Functional, structural, and phenotypic data fusion to predict developmental scores of pre-school children based on Canonical Polyadic Decomposition

Citation for published version:
Dron, N, Navarro, M, Chin, R & Escudero, J 2021, 'Functional, structural, and phenotypic data fusion to predict developmental scores of pre-school children based on Canonical Polyadic Decomposition', *Biomedical Signal Processing and Control*, vol. 70, 102889. https://doi.org/10.1016/j.bspc.2021.102889

Digital Object Identifier (DOI):
10.1016/j.bspc.2021.102889

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Biomedical Signal Processing and Control

Functional, structural, and phenotypic data fusion to predict developmental scores of pre-school children based on Canonical Polyadic Decomposition

Noramon Dron[a,b,*], Maria Navarro[c], Richard F.M. Chin[b,d,e], Javier Escudero[a,b]

[a] *School of Engineering, Institute for Digital Communications, University of Edinburgh, UK*
[b] *Muir Maxwell Epilepsy Centre, Edinburgh, UK*
[c] *Computer Sciences Department in University of Salamanca, Spain*
[d] *Royal Hospital for Sick Children, Edinburgh*
[e] *Centre for Clinical Brain Sciences at the University of Edinburgh, Edinburgh, UK*

Abstract

Recent technological advances enable the acquisition of diverse datasets that demand data-driven analysis. In this context, we seek to take advantage of diverse data modalities to explore the links between childhood development, structure and function of the brain. We deploy a data fusion model using coupled matrix-tensor decomposition of electroencephalography (EEG), structural magnetic resonance imaging (sMRI), and phenotypic score data to investigate how functional, structural, and phenotypic variables reflect development in young children with epilepsy. Our model is based on Canonical Polyadic Decomposition and optimised with grid search to predict developmental scores of preschool children. The model is promising and able to show relationships between modalities that agree with clinical expectations. The score prediction yields a high similarity at the group level and potential to predict laborious and time-consuming developmental scores from routinely collected sMRI and/or EEG data, thus becoming a stepping-stone towards more efficient clinical assessment of brain development in young children.

Keywords: `Data fusion`, tensor decomposition, matrix decomposition, EEG,

*Corresponding author
Email address: `Noramon.Dron@ed.ac.uk` (Noramon Dron)

1. Introduction

Due to recent advances in medical technology, it is now possible to record multiple sources of information from patients. Clinicians are expected to provide better care and more accurate diagnosis from these rich data. However, one of the difficulties is how to find a method to combine and analyse data from different sources. A particularly relevant example is the case of the human brain, especially the developing brain of children, as childhood is a critical developmental period and the foundation of a child's future development and health [1, 2, 3]. Clinicians can acquire information about the brain function and structure through different modalities, such as electroencephalography (EEG) or magnetic resonance imaging (MRI). They usually expect that patterns in the anatomy and function would reflect changes in the clinical scores used to diagnose and monitor disease. In this context, this paper presents a novel data fusion algorithm for combining brain function, structure, and phenotypic data with regards to the development of preschool children.

The EEG can record fast changes in the electrical activity of the brain using multiple electrodes attached on the scalp [4]. The EEG is a standard clinical tool to assess the functional activity of the brain and helps to diagnose a number of conditions [4]. The advantage of EEG is its high temporal resolution, which is capable of detecting changes in brain activity in the range of milliseconds. However, the downside is its low spatial resolution, in the range of centimetres. Furthermore, it is difficult to perform accurate EEG signal source localisation due to the ill-pose nature of the inverse problem [5].

Another standard tool used in the clinic is MRI. Among other MRI modalities, structural MRI (sMRI) allows us to study the anatomy of the brain. sMRI provides sliced images of the brain from which a three-dimensional volume can be reconstructed. Clinicians use sMRI to inspect the physical appearance of the brain [6]. The advantage of sMRI is its precise spatial resolution, in the range of millimetres. However, sMRI capture brain image at a specific time, so it lacks the ability to acquire temporal functional information [7].

Both modalities have complementary advantages considering the high temporal resolution of the EEG and the detailed anatomical images of the MRI. They both are commonly used in epilepsy diagnosis [8]. Epilepsy is a chronic neurological disease associated with unprovoked epileptic seizures. It can start at any point in life, as early as an infant. Children with early-onset epilepsy (CWEOE) are prone to be cognitively and behaviourally impaired compared to healthy children of the same age [9, 10]. These deficits occur in 20% to 57% of CWEOEs, and cause more disabilities to their life than the seizures themselves [10, 11, 12, 13, 14].

Early diagnosis of those developmental deficits is pivotal to the child's quality of life. The clinical gold-standard entails the use of paper questionnaires to appraise potential impairment in children. However, such questionnaires are time-consuming and labour-intensive [15]. Moreover, the tests only detect the deficits after they show their signs. During the diagnosis of epilepsy, EEG and sMRI are acquired from CWEOEs. Therefore, exploiting those already existing data to estimate the risk of developmental impairment at each CWEOE is an appealing proposition.

To that end, we propose a new approach to predict developmental scores in CWEOE by combining EEG, sMRI, and phenotypic data through coupled tensor-matrix-matrix decomposition [16]. We consider phenotypic scores – including age, cognitive and behavioural scores – and analyse the changes in function and structure of the brain that may relate to such deficits. We then use that model to predict the score of the new diagnosis CWEOEs. Our approach is further validated using publicly available dataset from healthy children [17]. We present our main contributions below.

- We explore, for the first time, links between phenotypic scores, EEG, and sMRI data in very young CWEOEs by extending a recent model [18] to fuse data via a tensor-matrix-matrix decomposition. In comparison with the previous state of the art, we include the third modality (sMRI) and behaviour scores into the decomposition. This extends our feasibility study in [19], where the preliminary results showed changes in the volume of brain regions related to changes in scores that were in agreement with previous clinical studies [20, 21, 22]. We also include a larger dataset and healthy participants.

- We improve the initialisation step of the joint decomposition in comparison with [19] to reduce model cost and improve stability.
- We investigate the variability across subjects in the results of the decomposition. We analyse the change in the model when one subject is removed from the population and inspect which subjects contribute to larger changes in the model.
- We present how to use the data fusion model to further perform developmental score prediction. We show that developmental score prediction can be determined from the components estimated of the data fusion model through linear projection from the joint decomposition of three modalities. In addition, we exploit the common interactions between data modes [23, 24, 25] to predict phenotypic scores in cases where patients may only have one type of data available.

2. Background

A widespread way to fuse data from different modalities is to carry out joint factorisation of data arrays representing such data. The main premise of this approach is that the joint analysis allows us to decompose the data into common factors. In this way, we can reveal complementary profiles from multiple data sources [26]. Several research studies have successfully performed joint matrix factorisation on EEG and MRI. To this end, EEG has been jointly analysed with functional MRI (fMRI) using, for example, independent component analysis (ICA) [27], joint ICA, and independent vector analysis (IVA) [28]. In [29], ICA and canonical correlation analysis (CCA) were used. There have also been successful attempts to fuse the three modalities of EEG, fMRI and sMRI, for instance, using joint ICA and transpose IVA [30] or multi-set CCA (MCCA)[31]. In these approaches, feature data are put into separate matrices, or concatenated into a single matrix. Then, the algorithms proceed to decompose those matrices.

However, a matrix form may not be the most appropriate way to represent data related to brain structure and function [32]. This is particularly the case for EEG data, which naturally have more than two ways. That is, EEG samples can be indexed according to space (e.g., EEG channel), time and/or frequency, and subject, among others. Unfolding the data for using matrix decomposition can result in the loss of its multi-way properties. In addition, when applying matrix decompositions, one has to apply additional constraints to achieve the unique results in the factorisation [33, 34, 35]. Therefore, coupled matrix-tensor decompositions have recently been proposed as a more compatible model for fusing data that has more than two ways without destroying the structure of higher-order dataset. In this context, tensor decomposition allows us to study multi-way data arrays without disrupting the natural organisation and dependencies in the data. This can facilitate the extraction of shared information between domains [36].

Coupled matrix-tensor factorisations decompose high-order data (tensors) with two-way data (matrices) so we can analyse the common relationships between them. A number of studies have carried out coupled matrix-tensor decompositions of EEG and MRI data successfully. Simultaneous recording EEG and fMRI were jointly analysed based on canonical polyadic decomposition (CPD), also known as PARAFAC [37], in [38, 39, 40, 41] to study neural activity. CPD was also used to fuse EEG and fMRI with the purpose of characterising neurological disease such as schizophrenia [42] and epilepsy [43, 44]. CPD was further used to fuse more modalities of fMRI, sMRI, and EEG in the study of biomarkers in schizophrenia [16, 45]. Other factorisation approaches such as PARAFAC2 are used to combine measures of brain function – EEG and fMRI – in [46, 47, 48] and block term decomposition (BTD) [49, 50] has been implemented to study epilepsy in [43].

In this paper, we focus on a tensor-matrix-matrix factorisation model to jointly analyse EEG and sMRI in order to predict the developmental scores of children with epilepsy. We choose EEG and sMRI because these two data modalities are acquired during the epilepsy diagnosis, as described in Section 1. Moreover, previous studies suggested relationships between either of EEG or sMRI and developmental scores. For example, correlations between EEG features and cognitive scores were reported in [51, 18]. A reduction in thalamic volume (sMRI) in epileptic children with cognitive problems was found in [20] and correlations between sMRI and behavioural scores have been reported too [52]. In this context, we consider phenotypic information including age, cognitive score, and behavioural score as a third data modality in the analysis. In this way, we seek to explore links between cognitive and behavioural scores and brain data. To the best of our knowledge, this is the first time EEG, sMRI and these type of phenotypic scores from very young CWEOEs are combined for the analysis through tensor-matrix-matrix decomposition.

Notations and definitions in this paper follow the descriptions in [23]. Bold lowercase, such as $\mathbf{a}$, represents 1-way tensors or vectors. The outer product between vector $\mathbf{a}$ and $\mathbf{b}$ appears as $\mathbf{a} \circ \mathbf{b}$. Matrices, or 2-way tensors, are represented by bold uppercase letters, $\mathbf{X} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_J] \in \mathbb{R}^{I \times J}$. A multiway data array, so-called tensor, is denoted by calligraphic upper case letter $\mathcal{X} \in \mathbb{R}^{I \times J \times \cdots \times N}$. Khatri-Rao product between two matrices is represented as $\mathbf{A} \odot \mathbf{B}$.

3. Methods

3.1. Dataset

This study used two complementary datasets that are analysed separately to enhance the replicability of our approach. One contains data from CWEOE and the other one includes healthy subjects. They are described in the following subsections.

3.1.1. *Child Mild Institute*

Data of 35 healthy children aged 5-6 years-old with completed phenotypic data are selected from the public available dataset provided by Child Mild Institute - Healthy Brain Network (CMI) [17]. Resting state brain activity are recorded from high-density 111-channel EEG. MRI T1-weighted image are acquired from three different MRI machines, 1.5T Siemens Avanto, 3T Siemens Trim Trio and 3T Siemens Prisma. For CMI, we chose four developmental scores which available for every subjects and try to account for cognitive and behavioural ability. Wechsler Individual Achievement Test-III (WIAT) [53] and Clinical Evaluation of Language Fundamentals (CELF) [54] were chosen to represent the cognitive ability while Child Behaviour Checklist (CBCL) [55] and Strengths and Weaknesses Assessment of ADHD and Normal Behavior (SWAN) [56] were chosen to represent the behaviour of the children. The number of the subjects categorised by their developmental score are presented on Table 1. Note that two out of three subjects have mild cognitive deficit, as indicated by at least one of the cognitive assessments. For example, two subjects had mild impairments in behaviour, as indicated by the SWAN, but not according to CBCL.

3.1.2. *NEUROPROFILE*

A prospective population-based study [57] provided a clinical data of 30 preschool children (<5 y.o.) who were diagnosed with epilepsy. This dataset is recorded as part of the patient's clinical care across NHS Fife and Lothian, UK with the written consent to study their data from patient's parents. Resting-state EEG recorded at 20 scalp electrodes and average referenced. The signal are processed to be seizure free for the analysis. Structural MRI T1-weighted image are recorded from 1.5T Siemens Espree. Children also participated in cognitive assessment with Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III) [58] and behavioural assessment with Adaptive Behavior Assessment System-General Adaptive Composite (ABAS-GAC) [59]. The total number of the subjects according to their psychometric score are showed on

Table 1: Grouping of healthy control ($n = 35$) according to their level of cognitive and behavioural impairment.into normal (± 1 SD), mild/moderate (-1 to -2 SD) and severe (< -2 SD) impairment.

