.3| PLMs, SVMs, and DTEs regression performance for children living without HIV.

Cognitive test	Model category	RMSE	MAE	r	R^2	Algorithm with lowest CV errors**
Learning	PLM	5.32	4.22	0.03	0.01	
	SVM	4.24	3.10	0.15	0.06	Linear SVM
	DTE	4.75	4.30	0.08	0.04	
Planning	PLM	7.16	6.85	0.03	-0.77	
	SVM	4.76	4.36	0.05	0.02	Cubic SVM
	DTE	4.91	4.52	0.07	0.01	
Simultaneous processing	PLM	4.85	4.03	-0.39	-0.10	
	SVM	4.89	4.07	-0.04	-0.08	Boosted DTE
	DTE	4.87	4.02	0.06	-0.07	
Sequential processing	PLM	6.04	5.15	-0.24	-1.00	
	SVM	4.43	3.88	-0.11	0.03	Elastic net PLM
	DTE	4.64	4.04	-0.11	-0.08	
Beery-VMI	PLM	2.45	2.12	-0.68	-0.15	
	SVM	2.39	2.07	0.14	-0.04	Linear SVM
	DTE	2.42	2.09	0.06	-0.08	
PPVT	PLM	16.37	12.04	-0.12	-1.32	
	SVM	17.15	12.44	0.17	-0.08	Lasso PLM
	DTE	17.19	12.36	0.14	-0.08	
SFT	PLM	3.48	3.02	-0.07	-0.38	
	SVM	3.19	2.74	-0.04	-0.08	Boosted DTE
	DTE	3.03	2.54	0.16	0.06	
TOVA API	PLM	3.27	2.86	-0.38	-0.09	
	SVM	3.24	2.78	-0.02	-0.05	Boosted DTE
	DTE	3.19	2.78	-0.03	-0.01	
PPT Preferred	PLM	2.64	2.32	-0.42	-0.10	
	SVM	2.45	2.15	0.05	0.12	Linear SVM
	DTE	2.57	2.24	0.07	0.01	
PPT non-preferred	PLM	2.51	2.16	-0.55	-0.11	
	SVM	2.51	2.17	-0.09	-0.08	Lasso PLM
	DTE	2.51	2.17	0.04	-0.09	
PPT Both hands	PLM	2.56	2.25	-0.36	-0.57	
	SVM	2.26	1.96	0.15	-0.05	Linear SVM
	DTE	2.28	1.98	0.12	-0.08	

**This is based on comparing regression errors between models. All the models produce poor regression performance as shown by the r and R-squared values.

These are the classification performance results for poorer and better performing children in each group of children living with (Table B.4) and without HIV (Table B.5). The categories were performed using hierarchical clustering as detailed in Chapter 3.

Table B.4| Classification performance of SVM, DTE, and PLM for distinguishing poorer (n = 24) vs better (n = 24) performing children living with HIV

Technique	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	Balanced accuracy (%)*	Relevant Features
SVM	0.61	68	52	54	60	60
DTE	0.55	52	53	51	53	79
PLM**	—	—	—	—	—	—

*Balanced accuracy was computed due to imbalances in category memberships. It is the arithmetic mean of sensitivity and specificity.
**Models failed to converge to a solution

Table B.5| Classification performance of SVM, DTE, and PLM for distinguishing poorer (n = 16) vs better (n = 25) performing children living without HIV

Technique	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	Balanced accuracy (%)*	Relevant Features
SVM	0.53	61	52	52	57	77
DTE	0.57	51	52	49	52	64
PLM**	—	—	—	—	—	—

*Balanced accuracy was computed due to imbalances in category memberships. It is the arithmetic mean of sensitivity and specificity.
**Models failed to converge to a solution

Feature selection: prediction of categories of performance at ages 7 and 9 years

In Chapters 3 and 4, we have presented feature selection for only the best performing models. Below are the selected features that were not presented for comparisons but referred to in the text.

In the prediction of cognitive performance at age 7, Table B.6 shows the selected features for RUS boosted decision tree ensemble (DTE). Table B.7 shows the selected features for lasso penalised linear model (PLM). In the prediction of cognitive performance at age 9, Table B.8 shows selected features using linear SVM on the MM7 dataset. Table B.9 shows selected features using linear SVM on the MM9 dataset. Table B.9 shows selected features using the CS7+MM79 dataset.