		Behavioural score (SWARN and CBCL)		
		Normal	Mild	Severe
	Normal	29	1	1
Cognitive score	Mild	3	1	0
(WAIT-III and CELF)	Severe	0	0	0

Table 2: Grouping of the CWEOE ($n = 30$) according to their level of cognitive and behavioural impairment into normal (± 1 SD), mild/moderate (-1 to -2 SD) and severe (< -2 SD) impairment.

		Behavioural score (ABAS-GAC)		
		Normal	Mild	Severe
	Normal	11	3	2
Cognitive score	Mild	4	2	1
(Bayley III)	Severe	0	2	5

Table 2.

3.2. Data Preparation and Tensor Construction

This section describes how data were prepared. Figure 1 illustrates the tensor and matrices construction to be used in this work. In order to perform data fusion, we need at least one shared domain across the modalities. By aligning the data in the same subject order, the *[Subject]* mode is treated as shared, which means data from different modality are matched and can be linked across the subjects.

Both datasets were processed as similarly as possible. The sampling rate of resting-state EEG from CMI and NEUROPROFILE are 500 Hz and 511 Hz, respectively. The reference and auxiliary channels are removed from the analysis, result in 111 channels for CMI and 20 channel for Neuroprofile. The data was filtered from 0.5Hz to 45Hz to remove power line noise. Manual and automatic rejection from Fieldtrip toolbox [60] are performed to remove other artefacts. Then, the EEG signals are split into non-overlap two-second long trials. The frequency spectrum was calculated and averaged over the trials to provide a general spectral profile of the resting-state EEG. Subsequently, tensor is constructed to store all EEG data in modes *[Subject]* $\times$ *[Spectral]* $\times$ *[Channel]* with size $35 \times 81 \times 111$ for the CMI and $30 \times 79 \times 20$ for the NEUROPROFILE data.

Due to the different scales in the developmental scores, we standardise them by converting them into z-score before placing in the phenotypic matrix together with the age of participants. The matrix construct in the same subject order as EEG tensor with mode *[Subject]*$\times$*[Score]*. For CMI, age, WIAT-III, CELF, SWARN, and CBCL are arranged into a score matrix at the size of 35×5. For the NEUROPROFILE, a score matrix size 30×3 is arranged by order of age, Bayley-III, and ABAS-GAC.

For MRI data, T1-weighted images from both datasets were processed and their quality was manually assessed using FMRIB Software Library (FSL) 5.0 [61]. After removing children data with severe motion artefacts, MRI sequences were segmented into eight brain regions following the study in [20], by FMRIB's Integrated Registration and Segmentation Tool (FIRST) [62] in FSL. We calculate the volume of the following eight segmented regions, which are the left and the right thalamus, caudate, putamen, and pallidum. Then each regions are normalised by head size of the subjects to be in the same standard space. Then MRI data are stored in the same order as EEG and score in a matrix with mode *[Subject]*$\times$*[Region]* at the size of 35×8 for the CMI and 30×8 for NEUROPROFILE.

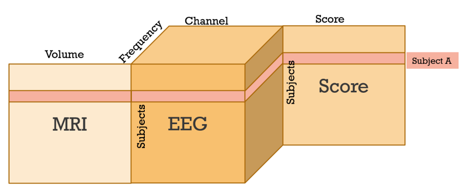

Figure 1: Joint coupled tensor-matrix representation used in this work. The data are linked across the subject domain through three different modalities.

3.3. Data Fusion and Tensor-matrix-matrix Decomposition

In order to fuse the EEG tensor with the MRI and phenotypic score matrices, we adapt a tensor factorisation model coined Joint EEG-Development Inference (JEDI) [18]. JEDI is a matrix-tensor factorisation model based on CPD that can preserve multi-dimensional relations in the data and can work with different type of data expressed as either matrix or tensor. In previous JEDI work, the model combined EEG and cognitive attributes of healthy pre-adolescent children (<11 y.o.). In this work, we decrease the age of subject to preschool children (<7 y.o.) on healthy cohorts and having clinical children (<5 y.o.) with record of epilepsy. Reducing age plays an important role in assist early detection of impairment.

To perform the fusion of three data types all at once, CPD [37] is chosen. The coupled tensor-matrix-matrix factorisation between EEG, phenotypic score, and sMRI will consider the EEG tensor $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$, the score matrix $\mathbf{S} \in \mathbb{R}^{I\times N}$, and the sMRI matrix $\mathbf{M} \in \mathbb{R}^{I\times M}$. The score matrix $\mathbf{S}$ is decomposed into component factors $\mathbf{A} \in \mathbb{R}^{I\times R}$, and $\mathbf{D} \in \mathbb{R}^{N\times R}$, respectively, and $\mathbf{S} \approx \mathbf{A}\mathbf{D}^T$. Similarly, the MRI volume matrix $\mathbf{M}$ has component factors $\mathbf{A} \in \mathbb{R}^{I\times R}$, and $\mathbf{E} \in \mathbb{R}^{M\times R}$, respectively, and $\mathbf{M} \approx \mathbf{A}\mathbf{E}^T$, where subject domain $\mathbf{A}$ is a shared mode among all three modalities.

CPD is a commonly used tensor decomposition model that decomposes a tensor into a sum of rank-1 tensors and has the advantage of being unique under mild constraints [63]. CPD is widely used in brain signal analysis [23, 24] and has been proven to extract developmental profiles in children's brain activity [24, 51, 18]. For example, in the CPD model, the three-way tensor $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$, with rank R can be written as

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{1}$$

where $r = 1, 2, \cdots, R$ with $\mathbf{a}_r \in \mathbb{R}^I$, $\mathbf{b}_r \in \mathbb{R}^J$, $\mathbf{c}_r \in \mathbb{R}^K$. This can also be written as

$$x_{ijk} = \sum a_{ir} b_{jr} c_{kr} + \epsilon_{ijk}, \tag{2}$$

where $i = 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots, K; r = 1, \cdots, R$ with x_{ijk}, a_{ir}, b_{jr}, c_{kr}, and ϵ_{ijk} as element of $\mathcal{X}$, domain $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}$, and residual $\epsilon \in \mathbb{R}^{I \times J \times K}$, respectively

The coupled tensor-matrix-matrix decomposition was carried out with the structured data fusion (SDF) framework [36] implemented in Tensorlab [64] in Matlab. JEDI initialised EEG and score factor components using multilinear SVD [64] and randomly, respectively. In contrast, we initialise the factor components in the coupled CPD via ordinary CPD method [65] to increase the fitness. The initialised step firstly compute CPD of EEG tensor $\mathcal{X}$ to obtain initial value for factor $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. The remaining factor $\mathbf{D}$, and $\mathbf{E}$ are computed by matrix multiplication with previously obtained factors $\mathbf{A}$.

We set the subject domain *[Subject]* $\mathbf{A} \in \mathbb{R}^{I \times R}$ as a shared domain for the SDF. Additionally, to improve the interpretation and ground the factors into the realistic boundary, we impose a non-negative constraint to every domain of EEG tensor and MRI matrix, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$. As the subject domain $\mathbf{A}$ is an overlap domain, this non-negativity also affects the subject domain of the score matrix. Next, we imposed regularisation to all the modes of tensors and matrices in order to reduce the overfitting. $L1$ regularisation was set on *[Subject]* modes across the domain to promote sparse responses. $L2$ regularisation was set on the domains other than *[Subject]*. Relative weights λ_{1-5}, also called hyperparameters, were set to define the contribution of EEG tensor $\mathcal{X}$, MRI matrix $\mathbf{M}$, score matrix $\mathbf{S}$, $L1$, and $L2$ regularisation toward the SDF structure. We impose all these hyperparameters to minimise the cost function which extended from [18] and can be written as

$$\begin{aligned}\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{E},R} \quad & (\lambda_1/2)||\mathcal{X} - M_{CPD}(\mathbf{A},\mathbf{B},\mathbf{C},R)||_F^2 + \\ & (\lambda_2/2)||\mathbf{S} - M_{CPD}(\mathbf{A},\mathbf{D},R)||_F^2 + \\ & (\lambda_3/2)||\mathbf{M} - M_{CPD}(\mathbf{A},\mathbf{E},R)||_F^2 + \\ & (\lambda_4/2)(||vec(\mathbf{B})||_F^2 + ||vec(\mathbf{C})||_F^2 + \\ & ||vec(\mathbf{D})||_F^2 + ||vec(\mathbf{E})||_F^2) + \\ & (\lambda_5/2)||vec(\mathbf{A})||_1, \qquad (3)\end{aligned}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$, are the component factor of each mode that previously mentioned. R and M_{CPD} represent the number of components and refer to joint CPD decomposition, respectively. The hyperparameters λ_{1-3} are the relative weights for increase the fittings to $\mathcal{X}$, $\mathbf{S}$, $\mathbf{M}$ while λ_4 and λ_5 are $L2$ and $L1$ regularisation, respectively.

The optimum values of λ_{1-5} are explored within the optimisation process. The value of hyperparameter λ_1, λ_4 and λ_5 are set to varied from 0.01 to 100. While λ_2 and λ_3 are varied from 0.1 to 10 then multiplied by λ_1 to prevent tensor or matrices to overwhelm the fusion. Moreover, we also varied number of components R from 2 to 6. The hyperparameter λ_{1-5} and R with lower costs are recorded to further evaluate the ability to predict the developmental score.

3.4. Variability across subjects

We also investigate the variability that affects the decomposition result due to changes in subjects within the population. We use a leave-one-out setting to assess that variability. This allows us to investigate the effect of having one of the subjects left out of the group used to compute the components. We hypothesise that the factor values will not change considerably due to this change in the composition of the dataset used in the decomposition and that the components will maintain the main trends. In particular, we remove one subject out from the decomposition at a time, then proceed with the same set of hyperparameters resulted from grid search, and compute the decomposition. Thus, we have a total of 35 new runs for CMI and 30 new runs for NEUROPROFILE, each of which has one corresponding subject left out of the analysis. Due to the fact that the order of the factors may vary from run to run, the components are manually matched into the same order obtained from the decomposition of the whole dataset (order shown in Fig. 3). The subject mode is rearranged by shifting the subject that was left out. Once the components have matched, the mean and SD of each component for each factor are calculated and plotted separately to assist visualisation. We also compute the average correlation coefficient across all components extracted from the complete dataset and from the factors computed in the leave-on-out procedure.

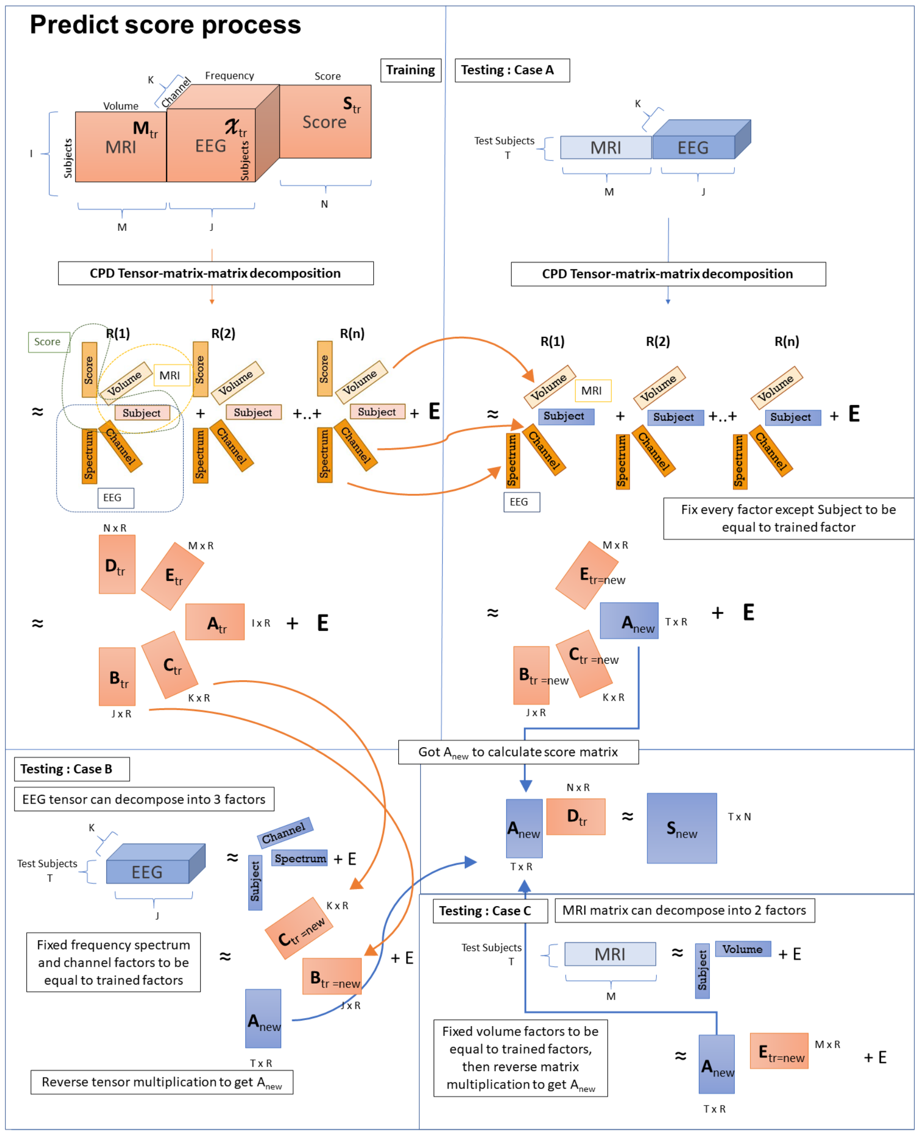

Figure 2: Visual presentation of score prediction through coupled tensor-matrix-matrix decomposition. **Step 1.)** Perform tensor-matrix-matrix decomposition of the training dataset and obtain training components $\mathbf{A}_{tr}$, $\mathbf{B}_{tr}$, $\mathbf{C}_{tr}$, $\mathbf{D}_{tr}$, and $\mathbf{E}_{tr}$. **Step 2.)**: <u>Case A</u> Both EEG and sMRI available from new patients. Perform tensor-matrix decomposition of the tested dataset by fixing mode $\mathbf{B}_{new}$, $\mathbf{C}_{new}$, $\mathbf{E}_{new}$ to be the same value as $\mathbf{B}_{tr}$, $\mathbf{C}_{tr}$, $\mathbf{E}_{tr}$, to get $\mathbf{A}_{new}$ that holds the information of new patients. **Step 2.)**: <u>Case B</u> When only EEG is available from new patients. We apply tensorial algebra to project $\mathbf{A}_{new}$ by fixed mode $\mathbf{B}_{new}$, and $\mathbf{C}_{new}$, to be equal to $\mathbf{B}_{tr}$, and $\mathbf{C}_{tr}$. **Step 2.)**: <u>Case C</u> Only sMRI is available from new patients. Similarly to case B, we fixed mode $\mathbf{E}_{new}$ to be equal to $\mathbf{E}_{tr}$. Then perform matrix multiplication with the pseudo-inverse to get $\mathbf{A}_{new}$ that hold the information of new patients. **Step 3.)** Once $\mathbf{A}_{new}$ is known, the score matrix $\mathbf{S}_{new}$ of the new patients can be estimated from $\mathbf{S}_{new} = \mathbf{A}_{new}\mathbf{D}_{tr}^{T}$

3.5. Score Prediction

3.5.1. Prediction via tensor-matrix-matrix factorisation

In this study, we predict the developmental score of the new children from three possible cases. First, when both EEG and MRI are available, and then where only one of them is known.

Assume that we have data from subjects for whom we know the developmental scores. This data can be considered as the training set: $\mathcal{X}_{tr}$, $\mathbf{M}_{tr}$ and $\mathbf{S}_{tr}$. Applying the data fusion model, we can estimate training components $\mathbf{A}_{tr}$, $\mathbf{B}_{tr}$, $\mathbf{C}_{tr}$, $\mathbf{D}_{tr}$, and $\mathbf{E}_{tr}$. Once new unseen (e.g., test) data is available, $\mathcal{X}_{new}$ and $\mathbf{M}_{new}$, the goal is to predict the developmental scores from only the EEG, only the MRI, or both EEG and MRI. We assume that the new kids are part of the same population as our trained dataset. Thus when new kids are added to the trained dataset and jointly decomposed, the factorised components are assumed to stay the same and be equal across the old and new datasets. Therefore, $\mathbf{B}_{tr} = \mathbf{B}_{new}$, $\mathbf{C}_{tr} = \mathbf{C}_{new}$, $\mathbf{D}_{tr} = \mathbf{D}_{new}$, $\mathbf{E}_{tr} = \mathbf{E}_{new}$ since common interactions held within *[Spectral]*, *[Channel]*, *[Score]* and *[Volume]* domains remain. With this assumption, developmental score matrix of the new kids can calculated in the following cases.

Firstly, when both sMRI and EEG are available for new subjects, $\mathcal{X}_{new}$ and $\mathbf{M}_{new}$ will be integrated into a CPD data fusion model without the score matrix and perform tensor-matrix decomposition with $\mathbf{B}$, $\mathbf{C}$, $\mathbf{E}$ fixed as $\mathbf{B}_{tr}$, $\mathbf{C}_{tr}$, and $\mathbf{E}_{tr}$. The cost function will be similar to the full decomposition equation but without the score matrix and its regularisation. The equation can be written as:

$$\begin{aligned}\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{E},R} \quad & (\lambda_1/2)||\mathcal{X}_{new} - M_{CPD}(\mathbf{A}_{new}, \mathbf{B}_{tr}, \mathbf{C}_{tr}, R)||_F^2 + \\ & (\lambda_3/2)||\mathbf{M}_{new} - M_{CPD}(\mathbf{A}_{new}, \mathbf{E}_{tr}, R)||_F^2 + \\ & (\lambda_4/2)(||vec(\mathbf{B}_{tr})||_F^2 + ||vec(\mathbf{C}_{tr})||_F^2 + \\ & ||vec(\mathbf{E}_{tr})||_F^2) + \\ & (\lambda_5/2)||vec(\mathbf{A}_{new})||_1, \end{aligned} \quad (4)$$

The model will estimate $\mathbf{A}_{new}$ that hold the information of the new subjects in the context that other modes are the same as the training population. Then, we can estimate the unknown score matrix as $\mathbf{S}_{new} = \mathbf{A}_{new}\mathbf{D}_{tr}^T$, as visually demonstrated in Fig. 2, case A.

In the case only one of the new data was available, for example, the EEG, the training data are firstly decomposed in the same way as the previous case. Then, the weight $\mathbf{W}$ for *[Subject]* domain to determine the predicted score can be calculated from training EEG data $\mathcal{X}_{tr}$ unfolded along the subject domain, as:

$$\mathbf{X}_{tr} = \mathbf{A}_{tr}(\mathbf{B}_{tr} \odot \mathbf{C}_{tr})^T \tag{5}$$

$$\mathbf{X}_{tr} = \mathbf{A}_{tr}\mathbf{W}, \tag{6}$$

where $\mathbf{W}$ is a weight matrix for *[Channel]*$\times$*[Frequency]*. Then, to get the estimated *[Subject]* domain $\mathbf{A}_{new}$ for score prediction, we calculate non-negative least square projection ($NN(\cdot)$) between $\mathbf{W}$ and the unfolded new EEG $\mathcal{X}_{new}$ along with the subject domain as:

$$\mathbf{X}_{new} = \mathbf{A}_{new}(\mathbf{B}_{new} \odot \mathbf{C}_{new})^T \tag{7}$$

$$\mathbf{A}_{new} = NN(\mathbf{X}_{new}\mathbf{W}), \tag{8}$$

where $\mathbf{A}_{new}$ is an estimation of the new subject component for this case. The $NN(\cdot)$ is used instead of standard regression because we assume non-negativity in extracted modes. Then $\mathbf{A}_{new}$ can be multiplied with the transposed factor matrix *[score]* from the training session to get the predicted score as $\mathbf{S}_{new} = \mathbf{A}_{new}\mathbf{D}_{tr}^T$. Fig. 2, case B illustrated this case where only EEG are available.

The similar approach can be applied if only the sMRI is available and represent in Fig. 2, case C. The training sMRI $\mathbf{M}_{tr}$ is unfold into $\mathbf{A}_{tr}\mathbf{E}_{tr}^T$ with the weight matrix $\mathbf{W} = \mathbf{E}_{tr}^T$. Then, $\mathbf{A}_{new}$ can be calculated from $NN(\mathbf{M}_{new}\mathbf{W})$. The predicted score is $\mathbf{S}_{new} = \mathbf{A}_{new}\mathbf{D}_{tr}^T$.

3.5.2. Evaluation of score prediction

As a benchmark, we used a support vector machine (SVM) [66] to predict scores from EEG and/or sMRI data. SVM is a supervised machine learning algorithm used in many fields for classification and regression analysis, and recently used in the analysis of epileptic EEG and MRI [67, 68, 69]. Therefore, SVM is adopted as a benchmark for out proposed CPD model. In this paper, we focus on the predictive ability of SVM regression in order to compare with the predicted result from the proposed model. We fit a SVM regression model into a ten-fold cross-validation setting to perform score prediction [70]. EEG tensor was unfolded along subject dimension into matrix size 30×8991 and 30×1580 (subjects $\times$ channel by frequency combinations) for CMI and NEUROPROFILE, respectively. Then, the unfolded EEG is concatenated along the subject domain with MRI matrix to create a trained matrix for SVM before performing score prediction. The parameters used in SVM were determined through automatic hyperparameter optimisation to find the best fit.

Moreover, we also tested the JEDI model as another benchmark for prediction. We also used grid search in this case to determine the optimal hyperparameters for JEDI, when only EEG is available as described in [18]. The SVM, JEDI, and our CPD model were compared using the same ten-fold cross-validated set-up, which will be described in section 3.6. Thus each fold can be compared side by side for their predictive performance.

In addition to the benchmark models; SVM and JEDI, a two-tailed t-test with a 5% significance level is tested to check the mean distribution of actual and predicted score.

3.6. Model Optimisation

In our experiments, the hyperparameters resulting in the least cost are selected to explore the common interaction between modalities. Grid search was adopted to explore the combination of hyperparameters λ_{1-5} and number of components R from equation 3.3. We fixed a regularisation to subject $\lambda_5 = 1$. Then λ_1 and λ_4 are varied from 0.01 to 100 in nine logarithmic steps while λ_2

and λ_3 are varied from 0.1 to 10 in 5 logarithmic steps then multiplied by λ_1 followed [18].

For the score prediction process, we adopted a ten-fold nested cross-validation [71], which was divided into outer and inner cross-validation. In outer cross-validation, ten percent of all data is considered a new test dataset with unknown developmental score and the rest of the dataset is assigned to the inner cross-validation. Then another ten and ninety percent of the remaining data in inner cross-validation are considered as validation and training, respectively. In this setting, the optimum hyperparameters yielding the most accurate score prediction from the inner training to validation is then transferred to compute the developmental score for new patients.

For model evaluation, JEDI is set to use the same grid search and cross-validation setting as our CPD model. Likewise, SVM is set to the same cross-validation. However, SVM cannot predict the whole score matrix all at once. Therefore each score is predicted separately before merge into the predicted matrix for later evaluation.

4. Results and Discussion

4.1. Profiles of the components

The component factors with the lowest cost from grid search is plotted in Fig. 3. The grid search resulted in optimal values of $R = 6$, $\lambda_{1-5} = 100$, 316.2278, 10, 0.01, and 1, respectively, for CMI dataset with $cost = 0.2403$; and $R = 6$, $\lambda_{1-5} = 100$, 31.6228, 10, 0.1, and 1, respectively, for the NEUROPROFILE with $cost = 0.2368$. Interestingly, both hyperparameter sets are nearly identical. The only difference is the relative weight imposed on the score matrix, 316.2278 for CMI and 31.6228 for the NEUROPROFILE. The higher value in score matrix weight in CMI indicated that the model needs higher weight to achieve a similar low cost. This may occur due to the fact that CMI have little number of subjects with score deficits when compare to NEUROPROFILE. By improving the initialisation step as mentioned in 3.3, the NEUROPROFILE

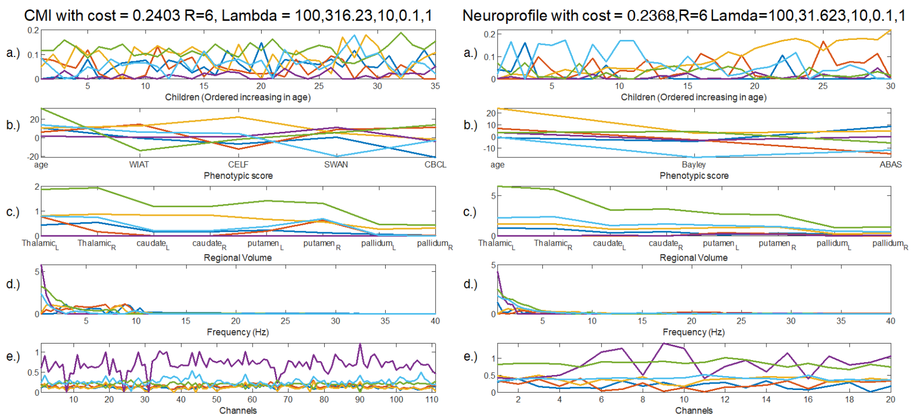

Figure 3: Plots of the components extracted by our data fusion model. The components in the left graph were retrieved from the CMI dataset by using hyperparameters $R = 6$, $\lambda_{1-5} =$ 100,316.23,10,0.1,1, respectively. The components in the right graph are from the NEUROPROFILE dataset using hyperparameters $R = 6$, $\lambda_{1-5} =$ 100,31.623,10,0.1,1, respectively. (a.) Components for the subject mode ordered in the horizontal axis from youngest to oldest. (b.) Phenotypic components ordered in the x-axis as age, followed by the psychological scores. (c.) Components for the volume of eight brain regions: Thalamic (L,R), caudate (L,R), putamen (L,R), and pallidum (L,R), as indicated on the horizontal axis. (d.) Components for the frequency spectrum (in Hz. (e.) Components corresponding to the EEG channels.