Table B.6| 48 features selected in the prediction of cognitive performance at age 7 using RUSBoost trees

Feature	Weight* $\times 10^{-4}$
Left pallidum volume	296.05
Right pallidum volume	166.35
LH precentral volume	54.85
Left lateral ventricle volume	39.54
Right superior parietal area	38.89
RD in right inferior longitudinal fasciculus	36.91
Right cerebellum white matter volume	24.34
RH middle temporal mean curvature	22.89
LH pars triangularis mean curvature	22.27
RH cerebral white matter volume	19.11
Right cuneus thickness	21.56
RD in left inferior longitudinal fasciculus	17.59
Right superior frontal volume	17.29
LH cerebral white matter volume	17.12
Left superior parietal area	16.80
Left transverse-temporal volume	16.29
Right paracentral volume	16.03
Left accumbens area volume	15.56
RH mean thickness	15.09
Right thalamus proper volume	14.80

Left frontal pole area	14.60
Right postcentral volume	14.33
Left cuneus thickness	13.81
Left supramarginal volume	13.66
Left thalamus proper volume	12.83
Left supramarginal thickness	12.11
Right fusiform volume	12.03
Right transverse-temporal volume	11.79
CC anterior volume	11.11
AD in left uncinate fasciculus	10.99
Right para-hippocampal volume	10.93
CC posterior volume	9.80
Right caudate volume	9.32
Right isthmus cingulate volume	8.51
Left superior parietal volume	7.79
Left lingual volume	7.57
Left inferior lateral ventricle volume	7.57
Left frontal pole volume	7.28
Left hippocampus volume	5.46
Left caudate volume	5.43
Right temporal pole LGI	5.43
CSF volume	5.41
Left isthmus cingulate volume	5.41
Right lingual volume	5.01
Left sulcal banks volume	5.00
CC mid-posterior volume	5.00
CC central volume	4.98
Right amygdala volume	4.92

*Weights or coefficients in the final SVM model after recursive feature elimination

Abbreviations

LH - left hemisphere,

RH - right hemisphere

CSF - cerebrospinal fluid

CC - corpus callosum

LGI - local gyrification index

Table B.7| 45 features selected in the prediction of cognitive performance at age 7 using lasso PLM.

Feature	Weight* $\times 10^{-4}$
Right superior parietal area	301.53
Right pallidum volume	171.83
LH precentral volume	60.33
Left lateral ventricle volume	45.02
Right cerebellum white matter volume	44.37
Left pallidum volume	42.39
RH middle temporal mean curvature	29.82
LH pars triangularis mean curvature	28.37
RH cerebral white matter volume	27.75
RD in left inferior longitudinal fasciculus	24.59
Right superior frontal volume	27.04
LH cerebral white matter volume	23.07
Left superior parietal area	22.77
Right paracentral volume	22.60
Left accumbens area volume	22.28
RH mean thickness	21.77
Left frontal pole area	21.51
Right postcentral volume	21.04
Left cuneus thickness	20.57
Left supramarginal volume	20.28
Left thalamus proper volume	20.08
Left supramarginal thickness	19.81
Right fusiform volume	19.29
Right transverse-temporal volume	19.14
CC anterior volume	18.31
AD in left uncinate fasciculus	17.59
Right para-hippocampal volume	17.51
CC posterior volume	17.27
Right isthmus cingulate volume	16.59
Left superior parietal volume	16.47
Left lingual volume	16.41
Left inferior lateral ventricle volume	15.28
Left frontal pole volume	14.80
Left hippocampus volume	13.99
Left caudate volume	13.27
Right temporal pole LGI	13.05

CSF volume	13.05
Left pallidum volume	12.76
Left sulcal banks volume	10.94
CC mid-posterior volume	10.91
CC central volume	10.88
Right amygdala volume	10.88
AD in right uncinate fasciculus	10.83
Right cuneus thickness	10.83
Right cuneus curvature	10.81

*Weights or coefficients in the final SVM model after recursive feature elimination

Abbreviations

LH - left hemisphere,
RH - right hemisphere
CSF - cerebrospinal fluid
CC - corpus callosum
LGI - local gyrification index

Table B.8| 26 features selected when predicting cognitive performance categories at age 9 using the MM7 dataset

Feature	Weight* $\times 10^{-4}$
PCr concentration	16.76
RH caudal anterior cingulate area	16.40
LH pars opercularis area	15.22
Glu concentration	9.63
RH superior frontal LGI	8.42
LH medial orbitofrontal LGI	8.36
AD in left cingulum	7.75
LH pars triangularis LGI	7.01
LH sulcal banks volume	6.66
RH superior parietal LGI	6.31
RH para-hippocampal thickness	6.01
Cr concentration	5.80
LH posterior cingulate mean curvature	5.52
RH caudal middle frontal LGI	5.43
RD in left inferior longitudinal fasciculus	5.01
Cho concentration	4.80
Left thalamus proper volume	4.73
RH caudal middle frontal area	4.67

NAA concentration	4.72
LH inferior temporal thickness	4.53
LH supramarginal thickness	4.16
MD in right uncinate fasciculus	3.98
RH supramarginal volume	3.86
MD in right superior longitudinal fasciculus	3.66
Right pallidum volume	3.56
Left pallidum volume	3.45

* Weights or coefficients in the final SVM model after recursive feature elimination

Abbreviations

LH/RH - left/right hemispheres, PCr - phosphocreatine, Glu - glutamate, AD - axial diffusivity, LGI - local gyrification index, Cho - choline, Cr - creatine, MD - mean diffusivity