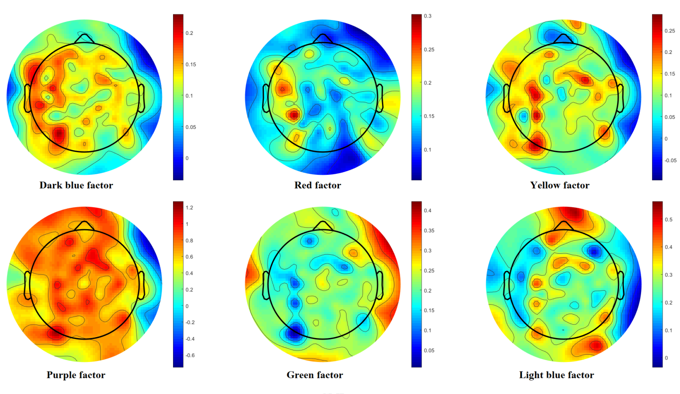

Figure 4: CMI topoplot of 111 EEG channels from channel factor components. (This is an alternative representation to panel e) on the left part of Fig. 3.)

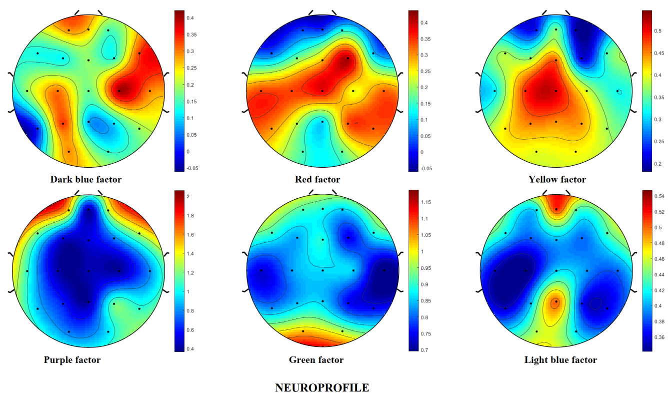

Figure 5: NEUROPROFILE topoplot of 20 EEG channels from channel factor components. (This is an alternative representation to panel e) on the right part of Fig. 3.)

lowest cost was reduced from 0.3906 [19] to 0.2368. Moreover, this initialised resulted in better stability. The cost variability was tested in the same manner as in [19], where we fixed the hyperparameters achieving from grid search and compute the resulting cost 200 times. The result shows the improvement in mean cost and overall cost compared to [19], where cost range from 0.3906 and can spike up to 2.500 on NEUROPROFILE data. The cost is more stable with a lower value and has no abrupt higher cost. The cost for the NEUROPROFILE dataset ranges from 0.2368 to 0.3287, while CMI ranges from 0.2403 to 0.3504.

Fig. 3 (a.) shows components for the subject mode, with increasing age on the x-axis. There is no trend associated with increased age in the profile components of CMI dataset. However, the yellow component on NEUROPROFILE demonstrates the component increasing with age.

Fig. 3 (b.) shows the components associated with the mode of the phenotypic scores as shown on the x-axis. The first value represents the age in the phenotypic score, which, in CMI, indicates that a green component is associated with age. However, when tracing back to Fig. 3 (a.), the green line does not show the obvious increasing trend when the children where ordered from younger to older. When looking at the age position on the x-axis in NEUROPROFILE, a yellow component is also associated with age. Now, when tracing back to section (a.), we can see the increasing trend with age in this case. The difference in both datasets' age gap may be the reason for the different trend in section (a.) because CMI subjects consist of 5-6 y.o., while NEUROPROFILE has children from 0.2-5 y.o., thus accounting for a broader age range.

The remaining positions of the x-axis of Fig. 3 (b.) feature developmental scores. The children are considered as normal, having better than normal performance, and impaired, when the graph lies around zero, positive, and negative values, respectively. For CMI, Fig. 3 (b.) WIAT and CELF are cognitive scores, and SWARN and CBCL represent behavioural scores. The yellow and purple components are associated with subjects who generally perform well in all developmental tests, with yellow suggesting that subjects have better cognition as of CELF and purple with slightly better behaviour in SWARN. The red component is associated with subjects with mild cognitive impairment indicated only by CELF but who perform well in other scores. The light blue component accounts for subjects with poor SWARN but in normal range with CBCL and perform well in cognitive test. The blue component represents subjects with poor CBCL but not in SWARN for behavioural ability, with slightly low in cognitive ability indicated by CELF. The green component represents subjects with lower cognition in WIAT but who appear normal in other scores. However, this component is heavily associated with age. For CMI, these components make sense because the majority of the subjects are in normal or better range of developmental scores resulting in better performance of different scores going into the same direction. Only a handful subjects indicated to be deficit by either one of the scores in cognitive or behavioural field and rarely both, making the graph only associate the deficit of each score separately.

When compared to NEUROPROFILE, where more subjects are labelled to have deficits in either or both developmental scores, we see a nearly parallel line in several components on Bayley and ABAS scores on Fig. 3 (b.) The yellow component represents subjects with normal developmental score and associated with age. The purple component represents subjects with normal developmental scores. The blue component represents subjects with normal cognitive score and who perform well in behavioural score, similarly to green component that represents normal behaviour and better scores in cognition. The red component represents subjects with behavioural impairment but normal in cognitive ability while light blue component represents subjects who have developmental impairment in both fields.

Fig. 3 (c.) illustrates the sMRI volume factors in eight regions. We can see similar trends in brain regional factors from both dataset, especially in the CMI green component and NEUROPROFILE yellow component. Both components are associated with volumes in thalamus, caudate and putamen that change with age when traced back to Fig. 3 (b.), something that agrees with previous studies [3, 72]. In CMI, the green component is also associated with lower cognitive ability from WIAT-III but with the heavily link to age. Therefore, it is hard to assume that this component can fully represent cognitive impairment. Thus, we focus on the yellow and red components which represent subjects who perform well and poorly in the language assessment CELF in CMI. In Fig. 3 (c.), the decomposition points out that cognitive ability is associated with differences in volume between right thalamus [72], caudate [73] and left putamen [21, 74]. Components blue and light blue represent behavioural impairment indicated by SWARN and CBCL, respectively. Both graphs are similar across the sMRI volume factor in Fig. 3 (c.), with differences only in right putamen, which follows the study [75] that found that children with ADHD or behavioural problems often have different volumes between left and right putamen. Finally, we compare this with the purple graph that represents subjects with normal behaviour and specifically good in SWARN. We can see that there is a volume different in thalamus [76], and right putamen [75] in subjects with deficits.

With the same context for NEUROPROFILE, we compare the blue and red components for better and worse behaviour as indicated by ABAS-GAC. Fig. 3 (c.) shows behavioural deficits associated with mainly lower volume in thalamus [76], and followed by caudate [77]. Then, we compare the green and light blue components for better and worse cognition in the Bayley score. Fig. 3 (c.) shows different in every brain regions except pallidum and significantly different in thalamus which is known to be associated with cognitive ability [20, 72]. The purple component that associated with subjects who are in normal range in both scores shows a very similar trajectory with the red component.

Fig. 3 (d.) illustrates the components for the frequency spectrum mode. Focusing on the frequency range 5-10 HZ in both groups of children, the healthy controls in the CMI have higher relative amplitude than CWEOE. The frequency profiles show most of the activity in low frequency band. This is in agreement with the fact that we have analysed EEG resting state activity from young children [78]. The components can be related to cognition and behaviour more clearly in the sMRI mode due to previous literature on the topic [20]. However, the inclusion of EEG allows us to evaluate their score predictive ability as it will be shown in the next subsection.

Fig. 3 (e.) demonstrates the profile the components in the channel mode. We plot the factors as topographic maps using the EEGlab toolbox[79] and match them with the colour line graph to assist the analysis. Fig. 4 depicts the topoplot of 111 channels from CMI data and Fig. 5 illustrates the 20 channels from the NEUROPROFILE data.

4.2. Variability Across Subject

We hypothesised that removing one subject from the population and then performing tensor-matrix-matrix decomposition with the same hyperparameters as in the previous section would yield a similar decomposition profile. This allows us to analyse the variability across each subject. The mean and SD graphs from these variability test for the CMI factors are plotted on Fig. 6 - 10, while the NEUROPROFILE factors are plotted on Fig. 11 - 15. The order of the factor component 1-6 follows the colour blue, red, yellow, purple, green and light blue of the component profile plots from Fig. 3. The major trends in the mean factor component graphs resemble those of Fig. 3, with some shifts in shape and intensity in some components. For both datasets, volume and score factor components are the closes resemble to the trend in Fig. 3, followed by

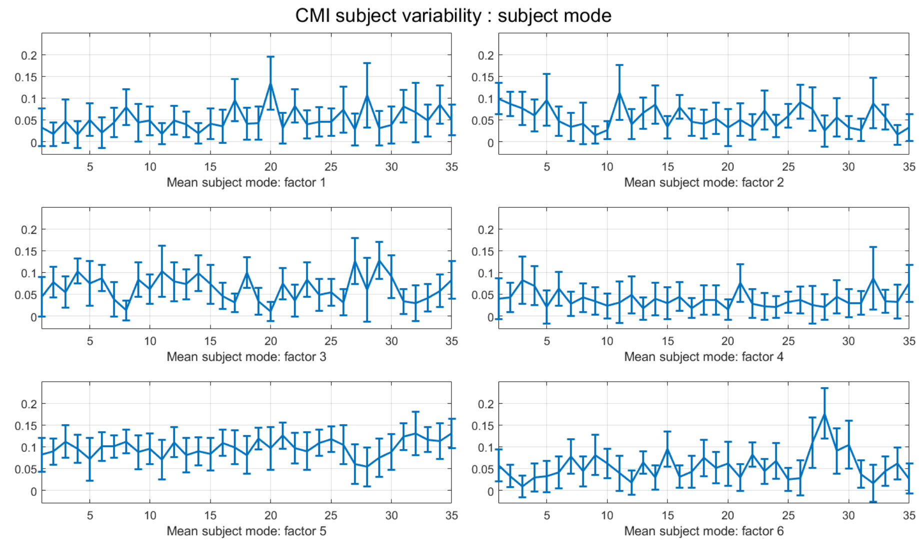

Figure 6: Mean and SD of CMI subject mode from the subject variability analysis.

subject, frequency and channel components.

The association between factors components linked in section 4.1 mostly remain the same. For CMI, blue (factor 1), yellow (factor 3), green (factor 5) and light blue (factor 6) have subtle change in trend from the mean variability graph. However, the association with developmental deficits and differences in brain volume remain the same as discussed in section 4.1. The red component (factor 2) remains representing children with poor language ability (CELF) and normal in other scores, but now only obviously associate with difference in caudate region. The volume difference in thalamus and putamen are not as noticeable when compared to the normal children on yellow (factor 3) component on Fig. 8. The purple component (factor 4) remained the same on volume trend in Fig. 8. However, instead of representing children with normal score in general, the mean graph is now bending toward lower WIAT on Fig. 7, factor 4. For NEUROPROFILE, all factor components in Fig. 12 represent the same developmental deficits as in previous section 4.1. The trend in MRI volume on Fig. 13 also stayed the same with slightly shift in Y-axis. However, this shift

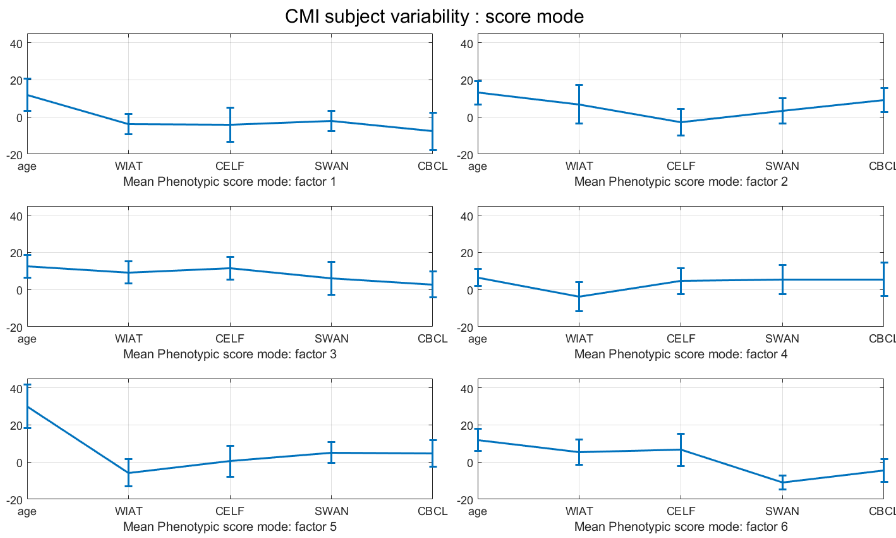

Figure 7: Mean and SD of CMI score mode from the subject variability analysis.

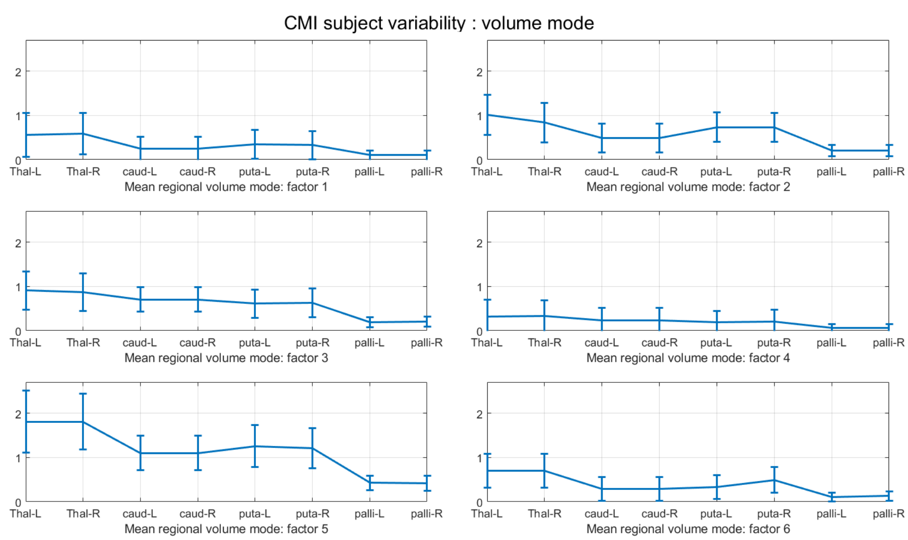

Figure 8: Mean and SD of CMI regional volume mode from the subject variability analysis.

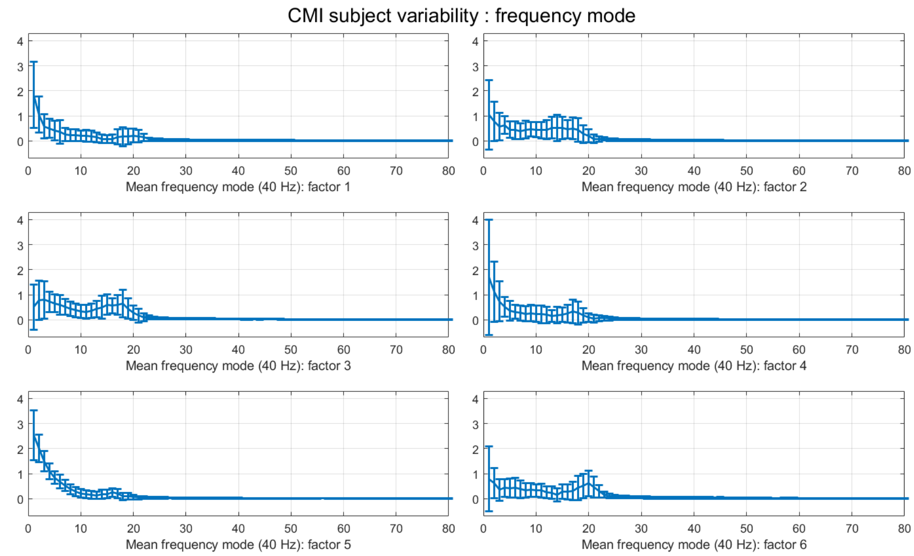

Figure 9: Mean and SD of CMI frequency mode from the subject variability analysis.

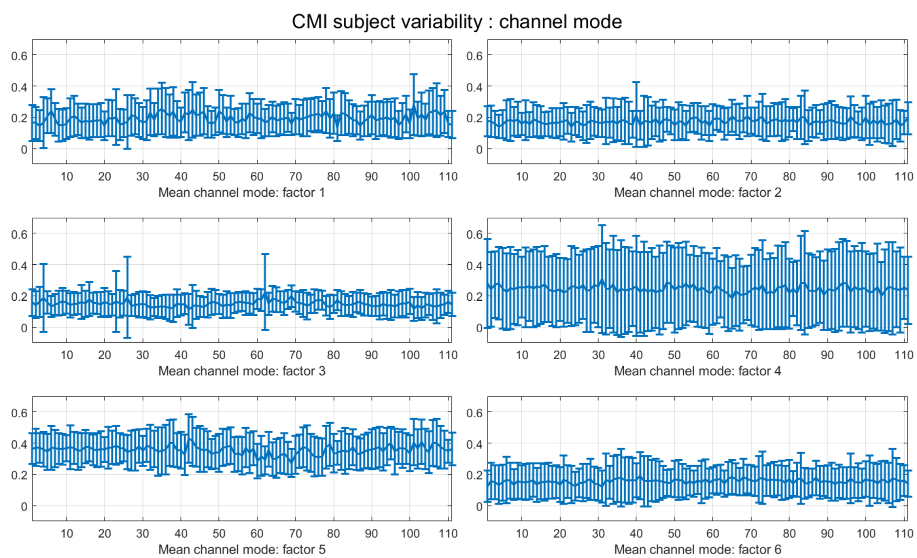

Figure 10: Mean and SD of CMI channel mode from the subject variability analysis.

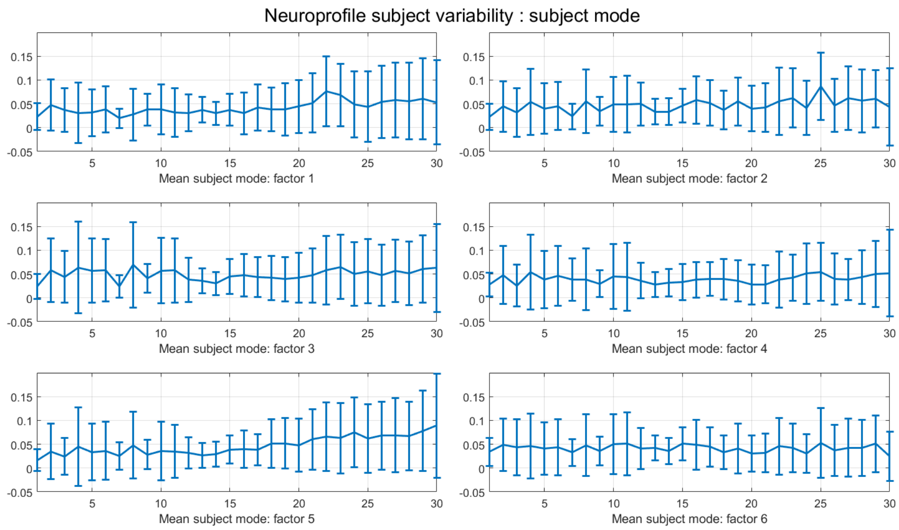

Figure 11: Mean and SD of NEUROPROFILE subject mode from the subject variability analysis.

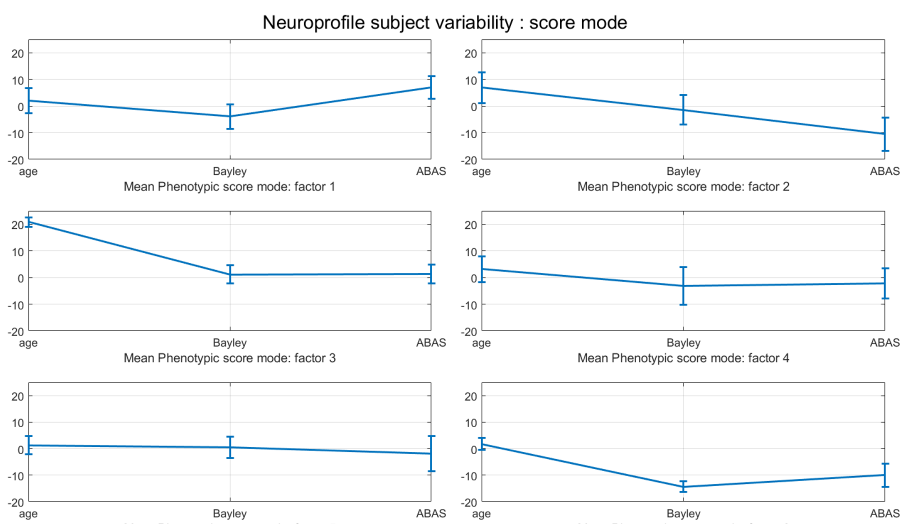

Figure 12: Mean and SD of NEUROPROFILE score mode from the subject variability analysis.

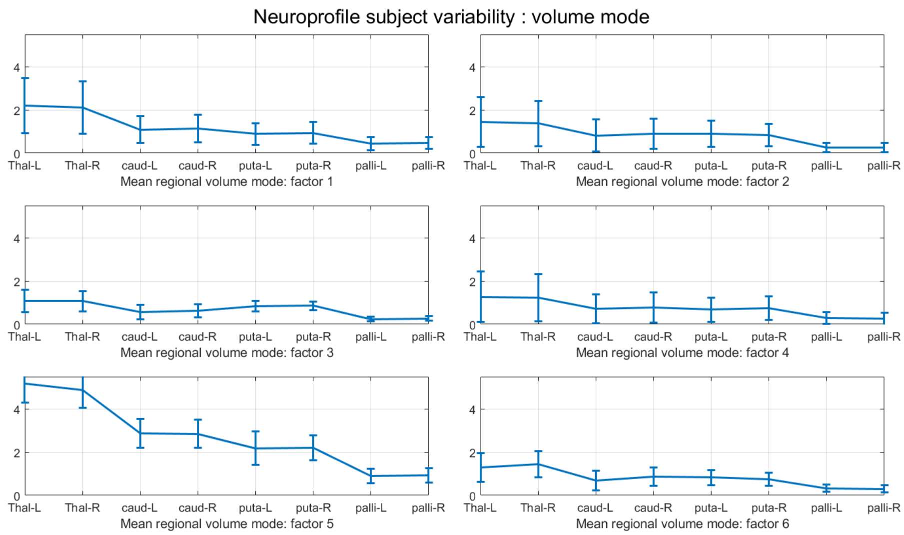

Figure 13: Mean and SD of NEUROPROFILE regional volume mode from the subject variability analysis.

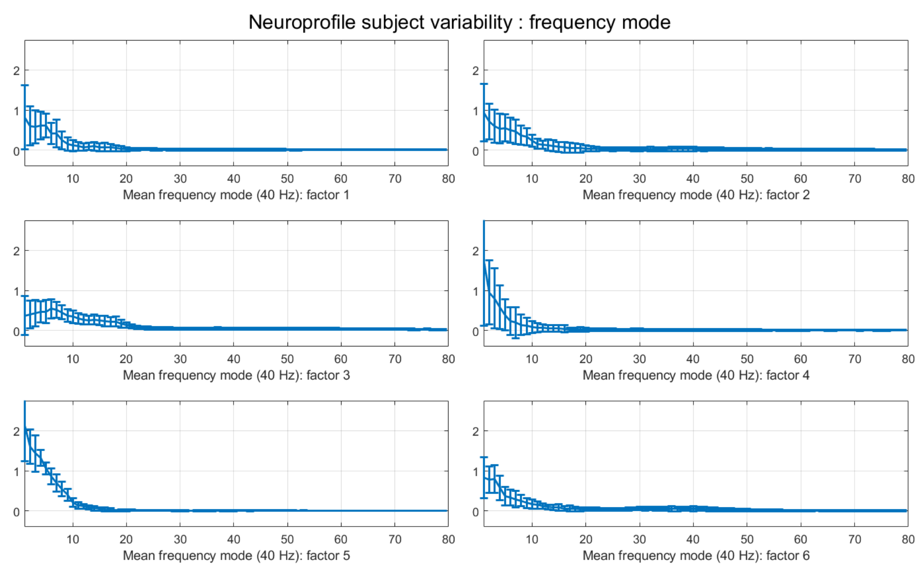

Figure 14: Mean and SD of NEUROPROFILE frequency mode from the subject variability analysis.