Table B.9| 24 features selected when predicting cognitive performance categories at age 9 using the MM9 dataset

Feature	Weight* $\times 10^{-4}$
NAA concentration	25.82
PCr concentration	25.46
RH caudal anterior cingulate area	24.28
RH superior parietal LGI	18.69
LH pars opercularis area	17.48
Glu concentration	17.42
RH superior frontal LGI	16.81
LH medial orbitofrontal LGI	16.07
AD in left cingulum	15.72
AD left corticospinal tract	15.37
LH pars triangularis LGI	15.07
LH sulcal banks volume	14.86
RH para-hippocampal thickness	14.58
RH supramarginal volume	14.49
Cr concentration	14.07
LH posterior cingulate mean thickness	13.86
LH caudal middle frontal curvature	13.79
MD in right inferior longitudinal fasciculus	13.73
Cho concentration	13.78
RH supramarginal thickness	13.59

RH caudal middle frontal curvature	13.22
Right pallidum volume	13.04
Left pallidum volume	12.92
LH inferior temporal thickness	12.72

* Weights or coefficients in the final SVM model after recursive feature elimination
Abbreviations
LH/RH - left/right hemispheres, PCr - phosphocreatine, Glu - glutamate, AD - axial diffusivity, LGI - local gyrification index, Cho - choline, Cr - creatine, MD - mean diffusivity

Table B.10| 43 features selected when predicting cognitive performance categories at age 9 using the CS7+MM79 dataset

Feature	Weight* $\times 10^{-4}$
Left pallidum volume age 7	304.70
Right pallidum volume age 9	175.00
LH precentral volume age 9	63.50
Left lateral ventricle volume age 9	48.19
Right superior parietal area age 7	47.54
RD in right inferior longitudinal fasciculus age 9	45.56
RH middle temporal mean curvature age 9	31.54
LH pars triangularis mean curvature age 9	30.92
RH cerebral white matter volume age 7	27.76
Right cuneus thickness age 7	30.21
RD in left inferior longitudinal fasciculus age 9	26.24
Right superior frontal volume age 9	25.94
LH cerebral white matter volume age 9	25.77
Left superior parietal area age 9	25.45
Left transverse-temporal volume age 9	24.94
Right paracentral volume age 9	24.68
Left accumbens area volume age 9	24.21
RH mean thickness age 7	23.74
Right thalamus proper volume age 9	23.45
Left frontal pole area age 7	23.25
Right postcentral volume age 9	22.98
Left cuneus thickness age 9	22.46
Left thalamus proper volume age 9	21.48

Left supramarginal thickness age 9	20.76
Right fusiform volume age 7	20.68
Right transverse-temporal volume age 9	20.44
CC anterior volume age 9	19.76
AD in left uncinate fasciculus age 9	19.64
Right para-hippocampal volume age 9	19.58
CC posterior volume age 9	18.45
Right caudate volume age 7	17.97
Left superior parietal volume age 9	16.44
Left lingual volume age 7	16.22
Left inferior lateral ventricle volume age 9	16.22
Left frontal pole volume age 9	15.93
Left hippocampus volume age 9	14.11
Left caudate volume age 9	14.08
CSF volume age 9	14.06
Left isthmus cingulate volume	14.06
Right lingual volume age 9	13.66
CC mid-posterior volume age 9	13.65
CC central volume age 9	13.63
Right amygdala volume age 7	13.57

*Weights or coefficients in the final SVM model after recursive feature elimination

Abbreviations

LH - left hemisphere,
RH - right hemisphere
CSF - cerebrospinal fluid
CC - corpus callosum
LGI - local gyrification index

Appendix C

Ethical approval

The Human Research Ethics Committee of the University of Cape Town approved this study. The letter from the committee is provided below.

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room G50- Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: hrec-submissions@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

23 July 2021

HREC REF: 453/2021

Dr F Robertson
Division of Biomedical Engineering
FHS
Email: Frances.robertson@uct.ac.za
Student: Khbisa001@myuct.ac.za

Dear Dr Robertson

PROJECT TITLE: T EXPLORING THE RELATIONSHIP BETWEEN MAGNETIC RESONANCE NEUROIMAGING AND NEUROCOGNITIVE OUTCOMES OF CHILDREN LIVING WITH HIV-PHD CANDIDATE-MR ISAAC L KHOBO-SUB-STUDY LINKED 448/2011

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

This approval is subject to strict adherence to the HREC recommendations regarding research involving human participants during COVID -19, dated 17 March 2020 & 06 July 2020.

Approval is granted for one year until the 30 July 2022.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

The HREC acknowledge that the student: Mr Isaac Khobo will also be involved in this study.

Please quote the HREC REF 453/2021 in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FACULTY OF HEALTH SCIENCES HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
NHREC-registration number: REC-210208-007
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2020), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI), and Declaration of Helsinki (2013) guidelines. The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.