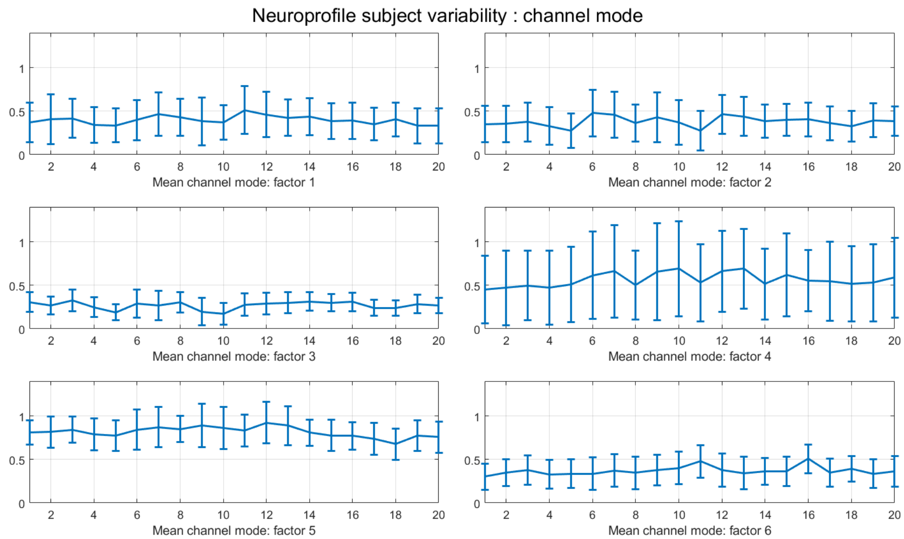

Figure 15: Mean and SD of NEUROPROFILE channel mode from the subject variability analysis.

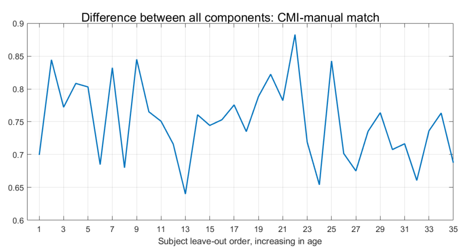

Figure 16: Correlation coefficients between the CMI whole dataset and the components estimated in the leave-one-out procedure.

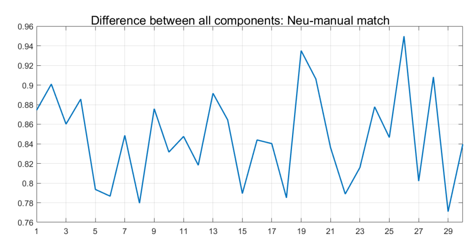

Figure 17: Correlation coefficients between the NEUROPROFILE whole dataset and the components estimated in the leave-one-out procedure.

did not affect the interpretation.

The mean frequency components showed trends in the same frequency ranges as in Fig. 3 for both dataset, with the change in intensity. The intensity change is more obvious in the mean channel components. For CMI, factor 3 and factor 4 from Fig. 10 showed the lowest and highest deviation, respectively. For NEUROPROFILE, factor 3 and 6 showed lower deviation while factor 4 showed higher variation in Fig. 15. The deviation happened because the factor components are matched base on their trended shape across five modes without regarding the intensity of the graph.

To further investigate the variability across subjects and how it affected the components, we calculated the correlation coefficients between the components computed from the whole dataset and the ones estimated during the leave-one-out procedure. We then plotted the average of the correlation coefficients against each subject left out of the data fusion decomposition as a way of visualising the effect of removing each subject from the decomposition. The CMI correlation coefficient is plotted on Fig. 16, and NEUROPROFILE on Fig. 17.

Overall, CMI had lower correlation coefficients, ranging from 0.64 to 0.88, compared to the NEUROPROFILE dataset whose coefficients ranged from 0.77 to 0.95. There was no clear association of the highest or lowest average correlation coefficient values with either children with or without deficits in any of the datasets. This means that the high correlation coefficient does not depend on the children developmental scores, and it is not associated with increasing age either. The subtle problems in CWEOE may lead to a higher correlation which needs further investigation.

4.3. Developmental score prediction

The predicted scores resulting from each fold are converted from z-score to the same standard space of mean 100 and standard deviation 10 to assist the analysis and visualisation. The overall percentage error of score prediction of each outer cross-validation fold are contained in Table 3.

Table 3: Predicted error of each validation folds

Model\Fold	1	2	3	4	5	6	7	8	9	10
CMI dataset										
SVM EEG prediction	9.5719	14.7207	11.5529	10.3688	11.2068	19.3905	21.2143	10.3974	13.0083	9.7004
SVM MRI prediction	12.64507	14.5719	13.2557	8.9693	10.6705	18.6596	21.4091	9.5546	11.1205	10.4657
SVM EEG/MRI prediction	9.3347	14.4806	11.3950	10.0944	11.2877	19.1832	21.1527	10.9920	13.1007	9.7145
JEDI prediction	12.0745	16.8612	15.0429	11.7953	15.9561	19.2568	16.0751	13.8068	12.4169	16.3152
Proposed model										
CMI EEG prediction	13.0641	20.7133	12.5610	16.4141	11.4748	19.0361	19.0869	10.2391	12.4695	13.1405
CMI MRI prediction	15.6760	15.0177	13.5070	8.6404	9.31443	18.0369	21.1521	8.8519	10.1490	10.1490
CMI EEG/MRI prediction	11.8388	12.5724	16.1760	9.0825	16.1360	15.5695	20.1891	10.9429	13.6734	11.0067
NEURPROFILE dataset										
SVM EEG prediction	24.1413	15.7197	13.9533	27.5374	17.8762	21.1055	9.7579	34.8469	23.4999	12.3570
SVM MRI prediction	25.8194	18.5142	14.0955	28.8177	21.8298	20.1382	9.3017	36.4359	22.2666	13.0785
SVM EEG/MRI prediction	24.3616	15.8613	14.559	27.4186	20.5457	21.6128	9.9125	34.6633	23.7996	12.1532
JEDI prediction	40.4933	8.8701	33.8156	39.4693	11.8312	16.1594	16.8124	21.3984	54.0610	16.5694
Proposed CPD model										
EEG prediction	27.4442	7.1149	17.2356	27.8665	10.5909	17.6620	9.6584	25.8922	39.0765	19.9491
MRI prediction	33.2168	8.1200	15.0377	19.9426	11.1727	16.1544	9.3504	15.8859	37.9849	14.5687
EEG/MRI prediction	30.7246	5.1344	12.1597	21.5247	9.9495	18.4134	10.1526	19.1759	48.6983	15.1373

For the CMI dataset, all three models have similar performance when comparing the percentage error. The percentage error of SVM ranged from 9.5719-21.2143%, 8.9693-21.4091% and 9.3347-21.1527% for the prediction from EEG, MRI, and both respectively. JEDI performed well with the CMI dataset. Note that this refers only to the use of EEG data to predict scores, as JEDI does not allow the inclusion of MRI. It yielded a percentage error from 11.7953-19.2568%. Our CPD-based model achieved an error range from 13.0641-20.7133%, 8.6404-21.1521% and, 9.0825-20.1891% for prediction from EEG, MRI, and both respectively.

For the NEUROPROFILE dataset, SVM achieved the percentage error ranging from 9.7579-34.8469%, 9.3017-36.4359%, 9.9125-34.6633% for prediction from EEG, MRI, and both modalities, respectively. JEDI yielded the highest overall error of all model ranging from 8.8701-54.0610%. The CPD model yielded error of 7.1149-39.0765%, 8.1200-37.9849%, and 5.1344-48.6983% for prediction from EEG, MRI, and both respectively. SVM and the CPD model achieved similar percentage error, with SVM having a slightly narrower gap of error. The CPD model could achieve lower error, while JEDI had the broader gap and highest error compared to the other two models.

We further investigated the predicted performance at the individual level of the scores by calculating the absolute difference between real and predicted scores from SVM and CPD model then plot on Fig. 18 and Fig. 19, respectively.

Then, we tested the difference between the mean distributions of actual and predicted scores through a two-tailed Student's t-test with a 5% significance level. For the CMI dataset, all predictions from SVM are not significantly different mean from the real scores. In contrast, the CELF and SWARN scores predicted with JEDI were rejected to have equal mean to the actual scores at $p = 0.0005$ and $p < 0.0001$, respectively. The scores predicted from CPD model are in the same distribution except SWARN and CBCL predicted using EEG only. In these cases, the null hypothesis was rejected at $p = 0.0389$ and $p = 0.0057$, respectively. This confirms the benefit of including both EEG and MRI in a data fusion model based on tensor factorisation. For NEUROPROFILE, all predictions from SVM were in the same mean as the real score. The JEDI Bayley prediction was rejected to have the same mean at $p = 0.038$. In the CPD model, t-test revealed that only the behaviour score prediction from EEG data alone yielded a significant difference at $p = 0.0199$. All other statistical comparisons showed no significant differences between the predicted and actual scores.

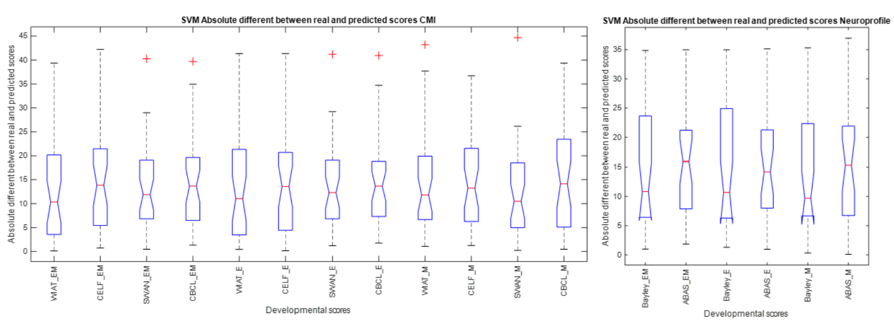

Figure 18: Absolute differences between real and predicted developmental scores from SVM at an individual level of CMI dataset (left) with WIAT-III, CELF, SWARN and CBCL scores, and NEUROPROFILE (right) with Bayley-III and ABAS-GAC scores. The score panels are arranged from left to right by the source of score estimation, predict from: both EEG and sMRI (_EM), only EEG (_E) and only sMRI (_M).

Table 4: Standard deviation of real and predicted phenotypic scores.

	CMI				NEUROPROFILE	
Scores	WIAT	CELF	SWARN	CBCL	Bayley	ABAS
Real score	17.757	17.459	14.939	17.912	17.638	17.406
SVM EEG prediction	7.411	6.416	5.479	6.637	2.137	2.705
SVM MRI prediction	1.684	1.723	1.071	1.052	1.629	2.231
SVM EEG/MRI prediction	7.479	6.362	5.365	6.412	2.245	3.211
JEDI	15.785	10.825	12.843	18.214	14.796	17.237
CPD EEG prediction	8.038	13.489	4.503	6.744	21.878	8.569
CPD MRI prediction	13.737	15.333	11.022	13.893	17.234	13.909
CPD EEG/MRI prediction	13.301	12.501	11.192	9.890	12.859	10.899

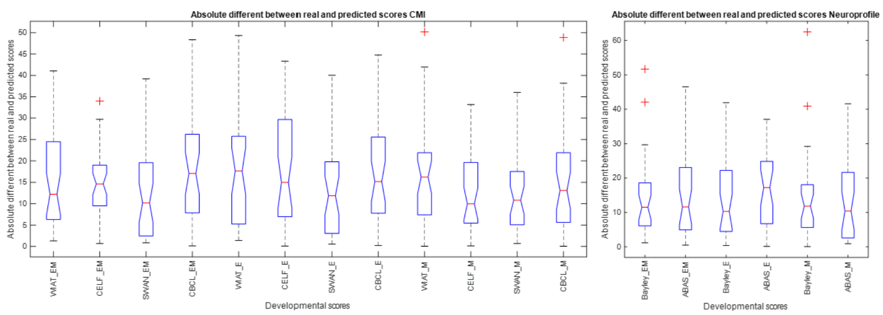

Figure 19: Absolute differences between real and predicted developmental scores from CPD model at an individual level of CMI dataset (left) with WIAT-III, CELF, SWARN and CBCL scores, and NEUROPROFILE (right) with Bayley-III and ABAS-GAC scores. The score panels are arranged from left to right by the source of score estimation, predict from: both EEG and sMRI (_EM), only EEG (_E) and only sMRI (_M).

When considering only the percentage error and mean values, SVM appears to show the lowest error across the two datasets, and it seems to be a promising approach, followed by a slightly higher error from our CPD model. However, it must be noted that the SVM results regressed toward the mean. This is demonstrated in Fig. 20 and Fig. 21 when plotting the predicted and actual scores side-by-side. This regression to the mean is more obvious in NEUROPROFILE children (Fig. 20). Moreover, when we look at the SD of the real and predict scores show in Table4, the scores estimated from SVM were significantly lower distribution compared to the real scores. This posts a problem when considering using SVM for prediction developmental score, all subject with severe deficit are predicted to be ‘normal’.

The model estimated WIAT and CBCL scores with smaller distribution than the real scores in healthy subjects, but it was able to achieve a similar distribution in others scores of both healthy and CWEOE as Fig. 21 shows.

The CPD model was quite stable when both EEG and sMRI are presented. Even though the estimated score from CPD model can point some of the deficits children out of the group and children who score 'severely low' can also be estimated by the model to be 'below average' or even 'normal'. The score prediction result from the CPD model is encouraging.

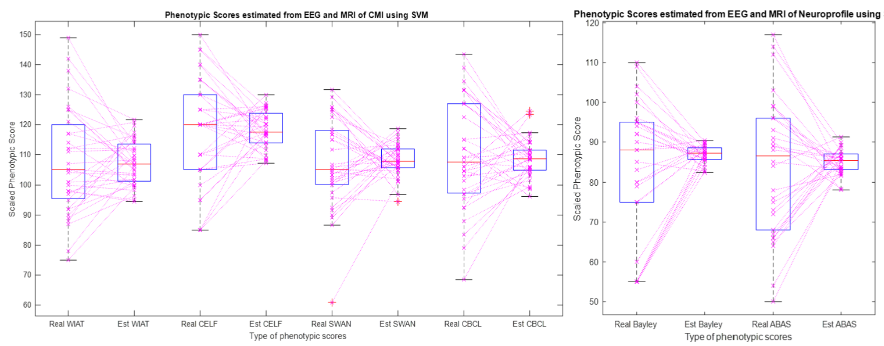

Figure 20: The scores predicted using SVM are plotted side-by-side with the real score on the left, and predicted score on the right. Both scores are linked by the pink dash line at the individual level. Left graph present the CMI dataset, while right graph present NEUROPROFILE dataset.

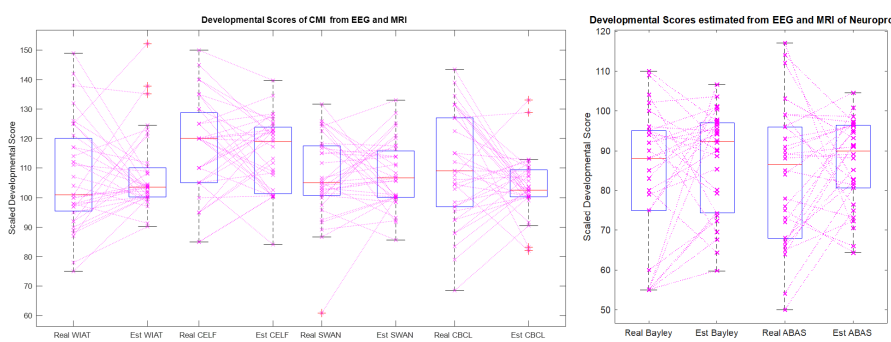

Figure 21: The side-by-side plot of scores predicted from the CPD model with the real score on the left, and predicted score on the right, linked by the pink dash line at the individual level. Left graph present the CMI dataset, while right graph present NEUROPROFILE dataset.

4.4. Limitations and future work

One can see opportunities to further refine and develop this model for clinical use. This would enable the clinician to pinpoint children at risk of developmental impairments at the time of epilepsy diagnosis without having to use time-consuming questionnaires. It would also open up the opportunity to monitoring changes over time in developmental status without concerning about the learning effects of repetitively administering the same questionnaire to a child.

Obtaining complete multimodal datasets from very young children is challenging. In this study, a proportion of data had to be excluded from analysis because EEG, phenotypic scores and/or sMRI values were missing. This results in limited sample sizes, which may affect the training process of score prediction. Future work will also try to address this issue by including additional flexibility in the model to consider missing data in some of the modalities.

In this work, SVM regression was used as the benchmark model to compare the performance of the score prediction. We let the parameters automatically optimised through its built-in function without close observation. Together with relatively small sample size, these factors might have affected the ability to generalise of the SVM regression and resulted in the predicted score regressing to the mean.

Other data fusion tools such as Coupled Matrix and Tensor Factorization (CMTF) [16] are interesting alternative options to explore and compare for the tasks of component estimation and score prediction. Additionally, alternative optimisations other than grid search can be considered. This would be particularly useful to improve the computational cost of the estimation of the hyperparameters. Moreover, the overall result seems to be affected by the strict CPD model with the assumption that the components for subject domain shared across modalities are identical, which may not be flexible enough for the nature of the data. Taking this into account, adopting the soft coupled decomposition [47] to allow the subject mode to be slightly different in order to increase the best fit of the model seem to be a promising next step.

Furthermore, this work aims to utilise existing data during the epilepsy diagnosis. Hence, we used resting-state EEG. However, our results show that resting-state EEG may not be the best option since the component profile plots and direct projection in score prediction demonstrate that it relies on sMRI more than the EEG. Further investigations are needed to confirm the use of task-based EEG and more flexible model.

5. Conclusion

Our work jointly decomposed three modalities of resting-state EEG, sMRI, and phenotypic scores to analyse the relationship between modalities. It then used this information to predict the developmental scores of unseen children. This work is motivated by the need to detect children with developmental impairments from routinely acquired clinical diagnostic modalities, thus avoiding time-consuming paper questionnaires and helping to prioritise patients for clinical follow-ups. This is particularly relevant in CWEOE. The data fusion based on the CPD decomposition revealed relationships that agreed with prior clinical knowledge in a data-driven way. The score prediction result is promising but it also points out the need for a bigger sample size with diverse distributions of developmental scores and the need for a more flexible data fusion model. However, this is a prominent first step toward the study between functional, structural, and phenotypic data of pre-school children to benefit the developmental score prediction from EEG and sMRI.

Acknowledgement

The authors would like to thank Dr. Eli Kinney-Lang for providing useful advice throughout the experimentation and all the children and families involved in the CMI and NEUROPROFILES.

This work was supported by a PhD studentship to N. Dron by the Royal Thai Government Scholarship. It has also been partially funded by the University of Salamanca and the Society of Spanish Researchers in the United Kingdom

(SRUK/CERU, www.sruk.org.uk) under the Mobility Program SRUK/CERU "On the Move".

References

[1] T. Gasser, R. Verleger, P. Bächer, L. Sroka, Development of the EEG of school-age children and adolescents. I. Analysis of band power, Electroencephalography and Clinical Neurophysiology 69 (2) (1988) 91–99. doi:10.1016/0013-4694(88)90204-0.

[2] R. J. Somsen, B. J. van't Klooster, M. W. van der Molen, H. M. van Leeuwen, R. Licht, Growth spurts in brain maturation during middle childhood as indexed by eeg power spectra, Biological psychology 44 (3) (1997) 187–209.

[3] R. C. Knickmeyer, S. Gouttard, C. Kang, D. Evans, K. Wilber, J. K. Smith, R. M. Hamer, W. Lin, G. Gerig, J. H. Gilmore, A structural MRI study of human brain development from birth to 2 years., The Journal of neuroscience 28 (47) (2008) 12176–82. doi:10.1523/JNEUROSCI.3479-08.2008.

[4] J. C. Henry, Electroencephalography: basic principles, clinical applications, and related fields, Neurology 67 (11) (2006) 2092–2092.

[5] M. D. Holmes, D. M. Tucker, J. M. Quiring, S. Hakimian, J. W. Miller, J. G. Ojemann, Comparing noninvasive dense array and intracranial electroencephalography for localization of seizures, Neurosurgery 66 (2) (2010) 354–362.

[6] T. M. Mayhew, D. R. Olsen, Magnetic resonance imaging (mri) and model-free estimates of brain volume determined using the cavalieri principle., Journal of anatomy 178 (1991) 133–144.

[7] S. Liu, W. Cai, S. Liu, F. Zhang, M. Fulham, D. Feng, S. Pujol, R. Kikinis, Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders, Brain informatics 2 (3) (2015) 167–180.

[8] F. Cendes, W. H. Theodore, B. H. Brinkmann, V. Sulc, G. D. Cascino, Neuroimaging of epilepsy, in: Handbook of clinical neurology, Vol. 136, Elsevier, 2016, pp. 985–1014.

[9] G. Motamedi, K. Meador, Epilepsy and cognition, Epilepsy & Behavior 4 (2003) 25–38.

[10] K. Rantanen, K. Eriksson, P. Nieminen, Cognitive impairment in preschool children with epilepsy, Epilepsia 52 (8) (2011) 1499–1505. `doi:10.1111/j.1528-1167.2011.03092.x`.

[11] B. Hermann, J. Jones, R. Sheth, C. Dow, M. Koehn, M. Seidenberg, Children with new-onset epilepsy: Neuropsychological status and brain structure, Brain 129 (10) (2006) 2609–2619. `doi:10.1093/brain/awl196`.

[12] M. P. Kerr, S. Mensah, F. Besag, B. De Toffol, A. Ettinger, K. Kanemoto, A. Kanner, S. Kemp, E. Krishnamoorthy, W. C. LaFrance Jr, et al., International consensus clinical practice statements for the treatment of neuropsychiatric conditions associated with epilepsy, Epilepsia 52 (11) (2011) 2133–2138.

[13] M. Yoong, Quantifying the deficit—imaging neurobehavioural impairment in childhood epilepsy, Quantitative imaging in medicine and surgery 5 (2) (2015) 225–237.

[14] H. M. Braakman, M. J. Vaessen, P. A. Hofman, M. H. Debeij-van Hall, W. H. Backes, J. S. Vles, A. P. Aldenkamp, Cognitive and behavioral complications of frontal lobe epilepsy in children: a review of the literature, Epilepsia 52 (5) (2011) 849–856.

[15] S. Baxendale, Neuropsychological assessment in epilepsy, Practical Neurology 18 (1) (2018) 43–48. `doi:10.1136/practneurol-2017-001827`.

[16] E. Acar, Y. Levin-Schwartz, V. D. Calhoun, T. Adali, Acmtf for fusion of multi-modal neuroimaging data and identification of biomarkers, in: 2017 25th European Signal Processing Conference (EUSIPCO), IEEE, 2017, pp. 643–647.

[17] L. M. Alexander, J. Escalera, L. Ai, C. Andreotti, K. Febre, A. Mangone, N. Vega-Potler, N. Langer, A. Alexander, M. Kovacs, et al., An open resource for transdiagnostic research in pediatric mental health and learning disorders, Scientific data 4 (2017) 170181.

[18] E. Kinney-Lang, A. Ebied, B. Auyeung, R. F. Chin, J. Escudero, Introducing the joint eeg-development inference (jedi) model: A multi-way, data fusion approach for estimating paediatric developmental scores via eeg, IEEE Transactions on Neural Systems and Rehabilitation Engineering 27 (3) (2019) 348–357.

[19] N. Dron, E. Kninney-Lang, R. Chin, J. Escudero, Preliminary fusion of eeg and mri with phenotypic scores in children with epilepsy based on the canonical polyadic decomposition, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 3884–3887.

[20] M. Yoong, M. Hunter, J. Stephen, A. Quigley, J. Jones, J. Shetty, A. McLellan, M. E. Bastin, R. F. Chin, Cognitive impairment in early onset epilepsy is associated with reduced left thalamic volume, Epilepsy & Behavior 80 (2018) 266–271.

[21] C. Garcia-Ramos, D. C. Jackson, J. J. Lin, K. Dabbs, J. E. Jones, D. A. Hsu, C. E. Stafstrom, L. Zawadzki, M. Seidenberg, V. Prabhakaran, et al., Cognition and brain development in children with benign epilepsy with centrotemporal spikes, Epilepsia 56 (10) (2015) 1615–1622.

[22] A. L. Krain, F. X. Castellanos, Brain development and adhd, Clinical psychology review 26 (4) (2006) 433–444.

[23] T. G. Kolda, B. W. Bader, Tensor Decompositions and Applications, SIAM Review 51 (3) (2009) 455–500. doi:10.1137/07070111X.

[24] E. Kinney-Lang, L. Spyrou, A. Ebied, R. F. Chin, J. Escudero, Tensor-driven extraction of developmental features from varying paediatric EEG datasets, Journal of Neural Engineering 15 (4). doi:10.1088/1741-2552/aac664.

[25] J. Escudero, E. Acar, A. Fernández, R. Bro, Multiscale entropy analysis of resting-state magnetoencephalogram with tensor factorisations in alzheimer's disease, Brain research bulletin 119 (2015) 136–144.

[26] E. Acar, T. G. Kolda, D. M. Dunlavy, All-at-once optimization for coupled matrix and tensor factorizations, arXiv preprint arXiv:1105.3422.

[27] V. D. Calhoun, T. Adali, G. D. Pearlson, K. A. Kiehl, Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data, Neuroimage 30 (2) (2006) 544–553.

[28] T. Adali, Y. Levin-Schwartz, V. D. Calhoun, Multimodal data fusion using source separation: Two effective models based on ica and iva and their properties, Proceedings of the IEEE 103 (9) (2015) 1478–1493.

[29] M. A. Akhonda, Y. Levin-Schwartz, S. Bhinge, V. D. Calhoun, T. Adali, Consecutive independence and correlation transform for multimodal fusion: Application to eeg and fmri data, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 2311–2315.

[30] T. Adali, Y. Levin-Schwartz, V. D. Calhoun, Multimodal Data Fusion Using Source Separation: Application to Medical Imaging, Proceedings of the IEEE 103 (9) (2015) 1494–1506. doi:10.1109/JPROC.2015.2461601.

[31] J. Sui, E. Castro, H. He, D. Bridwell, Y. Du, G. D. Pearlson, T. Jiang, V. D. Calhoun, Combination of fmri-smri-eeg data improves discrimination of

schizophrenia patients by ensemble feature selection, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 3889–3892.

[32] C. Stamile, F. Cotton, D. Sappey-Marinier, S. Van Huffel, Tensor Based Blind Source Separation in Longitudinal Magnetic Resonance Imaging Analysis, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 3879–3883.

[33] O. Alter, P. O. Brown, D. Botstein, Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms, Proceedings of the National Academy of Sciences 100 (6) (2003) 3351–3356.

[34] E. F. Lock, K. A. Hoadley, J. S. Marron, A. B. Nobel, Joint and individual variation explained (jive) for integrated analysis of multiple data types, The annals of applied statistics 7 (1) (2013) 523.

[35] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, T. Ristaniemi, Tensor decomposition of eeg signals: a brief review, Journal of neuroscience methods 248 (2015) 59–69.

[36] L. Sorber, M. Van Barel, L. De Lathauwer, Structured Data Fusion, IEEE Journal of Selected Topics in Signal Processing 9 (4) (2015) 586–600. `doi:10.1109/JSTSP.2015.2400415`.

[37] R. A. Harshman, et al., Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis.

[38] E. Martınez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, M. S. Cohen, Concurrent eeg/fmri analysis by multiway partial least squares, NeuroImage 22 (3) (2004) 1023–1034.

[39] S. Van Eyndhoven, B. Hunyadi, L. De Lathauwer, S. Van Huffel, Flexible fusion of electroencephalography and functional magnetic resonance imag-

ing: Revealing neural-hemodynamic coupling through structured matrix-tensor factorization, in: 2017 25th European Signal Processing Conference (EUSIPCO), IEEE, 2017, pp. 26–30.

[40] Y. Jonmohamadi, S. Muthukumaraswamy, J. Chen, J. Roberts, R. Crawford, A. Pandey, Extraction of common task features in eeg-fmri data using coupled tensor-tensor decomposition, Brain Topography 33 (5) (2020) 636–650.

[41] R. Mosayebi, G.-A. Hossein-Zadeh, Correlated coupled matrix tensor factorization method for simultaneous eeg-fmri data fusion, Biomedical Signal Processing and Control 62 (2020) 102071.

[42] E. Acar, Y. Levin-Schwartz, V. D. Calhoun, T. Adali, Tensor-based fusion of eeg and fmri to understand neurological changes in schizophrenia, in: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2017, pp. 1–4.

[43] B. Hunyadi, P. Dupont, W. Van Paesschen, S. Van Huffel, Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 7 (1) (2017) e1197.

[44] S. Van Eyndhoven, P. Dupont, S. Tousseyn, N. Vervliet, W. Van Paesschen, S. Van Huffel, B. Hunyadi, Augmenting interictal mapping with neurovascular coupling biomarkers by structured factorization of epileptic eeg and fmri data, NeuroImage 228 (2021) 117652.

[45] E. Acar, C. Schenker, Y. Levin-Schwartz, V. D. Calhoun, T. Adali, Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data, Frontiers in neuroscience 13 (2019) 416.

[46] S. Ferdowsi, V. Abolghasemi, S. Sanei, A new informed tensor factorization

approach to eeg–fmri fusion, Journal of neuroscience methods 254 (2015) 27–35.

[47] C. Chatzichristos, M. Davies, J. Escudero, E. Kofidis, S. Theodoridis, Fusion of eeg and fmri via soft coupled tensor decompositions, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 56–60.

[48] C. Chatzichristos, E. Kofidis, L. De Lathauwer, S. Theodoridis, S. Van Huffel, Early soft and flexible fusion of eeg and fmri via tensor decompositions, arXiv preprint arXiv:2005.07134.

[49] L. De Lathauwer, Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and Uniqueness, SIAM Journal on Matrix Analysis and Applications 30 (3) (2008) 1033–1066. doi:10.1137/070690729.
URL http://www.siam.org/journals/ojsa.php

[50] C. Chatzichristos, E. Kofidis, M. Morante, S. Theodoridis, Blind fmri source unmixing via higher-order tensor decompositions, Journal of neuroscience methods 315 (2019) 17–47.

[51] E. Kinney-Lang, L. Spyrou, A. Ebied, R. Chin, J. Escudero, Elucidating age-specific patterns from background electroencephalogram pediatric datasets via PARAFAC, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 3797–3800. doi:10.1109/EMBC.2017.8037684.

[52] A. Anderson, P. K. Douglas, W. T. Kerr, V. S. Haynes, A. L. Yuille, J. Xie, Y. N. Wu, J. A. Brown, M. S. Cohen, Non-negative matrix factorization of multimodal mri, fmri and phenotypic data reveals differential changes in default mode subnetworks in adhd, NeuroImage 102 (2014) 207–219.

[53] D. Wechsler, Wechsler Individual Achievement Test 3rd Edition (WIAT III), London: The Psychological Corporation, 2009a.

[54] E. H. Wiig, W. A. Secord, E. Semel, Clinical evaluation of language fundamentals: CELF-5, Pearson, 2013.

[55] T. M. Achenbach, Manual for the child behavior checklist/4-18 and 1991 profile, University of Vermont, Department of Psychiatry.

[56] J. Swanson, S. Schuck, M. Mann, C. Carlson, K. Hartman, J. Sergeant, R. McCleary, The swan rating scale.
URL `http://www.adhd.net`

[57] M. B. Hunter, R. Sumpter, K. Verity, M. Yoong, A. Mclellan, J. Shetty, R. F. Chin, Neurodevelopment in Preschool Children Of Fife and Lothian Epilepsy Study: Neuroprofiles-a population-based study, Developmental Medicine & Child Neurology (57) (2015) 56–57.

[58] N Bayley, Bayley scales of infant development: Manual, Psychological Corporation.

[59] P. L. Harrison, T. Oakland, Adaptive Behavior Assessment System: Third Edition, in: Encyclopedia of Clinical Neuropsychology, 2018, pp. 57–60. `doi:10.1007/978-3-319-57111-9{\_}1506`.

[60] R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data., Computational intelligence and neuroscience 2011. `doi:10.1155/2011/156869`.

[61] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, S. M. Smith, FSL, NeuroImage 62 (2) (2012) 782–790. `doi:10.1016/j.neuroimage.2011.09.015`.

[62] B. Patenaude, Bayesian Statistical Models of Shape and Appearance for Subcortical Brain Segmentation, NeuroImage 56 (3) (2011) 907–922.

[63] J. B. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, Linear algebra and its applications 18 (2) (1977) 95–138.

[64] N. Vervliet, O. Debals, L. De Lathauwer, Tensorlab 3.0—numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization, in: 2016 50th Asilomar Conference on Signals, Systems and Computers, IEEE, 2016, pp. 1733–1738.

[65] M. Sørensen, I. Domanov, L. De Lathauwer, Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-(l_r,n,l_r,n,1) terms—part ii: Algorithms, SIAM Journal on Matrix Analysis and Applications 36 (3) (2015) 1015–1045.

[66] C. Cortes, V. Vapnik, Support vector machine, Machine learning 20 (3) (1995) 273–297.

[67] M.-P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, H. Soltanian-Zadeh, Multimodal data analysis of epileptic eeg and rs-fmri via deep learning and edge computing, Artificial Intelligence in Medicine 104 (2020) 101813.

[68] Y. Höller, K. H. Butz, A. C. Thomschewski, E. V. Schmid, C. D. Hofer, A. Uhl, A. C. Bathke, W. Staffen, R. Nardone, F. Schwimmbeck, et al., Prediction of cognitive decline in temporal lobe epilepsy and mild cognitive impairment by eeg, mri, and neuropsychology, Computational intelligence and neuroscience 2020.

[69] G. Hwang, B. Hermann, V. A. Nair, L. L. Conant, K. Dabbs, J. Mathis, C. J. Cook, C. N. Rivera-Bonet, R. Mohanty, G. Zhao, et al., Brain aging in temporal lobe epilepsy: Chronological, structural, and functional, NeuroImage: Clinical 25 (2020) 102183.

[70] V. Kecman, T.-M. Huang, M. Vogt, Iterative single data algorithm for training kernel machines from huge data sets: Theory and performance, in: Support vector machines: Theory and Applications, Springer, 2005, pp. 255–274.

[71] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo,

Y. Schwartz, B. Thirion, Assessing and tuning brain decoders: cross-validation, caveats, and guidelines, NeuroImage 145 (2017) 166–179.

[72] Y. Xie, Y. A. Chen, M. D. De Bellis, The relationship of age, gender, and iq with the brainstem and thalamus in healthy children and adolescents: a magnetic resonance imaging volumetric study, Journal of child neurology 27 (3) (2012) 325–331.

[73] G. T. Voelbel, M. E. Bates, J. F. Buckman, G. Pandina, R. L. Hendren, Caudate nucleus volume and cognitive performance: are they related in childhood psychopathology?, Biological psychiatry 60 (9) (2006) 942–950.

[74] T. Sefcsik, D. Nemeth, K. Janacsek, I. Hoffmann, J. Scialabba, P. Klivenyi, G. G. Ambrus, G. Haden, L. Vecsei, The role of the putamen in cognitive functions—a case study, Learning & Perception 1 (2) (2009) 215–227.

[75] T. M. Wellington, M. Semrud-Clikeman, A. L. Gregory, J. M. Murphy, J. L. Lancaster, Magnetic resonance imaging volumetric analysis of the putamen in children with adhd: combined type versus control, Journal of attention disorders 10 (2) (2006) 171–180.

[76] I. Ivanov, R. Bansal, X. Hao, H. Zhu, C. Kellendonk, L. Miller, J. Sanchez-Pena, A. M. Miller, M. M. Chakravarty, K. Klahr, et al., Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder, American Journal of Psychiatry 167 (4) (2010) 397–408.

[77] F. Castanedo, A review of data fusion techniques, ScientificWorldJournal 2013 (2013) 704504. `doi:10.1155/2013/704504`.
URL `http://www.ncbi.nlm.nih.gov/pubmed/24288502`

[78] P. J. Marshall, Y. Bar-Haim, N. A. Fox, Development of the eeg from 5 months to 4 years of age, Clinical Neurophysiology 113 (8) (2002) 1199–1208.

[79] A. Delorme, S. Makeig, Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, Journal of neuroscience methods 134 (1) (2004) 9–21.